UNIVERSITY OF TROMSØ UIT

FACULTY OF SCIENCE
DEPARTMENT OF PHYSICS AND TECHNOLOGY

Statistical Analysis of Multilook Polarimetric Radar Images with the Mellin Transform

Stian Normann Anfinsen

A dissertation for the degree of
Philosophiae Doctor

May 2010

To my family

«The true sign of intelligence is
not knowledge but imagination.»
Albert Einstein

«Nobody creates anything. It's there,
and you just f****** grab a hold of it.»
Keith Richards

Abstract

This thesis presents methods for statistical analysis of the probability distributions used to model multilook polarimetric radar images. The methods are based on a matrix-variate version of Mellin's integral transform.

The proposed theoretical framework is referred to as Mellin kind statistics. It is an extension of a theory recently developed for single polarisation amplitude and intensity data to the complex matrix-variate case describing multilook polarimetric images. This generalisation is made possible by the rediscovery of a generalised Mellin transform, which is defined for functions of positive definite Hermitian matrices. The domain makes it suited for application to the distributions used to model the polarimetric covariance and coherency matrix.

The analysis tools include the matrix-variate Mellin kind characteristic function, which is defined with the Mellin transform in place of the conventional Fourier transform. Matrix log-moments and matrix log-cumulants are retrieved from this function. The matrix log-cumulants are used in a moment based approach to parameter estimation of the distribution parameters. The estimators make efficient use of all the statistical information in the polarimetric covariance matrix, and are superior to all known alternatives. The matrix log-cumulants are also used to construct the first known goodness-of-fit test for matrix distributions based on the multilook polarimetric product model. The algorithms are interpreted by means of a highly informative graphical visualisation tool displaying a space spanned by certain matrix log-cumulants.

It is demonstrated that the matrix-variate Mellin transform is the natural tool for analysing multilook polarimetric radar images. This conclusion is based on the simple and elegant mathematical expressions obtained, the superb statistical properties of developed estimators, as well as the intuitive interpretations offered by the Mellin kind statistics.

Acknowledgements

I would like to express my gratitude to my supervisor, Torbjørn Eltoft, for giving me the chance to study an exciting subject, both in terms of mathematical challenges and its relevance to important issues in environmental monitoring. He has allowed me unrestricted freedom in pursuing my own ideas and interests, provided knowledgable guidance when that was needed, and shown great consideration during the battle with my thesis. Through all stages, he has been a pleasant and good-tempered colleague, as much as boss.

Deep thanks go to fellow Ph.D. student Tony Doulgeris for being a great office mate and the best sparring partner one could desire, for unlimited sharing of his knowledge and ideas, and for the enthusiasm he has shown for my work. I can only try to return a fraction of it. Thanks also to Robert Jenssen and Camilla Brekke for inspiring collaborations that are hopefully only in an early phase, and to Robert for his tutoring on kernel methods and proofreading of my thesis. I also thank the members of the steadily expanding earth observation research group, as well as other members of the department, for contributing to a good working environment.

Special thanks go to Sebastiano Bruno Serpico for welcoming me to his research lab at the University of Genoa as a guest researcher. Thanks to both Bruno and Gabriele Moser for sharing of their insight and providing inspiration that shaped the further work on my thesis. Gabriele is also gratefully acknowledged for his comments on my manuscripts. I am indebted to Simona Castellana, Maciel Zortea and Michaela de Martino for their kind assistance on various practical issues, and would like to thank everyone affiliated with the lab for making my stay in Genoa a very enjoyable one.

I extend my thanks to Per Ivar Emanuelsen for technical support and Johan Werner Lyshaug for being a life saviour. I further thank Larry Page, Sergey Brin, Jimmy Wales and Larry Sanger for redefining the way research is conducted. I also want to thank Shaun Quegan, Florence Tupin and Alfred Hanssen for serving on my committee.

Equally important contributions have been made by those helping me to forget about the world of radar polarimetry and matrix statistics. In that concern, I am fortunate to have friends like Björn, Kevin, Asbjørn, Vegar, Anders and Lena with associates. You have all done a great job. Thanks also to members of the squash team and the squad of fallen football stars.

My family has supported me wholeheartedly in my endeavours from I first became student and to this day. My father-in-law has only added to this support. Most of all, I thank Ann-Tove for standing by my side throughout this journey, for all the good moments we have shared, for her patience, encouragement, comfort and entertainment, and for setting me straight whenever I started to talk in maths and buzz like a fridge. Finally, my biggest achievements during the course of my Ph.D. program are named Bjørnar and Sigve, who are the most pleasant diversions from science imaginable. I am forever grateful to you all.

Stian — Tromsø, 5th May 2010

Contents

List of Tables

List of Figures

Nomenclature

List of Notation

A	amplitude (of scattering coefficient)
$\mathcal{A}$	set of amplitude components
$\mathbf{C}$	polarimetric covariance matrix
$\mathbb{C}$	complex plane
d	number of polarimetric channels
e	Euler's number
E	electromagnetic field component
f	probability density function
$\mathcal{M}$	Fisher-Snedecor distribution
g	arbitrary scalar function defined on Ω_+
$\mathcal{G}^0$	G^0 distribution
h	arbitrary scalar function defined on Ω_+
h_{rms}	root mean square height of rough surface
I	intensity
I_L	multilook intensity
$\mathbf{I}_d$	identity matrix with dimension d
$\jmath$	imaginary unit
k	wave number
$\mathbf{k}$	Pauli basis scattering vector
$\mathcal{K}$	K distribution
L	equivalent/nominal number of looks
$\mathbf{M}$	multichannel texture matrix
$\mathcal{M}$	M distribution
$\mathfrak{M}$	Mellin transform
$\mathfrak{M}^{-1}$	inverse Mellin transform
N	number of scatterers
P_{tot}	Total measured power
r	received electromagnetic wave/field
$\mathbb{R}^+$	line of positive real numbers
s	complex transform variable
S	scattering coefficient

$\mathbf{s}$	lexicographic basis scattering vector
$\mathbf{S}$	scattering matrix
t	transmitted electromagnetic wave/field
T	texture variable
$\mathbf{U}$	unitary transformation matrix
$\mathcal{U}$	U distribution
$\mathbf{W}$	complex Wishart distributed matrix
$\widetilde{\mathbf{W}}$	scaled complex Wishart distributed matrix
$\mathcal{W}$	W distribution
$\mathcal{W}_d^{\mathbf{C}}$	complex Wishart distribution
$s\mathcal{W}_d^{\mathbf{C}}$	scaled complex Wishart distribution
$\mathbf{X}$	complex matrix in Ω_+
$\mathbf{Y}$	complex matrix in Ω_+
z	complex scalar
α	shape parameter
β	beta distribution
β^{-1}	inverse beta distribution
γ	gamma distribution
γ^{-1}	inverse gamma distribution
θ	phase (of scattering coefficient)
$\boldsymbol{\theta}$	vector of texture parameters
Θ	set of phase components
κ_ν	νth-order matrix-variate Mellin kind cumulant (matrix log-cumulant)
λ	shape parameter
μ	location parameter
μ_ν	νth-order matrix-variate Mellin kind moment (matrix log-moment)
ν	order parameter
ρ	distance between radar and target
σ^2	mean radar cross section
Σ	scale matrix of matrix distributions
ϕ	matrix-variate Mellin kind characteristic function
φ	matrix-variate Mellin kind cumulant generating function
Ω_+	cone of positive definite Hermitian matrices

List of Acronyms

AD	Anderson-Darling
ALOS	Advanced Land Observing Satellite
ASAR	Advanced Synthetic Aperture Radar
CCRS	Canadian Centre for Remote Sensing
CCT	Centre Canadien de télédétection
CDF	cumulative distribution function
CF	characteristic function
CGF	cumulant generating function
COSMO-SkyMed	COnstellation of small Satellites for the Mediterranean basin Observation
CvM	Cramér-von Mises
CW	continuous wave
EM	electromagnetic
ENL	equivalent (or effective) number of looks
ERS	European Remote Sensing satellite
FM-CW	frequency modulated continuous wave
FT	Fourier transform
GoF	goodness-of-fit
GPR	ground penetrating radar
IEEE	Institute of Electrical & Electronics Engineers
IID	independent and identically distributed
JERS	Japanese Earth Resources Satellite
JPL	Jet Propulsion Laboratory
KS	Kolmogorov-Smirnov
MKS	Mellin kind statistics
ML	maximum likelihood
MLC	matrix log-cumulant
MLC	multilook complex
MLM	matrix log-moment
MoLC	method of log-cumulants
MoMLC	method of matrix log-cumulants
MT	Mellin transform
NASA	National Aeronautics and Space Administration
PALSAR	Phased Array type L-band Synthetic Aperture Radar
PDF	probability density function
RAR	real aperture radar
RV	random variable
SAR	synthetic aperture radar
SF-CW	step frequency continuous wave
SLC	single-look complex

Chapter 1

Introduction

The first part of this chapter motivates the application of the Mellin Transform to the analysis of polarimetric radar images. The second and third part provides an extended summary of the chapters of the thesis, including three journal publications. A list of other publications produced is included at the end.

1.1 Motivation

The univariate *Mellin transform* is an integral transform named after the Finnish mathematician *Robert Hjalmar Mellin* (1854-1933), which has found many applications in mathematics, statistics, physics and engineering. It was first applied to statistical models of radar images by *Jean-Marie Nicolas*, a French professor of signal and image processing. In a milestone paper [Nicolas, 2002], he presented a new framework for statistical analysis of distributions[1] of single polarisation amplitude and intensity images. The paper was followed by a comprehensive technical report [Nicolas, 2006], which laid the foundation for a new approach to deduction of sophisticated distribution models, including their functional characterisation, the expressions for their statistical moments, and estimators of the model parameters.

In the years following the seminal publications, a large number of papers have appeared that use the analysis framework of Nicolas to solve many different radar image analysis problems. The list covers applications such as statistical modelling [Moser et al., 2006a, Moser et al., 2006b, Bombrun and Beaulieu, 2008], speckle filtering [Nicolas, 2003, Achim et al., 2006, Chen and Liu, 2008], image classification [Tison et al., 2004], image segmentation [Benboudjema et al., 2007, Galland et al., 2009], change detection [Bujor et al., 2004, Moser and Serpico, 2006, Moser and Serpico, 2009], estimation of interferometric coherence [Abdelfattah and Nicolas, 2006] and image compression [Valade and Nicolas, 2004]. Still it seems like the awareness about the strength of the method is increasing rapidly.

[1] The term *distribution* is used in this thesis as equivalent to probability density function.

Figure 1.1: Three innovators who have provided the theoretical underpinning for this thesis: Hjalmar Mellin, Jean-Marie Nicolas and Arak Mathai (left to right).

During the course of my research on radar image classification, practical requirements urged me to look into certain estimation problems for parameters of multilook polarimetric radar data distributions. The solutions I obtained bore strong similarities with those achieved by Mellin transform methods, but looked like matrix-variate extensions. This was the starting point of my quest for a matrix-variate generalisation of the Mellin transform, which proved successful. It was found in shape of the so-called *M-transform*, a lesser known integral transform for functions defined on positive definite Hermitian matrices. It was proposed by the Indian/Canadian statistician *Arakaparambil Mathai Mathai* in [Mathai, 1978] and extended to complex matrices in [Mathai, 1997]. Just like the univariate Mellin transform is tailor-made for distributions of real positive random variables, the matrix-variate M-transform went hand in glove with the covariance matrix distributions used to describe multilook polarimetric radar data.

The rediscovery of the M-transform opened the door for an extension of the *Mellin kind statistics*, which is the name used for the theoretical framework of Nicolas, to multilook polarimetric radar data. The prospect of achieving results with an impact comparable to those of Nicolas was sufficient to make me redefine my Ph.D. project completely. For one reason, the matrix-variate distributions describing polarimetric covariance matrices contain special functions and complicated forms that severely limit their mathematical tractability by conventional methods. Their applicability would certainly benefit from a method which provided a new perspective on functional characterisation, computation of statistical moments, model visualisation and parameter estimation. If the complexity of these distribution cannot be handled, the only rescue is to resort to the mathematically simpler but less accurate Wishart distribution model, which in many situations restricts the quality of the model based inference. Secondly, it was envisioned that the methods based on the Mellin transform would make possible the full use of the statistical information contained in the multilook polarimetric data, including the correlations between the polarimetric channels, and not only the single polarisation intensities. Furthermore, an extension of the Mellin kind statistics would provide a general

theory treating single polarisation images as a special case. The results presented in the current thesis proves in my opinion that the change of direction was a right decision.

It was advocated above that Mellin kind statistics have a positive effect on the applicability of certain distributions that are mathematically complex, but provide better fit with real data than the simpler alternative. The distributions referred to are those arising from the *doubly stochastic product model* for multilook polarimetric radar images. It is therefore relevant to comment on the necessity of such models. For low resolution radar images, each resolution cell contains a high number of microwave scatterers. The scattered electromagnetic field is the coherent sum of contributions from all these scatterers, and the central limit theorem asserts that it can be accurately modelled by Gaussian statistics. With the steadily improving spatial resolution of operational synthetic aperture radars, the Gaussian assumption is frequently challenged and often fails, in particular for scenes of urban environment, but also for natural surfaces such as forest and sea. Concerning the polarimetric aspect, it should be expected that the technological evolution will gradually replace single polarisation radars with instruments that have increasing capabilities for polarimetry. Hence, the need for adequate polarimetric distribution models will be more and more emergent.

After having argued in terms of technicalities, a more fundamental question naturally arises: Why is statistical modelling of radar data an important research task? When we zoom out and look at the benefit for society and mankind, the importance is connected to the value of the imaging radar as a remote sensing instrument. In the context of Earth observation, the radar is distinguished by its all-weather and all-season capabilities. It performs its measurements irrespective of cloud cover and sun conditions. Together with the wide spatial coverage and relatively good temporal resolution of the image acquisitions, these properties make spaceborne radar crucial for tasks such as monitoring of rain forest degradation and deforestation, change detection in Arctic glaciers, and mapping of sea ice conditions. A common aspect of these applications is that they require observations of inaccessible areas that are impossible to cover by on-site measurements. Radar remote sensing is both a practical and a cost effective alternative. The limiting factor of radar images is their content of strong speckle or clutter, an inherent feature of the coherent imaging process which complicates the interpretation and potential for information extraction. The most efficient remedy is to analyse the images within the context of a suitable statistical model, which closes the argument.

1.2 Chapter Review

Chapter 2 provides an introduction to radar imaging, which starts at the very fundamental by discussing key properties of active microwave sensors, describing different frequency bands and different types of imaging radars. The function of synthetic aperture radars and polarimetric radars is explained. We next look at the data formats delivered by polarimetric radars, starting with single-look complex data and moving on to multilook complex data, while explaining the concept of multilooking. The chapter

ends with the presentation of a physical model of radar speckle, and an explanation of its link to the traditional statistical models for the radar measurements. We define speckle and texture, the two factors of the doubly stochastic product model, which forms the basis of our statistical analysis.

Chapter 3 is opened by a historical review of the Mellin transform, with particular focus on its use in statistical distribution theory. The underlying ideas of Nicolas' univariate Mellin kind statistics are outlined, before we present Mathai's generalised Mellin transform. The original contribution of the author starts with the derivations of fundamental properties of the matrix-variate Mellin transform. It is followed by the new definitions of matrix-variate Mellin kind statistics. Finally, the Mellin kind statistics of a general multilook polarimetric product model are revealed, while leaving the detailed derivations to Paper 2.

Chapter 4 starts with the definitions of the special functions needed in the sequel. It further introduces candidate univariate distributions that can be used to model texture, together with their Mellin kind statistics, that were derived in [Nicolas, 2006]. The scaled complex Wishart distribution is also presented as a model for fully developed speckle, unmodulated by texture. The Mellin kind statistics given for this distribution are contributed in Paper 2. With the listed texture and speckle distributions as building block, we arrive at five compound distributions for the polarimetric covariance matrix. The Mellin kind statistics derived for all distributions are new.

Chapters 5–7 contain Papers 1–3. They are described separately in the next section.

Chapter 8 gives the conclusions. It lists the main results and points out future directions of research based on the work documented in the thesis.

Appendix A is a conference paper (referenced as Paper 4 in the list of Section 1.4) which contains ideas about statistical modelling of speckle filtered multilook polarimetric radar data. In is included for completeness and availability, since the results are mentioned and referenced in the thesis.
Appendix B is a translation of [Nicolas, 2002], which is a key reference for this thesis. The paper was originally published in French, but translated by the current author for the benefit of the research reported in this thesis. It is included here, with permission from the author, for the convenience of readers who are not proficient in French.

1.3 Publication Review

The following three publications are included in this thesis as Chapters 5, 6 and 7. A summary is given for each, highlighting the original contributions of the authors. The papers appear in chronological order and document the progress of my work in terms of maturity and depth of both exposition and content. Paper 1 is least mature in the sense that the theory of Mellin kind statistics for the polarimetric case had not been formulated yet. The connection of the results to the Mellin transform was discovered

later, as documented in Paper 2. Paper 3 is a first step in the direction of applications of the theory.

Paper 1

S.N. Anfinsen, A.P. Doulgeris and T. Eltoft, "**Estimation of the Equivalent Number of Looks in Polarimetric Synthetic Aperture Radar Imagery**", IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 11, pp. 3795–3809, December 2009.

The paper studies estimators for the equivalent number of looks (ENL), a parameter which is found in all distributions used to model multilook radar images. Still, the literature on the topic is very sparse. The relation $L = \mathrm{E}\{I\}^2/\,\mathrm{Var}\{I\}$, where I denotes intensity, is often mistaken for being the definition of the ENL in the single polarisation case, assuming fully developed speckle and no texture. In reality, it is just one particular way of resolving the ENL from moments relations.

In the paper, we propose two new estimators for the ENL that are adapted to multilook polarimetric radar data. The expressions are derived by examining different kinds of moments of the polarimetric covariance matrix. The first estimator is a generalisation of the expression given above. The second is found from moments of the log-determinant of the covariance matrix, and is also found to be the maximum likelihood (ML) estimator based on the Wishart distribution model for multilook polarimetric radar data. The proposed estimators are the first ones to take the full covariance matrix as input, thereby utilising all the available statistical information. This is reflected in the experimental results in terms of superior statistical properties. The ML estimator has the lowest bias and variance, and also most robust with respect to the assumption of no texture.

An approach to unsupervised estimation of the ENL is also presented, where the ML estimator is used to compute small sample estimates over the whole image, regardless of the homogeneity in the estimation window. It is shown that a robust estimate of the ENL can be extracted from a probability density function estimate of the collection of small sample estimates. This is only possible when using the novel low variance ML estimator.

There are minor differences between the published version of the paper and the version included in the thesis. Equation (23) has been corrected in the thesis version. The journal version gives an expression which is valid for real matrices, while complex matrices are considered. The paper has been reformatted to a different font size. Otherwise, the differences are mainly orthographical, due to different preferences of English style. Some symbols have also been changed to harmonise the notation of this paper with the others.

Paper 2

S.N. Anfinsen and T. Eltoft, "**Application of the Matrix-Variate Mellin Transform to Analysis of Polarimetric Radar Images**", submitted to IEEE Transactions on Geoscience and Remote Sensing.

This paper introduces the Mellin kind statistics framework for analysis of multilook polarimetric radar images. It builds on the equivalent framework derived by Nicolas for the single polarisation case, and the extension to multilook polarimetry rests upon Mathai's matrix-variate Mellin transform. The combination of these ideas, leading to a whole new set of definitions for the polarimetric case, is an original contribution.

The paper gives a thorough review of the univariate Mellin kind statistics. It explains Nicolas' idea of introducing a new kind of characteristic function for real positive random variables by replacing the Fourier transform with the Mellin transform in the definition of the conventional characteristic function. The moments and cumulants retrieved from this characteristic function are calculated on logarithmic scale, and are therefore called log-moments and log-cumulants. We also emphasise analogies between Mellin kind statistics and conventional (Fourier kind) statistics, showing that the Mellin kind statistics are the natural tools for analysis of a multiplicative signal model, just like the conventional statistics are for the familiar additive model. This care is taken to make the presentation pedagogical, but also to make Nicolas' theory available to a wider audience, since his most comprehensive derivations are only published in French [Nicolas, 2002, Nicolas, 2006].

The novel contributions include definitions of the Mellin kind characteristic function, cumulant generating function, moments and cumulants for the matrix-variate case describing multilook polarimetric radar images. We have further defined a matrix-variate Mellin convolution and correlation, and proved corresponding convolution and correlation theorems. The convolution theorem shows that the Mellin transform of the Mellin convolution of two functions decomposes as the product of Mellin transform of the individual functions. This result is needed to formulate the Mellin kind statistics under the multilook polarimetric product model, which is probably the contribution with the largest practical significance.

As an example of applications of the new theory, we have derived parameter estimators for some product model distributions for the polarimetric covariance matrix. The experimental results show that the estimators based on Mellin kind statistics are superior to all alternative estimators from the literature. The proposed estimation procedure is interpreted visually in terms of a diagram where we plot empirical matrix log-cumulants computed from data samples together with the population matrix log-cumulants of the distributions. The matrix log-cumulant diagram is a matrix-variate extension of the diagram Nicolas has used in the univariate case.

Paper 3

S.N. Anfinsen, A.P. Doulgeris and T. Eltoft, "**Goodness-of-Fit Tests for Multilook Polarimetric Radar Data Based on the Mellin Transform**", submitted to IEEE Transactions on Geoscience and Remote Sensing.

In this paper, the Mellin kind statistics framework is used to derive goodness-of-fit tests for distributions of the polarimetric covariance matrix derived under the multilook polarimetric product model. These are, to the best of our knowledge, the first formal statistical tests that have been devised for these complicated distributions. The test statistic is constructed from the matrix log-cumulants defined in Paper 2. In order to deduce sampling distributions for the test statistics, asymptotic statistics of the matrix log-cumulants have been derived. They proposed test statistic can be applied to both simple and composite hypothesis tests.

For the simple hypothesis, the sampling distribution of the test statistic is asymptotically χ^2 distributed. We demonstrate that this is a good approximation even for moderate sample sizes. For the composite test, we must resort to Monte Carlo simulations to find the sampling distribution. This approach has a higher computational cost, but produces the true sampling distribution regardless of sample size.

The simple and composite tests have been tested on simulated and real data. Assessments of the test powers show that we have found a useful method which meets the need for formal procedures of testing model fit for compound covariance matrix distributions. The matrix log-cumulant diagram introduced in Paper 2 is further promoted as an intuitive visualisation tool for interpretation of the test procedure. As a graphical aid for informal model selection and validation, it separates very well between different distributions whose differences are mainly manifested in the heavy tails.

The paper emphasises the coupling between the problems of goodness-of-fit testing and parameter estimation. A new estimation technique for parameters of the texture distributions is motivated by the proposed tests. The estimator maximises the asymptotic likelihood of the compounded matrix distribution, and is effectively a method of moment type procedure using multiple matrix log-cumulants. The estimator is tested in the experiments of Paper 2, and exhibits superior performance in terms of both bias and variance. It is also an intrinsic part of the composite test, which requires estimation of unknown distribution parameters.

1.4 Other Publications and Presentations

As first author:

1. S. N. Anfinsen, R. Jenssen and T. Eltoft, "Clustering of polarimetric SAR data with an information theoretic kernel method," presented at the *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS '06)*, Denver, U.S., 31 Jul.-4 Aug. 2006, not published.

2. S. N. Anfinsen, R. Jenssen and T. Eltoft, "Spectral clustering of polarimetric SAR data with Wishart-derived distance measures", *Proc. 3rd Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '07)*, Frascati, Italy, 22-26 Jan. 2007, ser. ESA SP-644, Mar. 2007, 8 pp.

3. S. N. Anfinsen, A. P. Doulgeris and T. Eltoft, "Estimation of the Equivalent Number of Looks in Polarimetric SAR Imagery", *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS '08)*, vol. 4, Boston, U.S., 6-11 Jul. 2008, pp. 487–490.

4. S. N. Anfinsen, T. Eltoft and A. P. Doulgeris, "A relaxed Wishart model for polarimetric SAR data (POLinSAR '09)", *Proc. 4th Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry*, Frascati, Italy, 26-30 Jan. 2009, ser. ESA SP-668, Apr. 2009, 8 pp.

5. S. N. Anfinsen and T. Eltoft, "Moment-based goodness-of-fit tests for polarimetric radar data", presented at the *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS '09)*, Cape Town, South Africa, 12-17 Jul. 2009, not published.

6. S. N. Anfinsen and T. Eltoft, "Analysis of multilook polarimetric data with the matrix-variate Mellin transform", *Proc. 8th Eur. Conf. Synthetic Aperture Radar (EUSAR 2010)*, Aachen, Germany, 7-10 Jun. 2010, in press.

As coauthor:

6. A. P. Doulgeris, S. N. Anfinsen and T. Eltoft, "Analysis of non-Gaussian PolSAR data", *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS '08)*, Barcelona, Spain, 23-27 Jul. 2008, pp. 160–163.

7. A. P. Doulgeris, S. N. Anfinsen and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data", *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.

8. A. P. Doulgeris, S. N. Anfinsen, Y. Larsen, K. Langley and T. Eltoft, "Evaluation of polarimetric configurations for glacier classification", in *Proc. 4th Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '09)*, Frascati, Italy, 26-30 Jan. 2009, ser. ESA SP-668, Apr. 2009, 8 pp.

9. T. Eltoft, A. P. Doulgeris and S. N. Anfinsen, "Model-based statistical analysis of PolSAR data (IGARSS '09)", *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Cape Town, South Africa., 12-17 Jul. 2009, 4 pp., in press.

10. C. Brekke, S. N. Anfinsen and T. Eltoft, "Marine target detection based on dual channel SAR images", *Proc. 8th Eur. Conf. Synthetic Aperture Radar (EUSAR 2010)*, Aachen, Germany, 7-10 Jun. 2010, in press.

Chapter 2

Radar Polarimetry

This chapter gives an overview of the fundamental properties of radar imaging, with emphasis on the potential of spaceborne polarimetric synthetic aperture radar. The foundation is laid for the subsequent theoretical developments by the definition of the multilook polarimetric product model. The connections to an underlying physical model of microwave scattering are explained.

2.1 Imaging Radar

An *imaging radar*[1] illuminates the target scene with directional pulses of electromagnetic (EM) energy, measures the backscattered energy and the round-trip time, and uses this information to form an image. The radar operates in the microwave region of the EM spectrum, and the backscatter depends on dielectric and geometrical properties (roughness and shape) of the target. If both amplitude and phase of the backscattered wave is measured, the image will constitute a two-dimensional map of the EM *scattering coefficient*, which is defined in Section 2.2.2. The scattering coefficient can be measured with different combinations of polarisations at the transmitter and receiver. The use of multiple polarisations gives rise to multidimensional image data known as polarimetric radar images. The principles of radar imaging are shown in Figure 2.1.

2.1.1 Frequency Bands

Radars are classified by the frequency band their emitted pulse belongs to. The frequency bands most commonly used by spaceborne earth observation radars are L-band (Seasat, JERS-1 and ALOS PALSAR), C-band (ERS-1, ERS-2, Envisat ASAR, Radarsat-1 and Radarsat-2) and X-band (TerraSAR-X and COSMO-SkyMed). P-band[2] radar has

[1] Radar was originally an acronym for the method of *radio detection and ranging*, but has been assimilated as a standard word in most languages, and now refers to the instrument itself.

[2] The P-band is not defined in the IEEE standard reported in Table 3, but refers to a band which lies partially in the UHF-band and partially in the VHF-band, according to this designation.

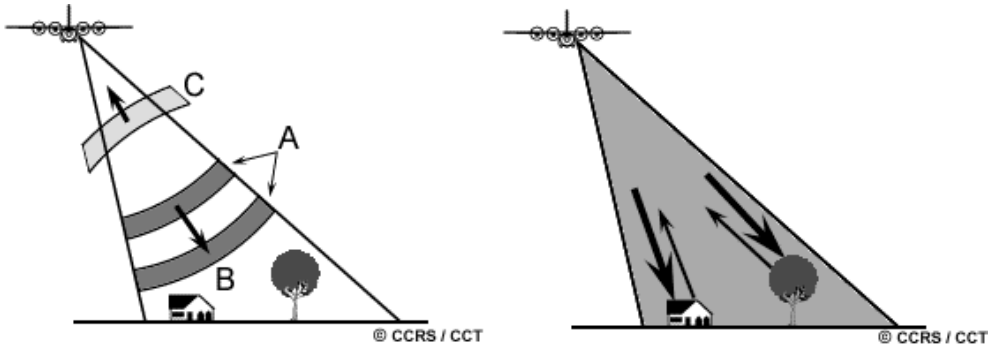

Figure 2.1: An imaging radar is an active microwave instrument which transmits electromagnetic pulses (A), focused by the antenna into a beam (B), and receives a portion of the reflected energy (C) backscattered from various objects.

only been mounted on airborne research missions, but a satellite mission named BIOMASS is planned. The letter designation applied to different frequency ranges is shown in Table 3, whose source is the IEEE Standard 521-2000(R2009) [IEEE, 2009].

The penetration depth of the microwave into a target medium, such as vegetation, glacier, sea ice and soil, increases with wavelength. It also depends on the moisture level of the medium. Thus, the usefulness of the different frequency bands vary with application. For instance, C-band is generally preferred for mapping of sea ice, because it provides the best contrast between sea and ice (although the contrast also depends largely on polarisation and incidence angle). L-band is preferred for studies of many types of vegetation, for which the wave penetrates the vegetation canopy and reaches the ground or surface level. P-band is required to penetrate rain forest, and thus to obtain meaningful estimates of biomass for this biotope, since the measurements saturate at shorter wavelengths, resulting in underestimation. The frequency band also determines the scale of roughness which interacts with the radar wave and influences the measurements. The backscattered energy is sensitive to surface curvature and roughness at length scales near the radar wavelength.

2.1.2 Properties of Imaging Radar

The major advantage of active microwave instruments is that they work independently of sunlight conditions and cloud cover. Unlike optical sensors, they operate equally well nighttime as daytime, and the attenuation of the signal by clouds and water vapour is negligible in most bands. L-band radars may experience disturbance by ionospheric Faraday rotation under certain conditions [Freeman and Saatchi, 2004], while X-band and Ku-band backscatter is sensitive to precipitation, which has been successfully retrieved from spaceborne SAR observations [Marzano and Weinman, 2008].

Radar systems give access to different parameters compared to optical systems. Their

[3]Frequencies from 216-450 MHz are sometimes called P-band.

Table 2.1: Microwave bands with letter designation, frequency range and wavelength range.

Band	Frequency range	wavelength range [cm]
HF	3-30 MHz	10-100 m
VHF[3]	30-300 MHz	1-10 m
UHF[3]	300-1000 MHz	30-100 cm
L	1-2 GHz	15-30 cm
S	2-4 GHz	7.5-15 cm
C	4-8 GHz	3.75-7.5 cm
X	8-12 GHz	2.5-3.75 cm
Ku	12-18 GHz	16.7-25 mm
K	18-27 GHz	11.1-16.7 mm
Ka	27-40 GHz	7.5-11.1 mm
V	40-75 GHz	4.0-7.5 mm
W	75-110 GHz	2.7-4.0 mm
mm	110-170 GHz	1.8-2.7 mm

measurements can be related to surface roughness, humidity and geometrical properties, as previously discussed. They cannot be used to retrieve biophysical parameters that require access to radiances, reflectances and brightness temperatures. Nevertheless, many mapping and classification products can be obtained with similar quality as for optical instruments, and the independence of solar illumination and cloud cover assures temporal consistence of the service. This property is especially attractive for monitoring of high latitude areas, where Arctic/Antarctic winter and frequent cloud cover limits the usefulness of optical sensors. Persistent cloud cover also severely restrict the capacity of optical monitoring of tropical rain forest, whereas radar instruments deliver consistent results, which is a vital requirement for operational services.

2.1.3 Instruments

Different types of radar can be classified as imaging radars. A ground-penetrating radar (GPR) is an active microwave instrument used to image the subsurface, which is often implemented as a continuous wave (CW) radar[4]. It is applied to a variety of media, including rock, soil, ice, snow, fresh water and man-made structures. The data are processed into a radargram which shows the depth profile and indicates boundaries between layers with different dielectric constant. A weather radar is normally a ground-based Doppler radar capable of locating precipitation, calculating its intensity

[4]A continuous wave radar transmits continuous waves instead of pulses. Range measurements are enabled by use of e.g. a frequency modulated (FM-CW) or step-frequency (SF-CW) transmitted wave.

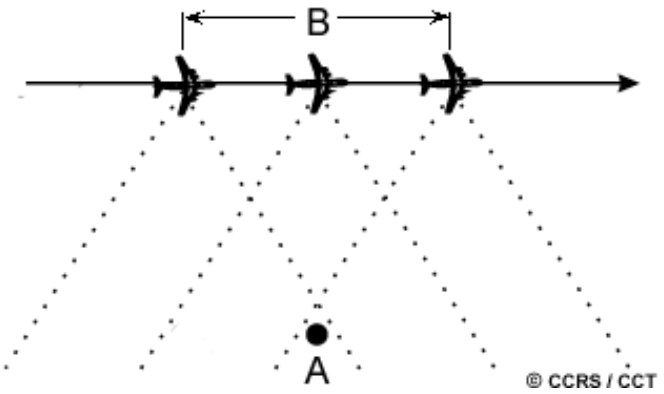

Figure 2.2: The SAR principle: A target (A) is illuminated by several pulses of the radar beam. The backscattered echoes of each pulse is recorded. The length of the synthesized antenna (B) is the the distance between the points where the target enters and leaves the radar beam.

and velocity, and identifying its type. It scans a volume of air around the radar station, and images are produced as different cross-sections of the scanned volume. In addition, we have real aperture radar (RAR) and *synthetic aperture radar* (SAR), whose function is explained in the next section. These can both be ground-based or airborne. SAR data are also widely available from spaceborne instruments that provide regular global coverage. These instruments are therefore important and reliable sources of information for various monitoration programs.

The methods presented in this thesis are aimed at multilook polarimetric data, a data format which is defined in Section 2.2.4. The most obvious subject for the proposed analysis tools is polarimetric SAR data. However, polarimetric GPRs and weather radars exist [Langley et al., 2009, Galetti et al., 2008], and their data can also be processed into the same format, even though other representations are more common. The developments of this thesis are presented as a generic theory for polarimetric radar data, since the theoretical framework can in principle be applied to any kind of multilook polarimetric radar data.

2.1.4 Synthetic Aperture Radar

The synthetic aperture processing technique [Oliver and Quegan, 2004, Cumming and Wong, 2005, Massonnet and Souyris, 2008] must be credited for the availability of high resolution radar images captured by spaceborne platforms. The difference between a RAR and a SAR sensor lies in the image resolution in the azimuth, or along-track direction. The azimuth resolution of a RAR is determined by the beamwidth of the antenna, which is limited by practical constraints. A SAR obtains multiple measurements of the scene at different azimuth angles, and uses advanced signal processing to combine these into an image with improved azimuth resolution. The effect is the same as using

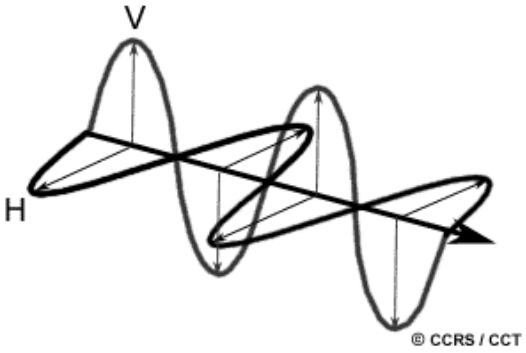

Figure 2.3: Electromagnetic wave with linear polarisation.

an aperture whose size is larger than the actual, hence the term *synthetic aperture*. The finest azimuth resolution achieved by the most recent spaceborne SAR instruments is 1 meter for TerraSAR-X and COSMO-SkyMED and 3 metres for Radarsat-2 and ALOS PALSAR. The SAR principle is explained by Figure 2.2.

2.1.5 Polarimetric Radar

An EM wave consists of electric and magnetic field components that oscillate in phase perpendicular to each other and perpendicular to the direction of energy propagation (see Figure 2.3). The *polarisation* of an EM wave describes the orientation of its oscillations. A fully polarimetric radar simultaneously transmits microwave pulses with two different orthogonal polarisations, it measures the electric field components at two orthogonal polarisations, and resolves the scattering coefficients for all four combinations of transmit and receive polarisation. Any polarisation can be synthesised as a superposition of two orthogonal polarisation. The fully polarimetric measurements thus constitute a complete description of the scattering characteristics of the resolution cell, which can be analysed for an arbitrary polarisation by a simple transformation of the orthogonal basis.

SAR instruments commonly use linear polarisations, where the electrical field is oriented in a single direction, normally horizontally and vertically. Another option is to use circular or elliptical polarisations, where the electric field rotates rightwards or leftwards in the direction of propagation. The polarisations at the transmitter and receiver need not be the same, and can be chosen to optimise system performance or in accordance with given restrictions or requirements.

In *radar polarimetry*, we use the complex scattering coefficients measured at different polarisations to characterise the target. A polarimetric radar extends the capabilities of normal single polarisation radar, not only because it provides multichannel data. Most importantly, it provides a strong link to the physics of the scattering process and allows us to identify distinct scattering mechanisms. Coherent scattering from point sources can be resolved as scattering from elementary geometrical objects, such

as spheres, cylinders, dipoles, diplanes, dihedrals and trihedrals. Incoherent scattering can be resolved as surface (single bounce) scattering, double bounce scattering and volume scattering. The methods used are known as *polarimetric decompositions*, and a rich literature has emerged on the topic [Mott, 2007, Lee and Pottier, 2009, Cloude, 2010].

2.2 Data Formats

This section describes the data formats encountered for polarimetric radar data. We start from a mathematical description of the most elementary measurable in polarimetric radar imaging, the matrix holding the scattering coefficients of all polarimetric channels. We then describe the multilooking process and transformation of the data into the intensity domain.

2.2.1 Single Look Complex Data

As described in the previous section, the fully polarimetric SAR instrument separately transmits orthogonally polarised microwaves pulses, and measures orthogonal components of the received signal. For each pixel, the measurements result in a matrix of scattering coefficients. These are complex-valued, dimensionless numbers that describe the transformation of the transmitted (incoming) EM field to the received (backscattered) EM field for all combinations of transmit and receive polarisation.

The transformation can be expressed as

$$\begin{bmatrix} E_x^r \\ E_y^r \end{bmatrix} = \frac{e^{\jmath k\rho}}{\rho} \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix} \begin{bmatrix} E_x^t \\ E_y^t \end{bmatrix} \tag{2.1}$$

where $\jmath = \sqrt{1}$ is the imaginary unit, k denotes wavenumber and ρ is the distance between radar and target. The subscript of the EM field component E_i^j, where $i \in \{x, y\}$, refers to the polarisation it is associated with. The superscript of E_i^j, where $j \in \{r, t\}$, indicates if it is the transmitted or received field component. The orthogonal polarisations are denoted x and y for generality, although it has been assumed that the same polarisations are used at the transmitter and the receiver. The scattering coefficients $S_{ij},\ i, j \in \{x, y\}$, are subscripted with the associated receive and transmit polarisation, in that order.

The scattering coefficients are complex-valued, and they comprise what is known in radar imaging terminology as a *look*. The SAR processor may split the full synthetic aperture into several subapertures. This is done by splitting the Doppler bandwidth into sub-bands and extracting the band-limited signal. The portion of the SAR signal associated with one subaperture or subband represents an individual look of the scene. The scattering coefficients represent one such look after the image has been focused. Data structures that contain the scattering coefficients are therefore referred to as *single-look complex* (SLC) data. The single-look term is used also for sensors without the synthetic aperture capacity, even though the look extraction aspect is missing.

2.2.2 Scattering Matrix

The choice of polarisations is from now on restricted to the linear pair, that is, the horizontal or vertical polarisation. This is the most commonly used set of orthogonal polarisations for SAR systems, and the restriction can be done without loss of generality for the methods subsequently derived. The exception is for interpretations of scattering mechanisms, that rely explicitly on the choice of polarisation basis.

We extract from (2.1) the *scattering matrix*

$$\mathbf{S} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \in \mathbb{C}^{2\times 2}, \tag{2.2}$$

where the subscripts h and v denote horizontal and vertical polarisation, respectively. This matrix is also known as the *Sinclair matrix*, and holds all scattering coefficients measured by the *fully polarimetric radar*.

2.2.3 Scattering Vector

Lexicographic Basis

The scattering vector is simply the vectorised version of the scattering matrix, defined as

$$\mathbf{s} = \mathrm{vec}(\mathbf{S}^T) = \begin{bmatrix} S_{hh} \\ S_{hv} \\ S_{vh} \\ S_{vv} \end{bmatrix} \in \mathbb{C}^{4\times 1}, \tag{2.3}$$

where $\mathrm{vec}(\cdot)$ is the column stacking vectorisation operator. The vector elements can also be seen as coefficients of the lexicographic decomposition of the scattering matrix:

$$\mathbf{S} = S_{hh}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + S_{hv}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + S_{vh}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + S_{vv}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.4}$$

This vector is therefore known as the *lexicographic basis scattering vector*.

Pauli Basis

Another representation of the scattering vector is obtained by a linear transformation of the lexicographic basis vector. The *Pauli basis scattering vector* is obtained as

$$\mathbf{k} = \mathbf{U}\,\mathbf{s} = \frac{1}{\sqrt{2}}\begin{bmatrix} S_{hh} + S_{vv} \\ S_{hh} - S_{vv} \\ S_{hv} + S_{vh} \\ \jmath(S_{hv} - S_{vh}) \end{bmatrix} \in \mathbb{C}^{4\times 1}, \tag{2.5}$$

where $\mathbf{U}$ is the unitary transformation matrix

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & \jmath & -\jmath & 0 \end{bmatrix} . \tag{2.6}$$

The vector elements are the coefficients in the Pauli decomposition of the scattering matrix, given by

$$\begin{aligned} \mathbf{S} = & \frac{S_{hh} + S_{vv}}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{S_{hh} - S_{vv}}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ & + \frac{S_{hv} + S_{vh}}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \frac{\jmath(S_{hv} - S_{vh})}{\sqrt{2}} \begin{bmatrix} 0 & -\jmath \\ \jmath & 0 \end{bmatrix} . \end{aligned} \tag{2.7}$$

The basis of this decomposition contains the three 2×2 Pauli matrices, that were originally introduced in quantum mechanics to describe the spin of a spin $1/2$ particle in three spatial directions. The fourth basis matrix is the 2×2 identity matrix, which is associated with the first element of $\mathbf{k}$.

The advantage of the Pauli basis scattering vector is that it provides physical interpretations of its elements in terms of elementary scattering mechanisms. The respective elements of $\mathbf{k}$, denoted $\{k_1, \cdots, k_4\}$, can be related to: single or odd-bounce scattering from a plane surface (k_1), diplane scattering (double-bounce or even-bounce) from corners with a relative orientation of $0°$ (k_2) and $45°$ (k_3), and the residue of antisymmetric components (k_4) [Lee and Pottier, 2009, Cloude, 2010].

Reduced Dimension Scattering Vectors

Let d be the dimension of $\mathbf{s}$, which is equivalent to the number of polarimetric channels. It will be referred to as the polarimetric dimension. The polarimetric dimension can be reduced compared to the quadrature polarimetric case of $d = 4$. If only one polarisation is used at the transmitter (or receiver), then only a subset of the scattering coefficients can be measured, and we obtain *dual polarisation* data with $d = 2$. Single polarisation data are obviously also a special case, with $d = 1$.

The case of $d = 3$ is encountered when we assume *reciprocity* of the target, defined as follows: For natural terrain it can safely be assumed that the cross-polarised channels are approximately equal: $S_{hv} \simeq S_{vh}$. These measurements can then be averaged to reduce uncertainty, and we obtain the reduced scattering vector

$$\mathbf{s} = \begin{bmatrix} S_{hh} \\ (S_{hv} + S_{vh})/\sqrt{2} \\ S_{vv} \end{bmatrix} \in \mathbb{C}^{3\times 1} . \tag{2.8}$$

The term $\sqrt{2}$ ensures that the total measured power:

$$\mathrm{P}_{\text{tot}} = |S_{hh}|^2 + |S_{hv}|^2 + |S_{vh}|^2 + |S_{vv}|^2$$

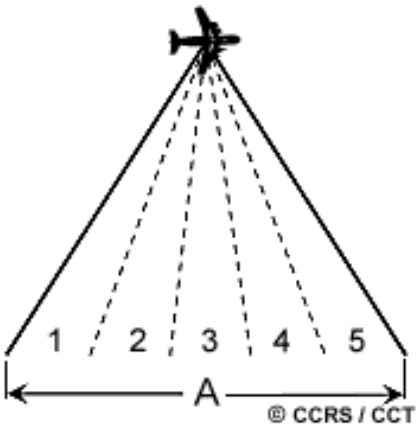

Figure 2.4: In multilook processing, the radar beam (A) is divided into several subbeams, each providing an independent (in an ideal case) look at the illuminated scene.

is maintained regardless of a change of basis. The reciprocity assumption breaks down in urban environments, or generally when the target geometry is such that the amount of cross-polarised power depends on the radar look angle, which is typical for man-made targets and structures with non-random orientation.

The equivalent version of the Pauli basis scattering vector is

$$\mathbf{k} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + S_{vv} \\ S_{hh} - S_{vv} \\ S_{hv} + S_{vh} \end{bmatrix} \in \mathbb{C}^{3\times 1} . \tag{2.9}$$

2.2.4 Multilook Complex Data

The *look* term was defined in Section 2.2.1 as a portion of the SAR signal recorded by a part of the synthetic aperture, known as a subaperture. The subaperture signal can be extracted from the total signal by filtering in the frequency domain. Multiple looks can be summed incoherently to produce a multilook image, an averaging operation known as *multilooking* [Cumming and Wong, 2005,Massonnet and Souyris, 2008]. Multilooking can also be done in the spatial domain, after the image has been focused. This is done by computing the mean value of a group of adjacent pixels, and must also be done incoherently, since the single-look data are complex-valued. The multilook principle is illustrated by Figure 2.4.

Multilooking of single polarisation radar data produces amplitude or intensity data that are real-valued. All phase information is discarded. In the polarimetric case, the multilooking process creates complex data which preserves information on the mean phase difference between the polarimetric channels. The format is known as *multilook complex* (MLC) data.

We remark that the averaged looks are correlated. For multilooking in the frequency domain, the correlation occurs because the filters used to split the Doppler bandwidth

into subbands have a slight overlap. In spatial domain multilooking, we average neighbour pixels that are correlated because they share a certain amount of information from the focusing process, due to the radar point spread function [Rignot and Chellappa, 1993]. This has an impact on the statistical modelling of MLC data. The exact form of the PDF for correlated data does not have a simple closed form expression [Goodman, 1975, Rignot and Chellappa, 1993, Gierull and Sikaneta, 2002]. The practical approach has been to derive the distribution as if the multilook samples were statistically independent, and replace the *nominal number of looks*, equivalent to the actual number of samples, with an *equivalent number of looks* (ENL). The ENL is must be estimated from the data, which is done by equating certain empirical sample moments with the corresponding theoretical population moments under the assumed statistical distribution model. This is the topic of Paper 1 in Chapter 5.

2.2.5 Covariance Matrix

Assume that L looks are available, in the form of the scattering vector sample $\{\mathbf{s}_\ell\}_{\ell=1}^{L}$. We refer to L as the nominal number of looks. The lexicographic basis scattering vector is multilooked by computing its sample covariance matrix, under the assumption that the $\{\mathbf{s}_\ell\}$ are zero mean, a condition discussed in Section 2.3.1. The multilooking operation is formulated as

$$\mathbf{C} = \frac{1}{L}\sum_{\ell=1}^{L}\mathbf{s}_\ell\mathbf{s}_\ell^H = \begin{bmatrix} \langle S_{hh}S_{hh}^*\rangle & \langle S_{hh}S_{hv}^*\rangle & \langle S_{hh}S_{vh}^*\rangle & \langle S_{hh}S_{vv}^*\rangle \\ \langle S_{hv}S_{hh}^*\rangle & \langle S_{hv}S_{hv}^*\rangle & \langle S_{hv}S_{vh}^*\rangle & \langle S_{hv}S_{vv}^*\rangle \\ \langle S_{vh}S_{hh}^*\rangle & \langle S_{vh}S_{hv}^*\rangle & \langle S_{vh}S_{vh}^*\rangle & \langle S_{vh}S_{vv}^*\rangle \\ \langle S_{vv}S_{hh}^*\rangle & \langle S_{vv}S_{hv}^*\rangle & \langle S_{vv}S_{vh}^*\rangle & \langle S_{vv}S_{vv}^*\rangle \end{bmatrix}, \tag{2.10}$$

where $(\cdot)^*$ and $(\cdot)^H$ are the complex conjugation and Hermitian transposition operators, respectively, and $\langle\cdot\rangle$ denotes a sample mean over all single-look measurements. This produces the *polarimetric covariance matrix* $\mathbf{C}$, which is positive definite and Hermitian symmetric. It has the real-valued intensities of the polarimetric channels on the main diagonal, and their complex covariances off the diagonal. The matrix $\mathbf{C}$ is the measurable when working with multilook polarimetric radar data, and it is the subject of the subsequent analysis.

2.2.6 Coherency Matrix

The covariance matrix $\mathbf{C}$ can be replaced by the *polarimetric coherency matrix* $\mathbf{\Omega}$, which is the term used for the sample covariance matrix of the Pauli basis scattering vector. It is computed from the sample $\{\mathbf{k}_\ell\}_\ell^L$ by

$$\mathbf{\Omega} = \frac{1}{L}\sum_{\ell=1}^{L}\mathbf{k}_\ell\mathbf{k}_\ell^H = \mathbf{UCU}\,. \tag{2.11}$$

The coherency matrix is often the preferred representation, because of its the physical interpretation. In the quadrature polarimetric case, the first three elements on the diagonal are the intensities:

$$[\mathbf{\Omega}]_{11} = \langle |S_{hh} + S_{vv}|^2 \rangle \,, \tag{2.12}$$
$$[\mathbf{\Omega}]_{22} = \langle |S_{hh} - S_{vv}|^2 \rangle \,, \tag{2.13}$$
$$[\mathbf{\Omega}]_{33} = \langle |S_{hv} + S_{vh}|^2 \rangle \,, \tag{2.14}$$

that can be interpreted in terms of the same elementary scattering mechanisms as the elements of $\mathbf{k}$. Incoherent polarimetric decompositions, such as the Freeman decomposition [Freeman and Durden, 1998] and Yamaguchi decomposition [Yamaguchi et al., 2005], go even further in extracting entities that can be directly related to surface scattering, double-bounce scattering and volume scattering and their relative proportion.

For the methods developed in this thesis, it is of no concern whether we use the covariance matrix or the coherency matrix. The results are equally valid and useful for both data formats.

2.3 Polarimetric Radar Statistics

In this section we review statistical models for the polarimetric radar measurements. We specifically look at probability density functions (PDFs) for the different data formats based on a physical description of the scattering process. The EM field measured at the radar is a superposition in the far field of coherent microwave components, each produced by a reflection from a unique surface element called a *scatterer*. If the resolution cell contains only a limited number of scatterers, whose position could be accurately determined, then a deterministic description is in theory possible. In practice, we must

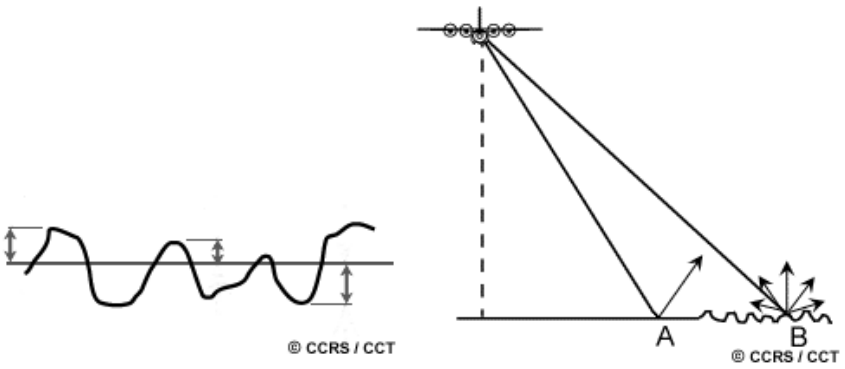

Figure 2.5: The roughness of the surface (left) determines the type of scattering produced (right). A smooth surface (A) causes specular reflection, while a rough surface scatters diffusely (B). The scattering can also be characterised as coherent (A) and incoherent (B).

resort to a stochastic model to describe the EM field.

A *rough surface* is a surface with height variation on a scale larger than the wavelength used to illuminate it [Delignon et al., 2001, Lee, 2005][5]. In a radar remote sensing context, many natural surfaces can be considered as rough. The scatterers vary in number and geometry according to the surface type and the spatial resolution of the observation, but the number is generally assumed to be high. The nature of the scattering depends on the relative strength of the reflections from the scatterers and the distribution of the phase shift they induce. The next sections establish a mathematical description of the scattering process and definitions that characterise the scattering as *strong* or *weak*. Surface roughness and different types of scattering is shown in Figure 2.5.

At this point, it is pertinent to define some other terms. In physics, *coherence* is a property of two or more waves that are in phase both temporally and spatially. More generally, it describes the correlation between all physical quantities of the wave [Born and Wolf, 1999, Glickman, 2000]. This is the property which enables stationary interference. *Incoherent scattering* is defined as the scattering produced when an incident wave encounters scatterers that cause the scattered EM field to exhibit random variations in phase and amplitude due to lack of coherence. *Coherent scattering*, on the other hand, produces a deterministic scattered EM field. This happens when the incident wave is scattered by a fixed point target or a distributed target with scatterers whose relative position is fixed [Glickman, 2000].

2.3.1 Random Walk Model of Scattering

The scattering process is often described by a random walk model [Goodman, 2007, Lopès et al., 2008]. Let the EM field measured at the sensor be a sum of the field components reflected by N scatterers in the resolution cell. The scattering coefficient S of a general polarimetric channel thus represents the total scattering experienced by the microwave after interaction with N scatterers. This can be written as the sum

$$S = \sum_{k=1}^{N} S^{(k)} = \sum_{k=1}^{N} A^{(k)} e^{\jmath\theta^{(k)}} \,. \tag{2.15}$$

Here $S^{(k)}$ is the scattering coefficient associated with the kth scatterer, whose polar decomposition yields the amplitude component $A^{(k)} = |S^{(k)}|$ and phase component $\theta^{(k)} = \angle S^{(k)}$. These represent the attenuation and phase shift imposed on the incident wave by the kth scatterer. We initially make three fundamental assumptions:

1. The amplitudes $\{A^{(k)}\}_{k=1}^{N}$ are independent and identically distributed (IID) random variables. So are the phases $\{\theta^{(k)}\}_{k=1}^{N}$.

[5]Surface roughness can be measured by the product of the wavenumber k and the root mean square height h_{rms} of the roughness. Moderate roughness starts at $kh_{\text{rms}} > 1$ and high roughness at $kh_{\text{rms}} > 5$.

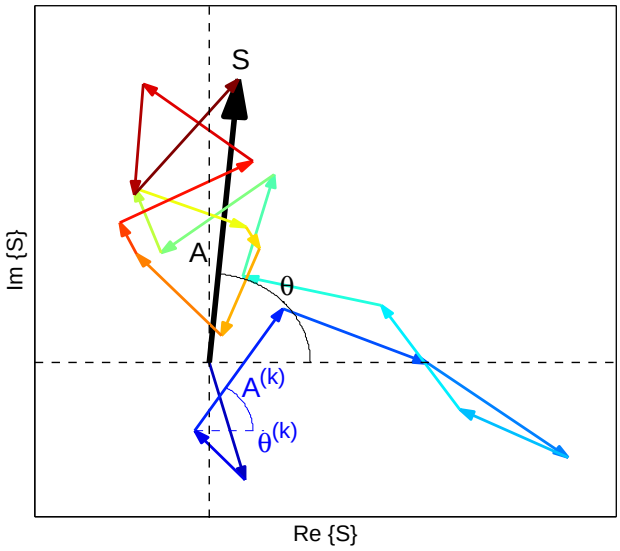

Figure 2.6: Random walk model of strong scattering. The scattering coefficient S with amplitude A and phase θ is shown as a phasor in the complex plane (black arrow). It is decomposed as the coherent sum of the scattering coefficients $S^{(k)}$ corresponding to N individual scatterers with amplitude $A^{(k)}$ and phase $\theta^{(k)}$. The accumulated phasor sum is indicated by the coloured arrows.

2. The $\{A^{(k)}\}$ and the $\{\theta^{(k)}\}$ are independent.

3. The $\{\theta^{(k)}\}$ are uniformly distributed over all angles.

Figure 2.6 illustrates equation (2.15) under the three assumptions listed above. It pictures the coherent summation as a two-dimensional random walk in the complex plane, with the in-phase component along the first axis and the quadrature phase along the second. We shall use the figure as a starting point for a discussion of different scattering regimes.

Strong Scattering

Figure 2.6 shows an example of a random walk with $N = 20$ steps, each representing the reflection by an individual scatterer. The coloured arrows show the accumulated coherent sum, which ends up as the total scattering coefficient S, shown as the thick black arrow annotated with amplitude A and phase θ. The function of the colour coding is just to show the progress of the vector summation of the scattering coefficient components $S^{(k)}$ with amplitude $A^{(k)}$ and phase $\theta^{(k)}$.

We have assumed that the phase components are uniformly distributed over all angles, denoted as: $\theta^{(k)} \sim U[0, 2\pi]$. This implies that the angle of the coherent sum is also uniform over the same interval: $\theta \sim U[0, 2\pi]$. This condition defines the *strong scattering* regime [Barakat, 1986, Jakeman and Tough, 1987]. It means that the random walk has

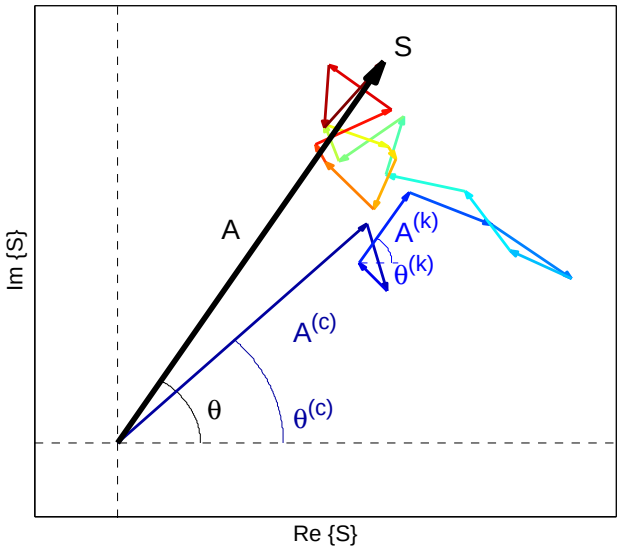

Figure 2.7: Random walk model of weak scattering. The scattering coefficient S with amplitude A and phase θ is shown as a phasor in the complex plane (black arrow). It is decomposed as the sum of a dominating coherent term S_c, with amplitude $A^{(c)}$ and phase $\theta^{(c)}$, and an incoherent term equal to the sum of random phasors described in Figure 2.6, contributed by N distributed scatterers.

an equal chance of ending up in any angular sector of the complex plane, and moreover that the coherent sum is a random variable with zero mean, even though the amplitude is positive. The zero mean of S is sometimes expressed as vanishing of the coherent, or specular, component [Ruffing and Fleischer, 1985]. This is an equivalent condition for strong scattering. A scattering process with the zero mean property produces *fully developed speckle*, a term which is explained in more detail in Section 2.3.4.

Weak Scattering

A non-uniform distribution of the phase components, written as $\theta^{(k)} \nsim U[0, 2\pi]$, results in a non-uniform distribution of θ and a non-zero mean of S, which thus has a non-vanishing incoherent or specular component. These are the characteristics of a *weak scattering* process, which produces a *partially developed speckle* pattern [Ruffing and Fleischer, 1985, Barakat, 1986]. The weak scattering regime is illustrated by Figure 2.7. Compared to Figure 2.6 and Equation (2.15), a dominant scatterer has been added, meaning a scatterer whose reflection is much stronger than the others. The random walk model has simply been translated in the complex plane by the first vector component, which represents the dominant scatterer.

The described weak scattering process can be expressed as

$$S = S^{(c)} + S^{(i)} = A^{(c)} e^{\jmath\theta^{(c)}} + \sum_{k=1}^{N} A^{(k)} e^{\jmath\theta^{(k)}} = A e^{\jmath\theta}\,, \tag{2.16}$$

where S is a sum of a coherent term, $S^{(c)}$, and an incoherent term, $S^{(i)}$. The coherent term has deterministic amplitude $A^{(c)}$ and phase $\theta^{(c)}$, that can be modelled with degenerate distributions[6]. The incoherent term represents the stochastic model of strong scattering from (2.15), and contains the same random amplitude and phase components. We thus have a set of amplitude components, $\mathcal{A} = \{A^{(k)}, A^{(1)}, \cdots, A^{(N)}\}$, and a set of phase components, $\Theta = \{\theta^{(c)}, \theta^{(1)}, \cdots, \theta^{(N)}\}$. The distribution of Θ becomes non-uniform, since $\theta^{(c)}$ is a deterministic component. In the described case, the IID assumptions are also violated. However, weak scattering can occur without the presence of dominant scatterers, if only the phase component distribution is non-uniform. See for instance [Jakeman and Tough, 1987], where the $\{\theta^{(k)}\}$ are assumed to follow a von Mises distribution.

2.3.2 Gaussian Model

Single-look Complex Data

The simplest statistical model for the radar return is based on the random walk model in (2.15) under all three assumptions listed in Section 2.3.1. Further assume that the number of scatterers, N, is large and constant. It follows from the central limit theorem that the scattering coefficient S converges in distribution to a complex Gaussian random variable with zero mean and variance σ^2 as $N \to \infty$, where σ^2 is the mean radar cross section of the scattering medium. The PDF of S is given by

$$f_S(S; \sigma^2) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{S^2}{\sigma^2}\right) . \tag{2.17}$$

In the polarimetric case, the zero mean scattering vector $\mathbf{s}$ follows a multivariate circular complex Gaussian distribution with covariance matrix $\mathbf{\Sigma} = \mathrm{E}\{\mathbf{s}\mathbf{s}^H\}$. The covariance matrix contains the complex cross-correlation between scattering coefficients at different polarimetric channels. Let S_i and S_k be the scattering coefficients of two polarimetric channels. They are decomposed into their real and complex parts by: $S_i = \mathcal{R}\mathrm{e}(S_i) + \jmath\,\mathcal{I}\mathrm{m}(S_i)$. The *circularity* property of complex Gaussian distribution is defined by the correlations [van den Bos, 1994, Goodman, 2007]

$$\mathrm{E}\left\{\mathcal{R}\mathrm{e}(S_i)\,\mathcal{R}\mathrm{e}(S_k)\right\} = \mathrm{E}\left\{\mathcal{I}\mathrm{m}(S_i)\,\mathcal{I}\mathrm{m}(S_k)\right\} , \tag{2.18}$$

$$\mathrm{E}\left\{\mathcal{R}\mathrm{e}(S_i)\,\mathcal{I}\mathrm{m}(S_k)\right\} = -\,\mathrm{E}\left\{\mathcal{R}\mathrm{e}(S_k)\,\mathcal{I}\mathrm{m}(S_i)\right\} , \tag{2.19}$$

that follow from the assumptions on the $\{A_k\}$ and the $\{\theta_k\}$ in the random walk model. The cross-correlations also imply that

$$\mathrm{E}\{S_i S_k\} = 0\,. \tag{2.20}$$

The PDF of $\mathbf{s}$ is [van den Bos, 1994]

$$f_{\mathbf{s}}(\mathbf{s}; \mathbf{\Sigma}) = \frac{1}{\pi^d |\mathbf{\Sigma}|} \exp(-\mathbf{s}^H \mathbf{\Sigma}^{-1} \mathbf{s})\,. \tag{2.21}$$

[6] A deterministic value x_0 can be modelled as a random variable with the Dirac delta function, $\delta(x - x_0)$, as its PDF. This is referred to as a *degenerate* distribution.

The derivations for the lexicographic basis and Pauli basis cases are analogous. The following presentation is therefore restricted to results for the former, without loss of generality.

We see that the random walk model of strong scattering leads to complex Gaussian distributions for SLC data. We will now present distributions for MLC data, but note that the term *Gaussian model* is often used somewhat imprecisely about other data formats, whose distributions are derived directly from the complex Gaussian distributions of the SLC data by appropriate transformations of the random variates. Thus, a Rayleigh distributed amplitude: $A = |S^2|$, an exponentially distributed single-look intensity: $I = A^2$, and a gamma distributed multilook intensity: $I_L = \langle I \rangle = \langle A^2 \rangle$ are considered to belong to the Gaussian model. The derivation of all these distributions are shown in [Oliver and Quegan, 2004]. In line with this usage of terminology, the scaled complex Wishart distribution, presented in the next section, can be considered as the multilook polarimetric extension of the Gaussian model [Doulgeris et al., 2008,Vasile et al., 2009].

Multilook Complex Data

Assume that we have a multilook sample of L independent scattering vectors that are multivariate circular complex Gaussian with dimension d, and that $L \geq d$. The sample covariance matrix $\mathbf{C}$ computed from this sample will then be non-singular, and belongs to the cone of positive definite complex Hermitian matrices, denoted as Ω_+. It follows what we refer to as the *scaled complex Wishart distribution*. The PDF of $\mathbf{C}$ under the Gaussian model is given by

$$f_{\mathbf{C}}(\mathbf{C}; L, \boldsymbol{\Sigma}) = \frac{L^{Ld}}{\Gamma_d(L)} \frac{|\mathbf{C}|^{L-d}}{|\boldsymbol{\Sigma}|^L} \exp(\mathrm{tr}(-L\boldsymbol{\Sigma}^{-1}\mathbf{C})) \,, \tag{2.22}$$

which is defined on Ω_+, where $|\cdot|$ is the determinant operator, $\mathrm{tr}(\cdot)$ is the trace operator and $\Gamma_d(L)$ is the multivariate gamma function of the complex kind, defined in (4.3). The true complex Wishart distribution [Goodman, 1963] describes the matrix variable $\mathbf{W} = L\mathbf{C}$, hence (2.22) follows by a linear transformation.

The distribution parameters are the shape parameter L and the scale matrix $\boldsymbol{\Sigma} = \mathrm{E}\{\mathbf{C}\}$. In radar statistics, L is recognised as the number of looks, with reference to the multilooking process. In other contexts, it is often referred to as the degrees of freedom. The equivalent number of looks, discussed in 2.2.4, should be inserted for L to account for correlation between the looks. The substituted parameter is still subject to the condition of non-singularity.

2.3.3 Non-Gaussian Model

It has been shown experimentally that the Gaussian model, with its bundle of equivalent distributions for different data formats, presents a good fit to real radar image data

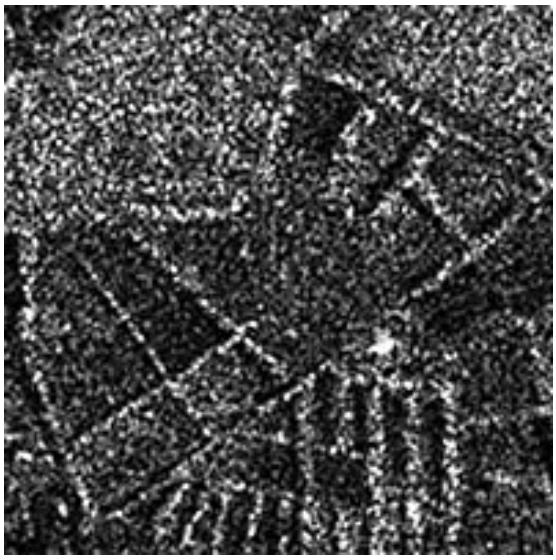

Figure 2.8: Example of the appearance of the speckle pattern in a radar image.

when the scene is homogeneous, with low to moderate roughness and a high number of scatterers. Nevertheless, there is abundant empirical evidence that real data deviate from the model too, especially for images of urban environment, but also for natural terrain, such as rough sea and forest in general (See e.g. [Jakeman and Pusey, 1976, Oliver, 2000, Tison et al., 2004]). This is generally explained by the notion of *texture*, thought of as variations in the mean radar reflectivity between pixels with the same thematic content, which is not accounted for in the Gaussian model. The texture term is defined and discussed in more detail in Section 2.3.5. The Gaussian model only encompasses statistical variation attributed to *speckle*, the interference pattern produced by the coherent sum in the random walk model, which is the topic of the next section.

Several distributions have been proposed for single polarisation amplitude and intensity, that imply non-Gaussian statistics for the scattering coefficient. The Weibull distribution and the log-normal distribution are two of the most popular examples [Oliver and Quegan, 2004]. Even though neither of them bear links to physical modelling of the scattering process, they have been shown to provide reasonably good fit to real data covering rough surfaces. However, they have not yet been extended to matrix-variate statistics and cannot be used to model the polarimetric covariance matrix. This has only been achieved for distributions derived from the doubly stochastic product model, which models the contribution of speckle and texture as independent random variates. The multilook polarimetric version of this model is presented in Section 2.3.6.

2.3.4 Speckle

Speckle is often referred to as noise, but this strictly is a misnomer. Although the appearance of the speckle patterns is granular, noise-like, and strongly limits the interpretabil-

ity of the radar image, the interference causing it is an inherent property of the measurement process that is common to all kind of coherent imaging. The same effect is found in laser images, and much of the fundamental research utilised in statistical analysis of radar images is taken from the field of laser optics (see e.g. [Goodman, 1975, Jakeman and Pusey, 1976]). Speckle is also an artifact of B-mode ultrasound sonography, where a linear array of acoustic transducers simultaneously scans a plane through an object that can be viewed as a two-dimensional image (see e.g. [Wagner et al., 1987, Eltoft, 2006]). In all these cases, the speckle pattern occurs because the measured field results from an incoherent scattering process. An example of a radar speckle is shown in Figure 2.8.

Radar measurements are commonly described by a multiplicative signal model[7]. The first factor in the product is the mean radar reflectivity, defined as the mean fraction of electromagnetic energy reflected by the scattering medium. The other factor is a random variable which models the variation due to speckle. The main characteristic of the product model is that the statistical variation is proportional to the reflectivity. This has a severely distorting effect on the image analysis, considering that the information we normally want to infer is related to the mean reflectivity. In the *doubly stochastic product model*, we also treat the mean reflectivity as a random variable, to obtain a more realistic description of the radar measurements.

The distinction between fully developed and partially developed speckle was made in Section 2.3.1. Fully developed speckle is described by the Gaussian model. Partially developed speckle is modelled by the Rician distribution for the single polarisation amplitude [Goodman, 2000]. The doubly stochastic product model for polarimetric SLC data has been extended to allow for a coherent component in [Eltoft et al., 2006], but has yet to be amended for MLC data, at least when it comes to deriving a closed form expression for the PDF.

2.3.5 Texture

In the field of image processing, the *texture* term is commonly used when referring to the deterministic or stochastic structure of an image region, characterised in terms of the spatial arrangement and directional alignment of the pixel intensities (in gray-scale or colour). As noted in [Tuceryan and Jain, 1994], «we recognize texture when we see it but it is very difficult to define». They list a number of proposed universal definitions, but it may prove more productive to search for a description of texture which is appropriate for the application at hand. In radar imaging, for instance, the texture term has indeed received it own meaning.

When we talk about texture in the context of radar images and the distributions of their data, it normally refers to variations in the mean radar reflectivity, as discussed in Section 2.3.4. We can model the mean reflectivity as a random variable, which is then referred to as a *texture variable*. The texture variable may also absorb other types

[7] According to commonly used terminology, which was argued against in the above, the radar images are subject to multiplicative noise.

of variation, such as the effect of inhomogeneity or mixed targets within a resolution cell. For instance, urban areas exhibit large variations that are badly described by the Gaussian model. This variation can be attributed to the mixture of different scattering media, just as much as the spatial variation of reflectivity within a single medium.

In the doubly stochastic product model, we assign a distribution model to the texture variable and estimate the parameters of the distribution. Textural features, defined in a multitude of ways, contain valuable information for delineation and recognition of image segments and classes in many pattern recognition problems. Classical texture features have also been used in analysis of radar images [Clausi and Yue, 2004, De Grandi et al., 2009]. A less travelled path is to exploit the parameters of the texture distributions, that contain radar specific textural information, as features in radar image analysis tasks, such as segmentation and classification [Oliver, 2000].

The distinction between *texture modulated speckle* and *pure speckle* is finally introduced, where the latter refers to the variation pattern created by the interference phenomenon alone. By these terms, we establish terminology to express clearly whether texture is affecting the speckle pattern and has been included in the signal model.

2.3.6 Multilook Polarimetric Product Model

Let $\widetilde{\mathbf{W}}$ be a scaled complex Wishart distributed random matrix, written as $\widetilde{\mathbf{W}} \sim s\mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma})$. The doubly stochastic product model for multilook polarimetric data is given by

$$\mathbf{C} = T \cdot \widetilde{\mathbf{W}} \,. \tag{2.23}$$

The variation of the polarimetric covariance matrix $\mathbf{C}$ is decomposed into the contribution of texture and fully developed speckle, where the scalar random variable $T \in \mathbb{R}^+$ represents texture and the random matrix $\widetilde{\mathbf{W}} \in \Omega_+$ represents speckle. The texture variable must be strictly positive, and is normally assumed to have unit mean, hence all scale information is put into $\widetilde{\mathbf{W}}$ (or $|\widetilde{\mathbf{W}}|$). This is the convention used in this thesis. The alternative solution is to normalise $\widetilde{\mathbf{W}}$, as it is done in [Doulgeris and Eltoft, 2010]. This yields a matrix variable which contains only information about polarimetry, in form of covariance structure between the channels, while all scale information is put into the texture variable.

By using a scalar random texture variable, it is implicitly assumed that the textural variation is the same in all polarimetric channels. This is not necessarily supported by real data, and is a limitation of the model. It has been suggested that one should at least use distinct texture variables for co-polarised and cross-polarised channels. An extended doubly stochastic product model with multichannel texture has been proposed in [Yu, 1998, Zou et al., 2000] on the form

$$\mathbf{C} = \mathbf{T} \cdot \widetilde{\mathbf{W}} \,, \tag{2.24}$$

with the random diagonal multi-texture matrix defined as

$$\mathbf{T} = \mathrm{diag}(T_1, T_2 \cdots, T_d) = \begin{bmatrix} T_1 & 0 & \cdots & 0 \\ 0 & T_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & T_d \end{bmatrix} . \tag{2.25}$$

Another objection to the doubly stochastic product model is that it is not adequate for co-polarised channels. Analysis of real data has shown that there is an additive term superimposed on the product of texture and fully developed speckle [Séry and Lopès, 1997, López-Martínez and Fàbregas, 2005]. The extensions to multi-texture and a mixed multiplicative additive signal model are not considered in this thesis, but left as topics of future research.

The distribution of $\mathbf{C}$ depends on the chosen distribution for T. For a general texture distribution $f_T(T; \boldsymbol{\theta})$, whose parameters are stored in the texture parameter vector $\boldsymbol{\theta}$, the PDF is given by

$$f_{\mathbf{C}}(\mathbf{C}; L, \boldsymbol{\Sigma}, \boldsymbol{\theta}) = \int_0^\infty f_{\mathbf{C}|T}(\mathbf{C}|T; L, \boldsymbol{\Sigma}) f_T(T; \boldsymbol{\theta})\, dT\,. \tag{2.26}$$

It will be seen that this integral is a matrix-variate Mellin convolution, as defined in Chapter 3. The distribution of $\mathbf{C}|T$ is recognised as the scaled complex Wishart distribution from (2.22). Closed form expressions of $f_{\mathbf{C}}(\mathbf{C}; L, \boldsymbol{\Sigma}, \boldsymbol{\theta})$ has been obtained for a limited number of texture distributions. For gamma distributed texture, the matrix-variate $\mathcal{K}$ distribution was derived in [Lee et al., 1994]. This distribution has shown the same merits as the $\mathcal{K}$ distributions for single polarisation amplitude and intensity (See [Oliver and Quegan, 2004]). It is versatile and provides good fit for natural surfaces with a certain degree of heterogeneity, such as rough sea and forest. For inverse gamma distributed texture, the matrix-variate $\mathcal{G}^0$ distribution was derived in [Freitas et al., 2005]. It has its strength for extremely heterogeneous surfaces, such as urban areas. Both the mentioned distributions are special cases of the matrix-variate $\mathcal{U}$ distribution, derived in [Bombrun and Beaulieu, 2008] from Fisher-Snedecor distributed texture. Because it has two texture parameters, one more than the $\mathcal{K}$ and $\mathcal{G}^0$ distributions, it covers many types of surfaces between the heterogeneous and the extremely heterogeneous case. The added flexibility in terms of modelling capability comes at the cost of more complicated and less accurate parameter estimation, as demonstrated in the paper of Chapter 6.

The mathematical details of the described distributions for the polarimetric covariance matrix is given in Chapter 4. For each distribution the PDF is presented, together with the matrix-variate Mellin kind statistics, that are defined in the next chapter.

Chapter 3

The Mellin Transform

This chapter gives an historical overview which motivates the application of the Mellin transform to statistical analysis of radar data distributions. The univariate and matrix-variate Mellin transform are defined, and the fundamental properties of the latter derived. Some of the main contributions of the thesis are revealed as the framework of matrix-variate Mellin kind statistics is presented.

3.1 An Historical Note

The Mellin transform first appeared in a memoir by the German mathematician Bernard Riemann (1826-1866) about the zeta function, but it is named after the Finnish mathematician *Robert Hjalmar Mellin* (1854-1933), who gave a systematic treatment of the transform and its inverse [Butzer and Jansche, 1997]. In contrast to the Fourier transform and the Laplace transforms, it arose in a mathematical context. Since then, it has also found numerous applications in many areas of physics, statistics and engineering, as reviewed in [Bertrand et al., 2000].

The Mellin transform has been applied to a wealth of problems in analytic combinatorics and analysis of algorithms [Flajolet et al., 1995, Szpankowski, 2001]. It is a key component in systematic methods for evaluation of integrals [Marichev, 1982, Fikioris, 2006, Fikioris, 2007]. In signal processing, it has been applied to time-frequency analysis on logarithmic scale, and its scale invariance property has been utilised to construct affine transformations [Bertrand et al., 1990, Ovarlez et al., 1992, Cohen, 1993, Ruth and Gilbert, 1994, Nelson, 1995, Kaiser, 1996].

The application of the Mellin transform to analysis of matrix-variate radar data builds on results in probability and the theory of statistical distributions. Benjamin Epstein was first to note that the Mellin transform is the «natural analytical tool to use in studying the distribution of products and quotients of independent random variables» [Epstein, 1948], after its utility in analysis of multivariate problems was indicated in [Nair, 1939]. This spurred a series of papers deriving distributions of product and quotients in one or more variables [Dolan, 1964, Springer and Thompson, 1966, Lom-

nicki, 1967, Springer and Thompson, 1970]. Products of dependent random variables were covered in [Subrahmaniam, 1970]. A good overview of the early literature is given in [Cook, 1981].

It will be seen in the following that the Mellin transform is related to the Fourier transform and the bilateral Laplace transform applied to logarithmically transformed data. It is an efficient tool to derive logarithmic moments and cumulants for products and quotients of random variables, and their matrix-variate generalisations, from which we obtain estimators for the distribution parameters with excellent statistical properties. Such logarithmic statistics and parameter estimators have been derived earlier in the univariate case, without explicit reference to the Mellin transform [Stacy, 1962, Stacy and Mihram, 1965, Hoekman, 1991, Kreithen and Hogan, 1991, Blacknell, 1994]. It was *Jean-Marie Nicolas* who developed these ideas into a systematic theory on logarithmic statistics, characterisation of data radar distributions, and estimation of their parameters, with the Mellin transform as the cornerstone [Nicolas, 2002, Nicolas, 2006]. The framework offered by Nicolas has triggered much research activity and many new applications to analysis of radar data. An extensive reference list is given in Paper 2 of Chapter 6.

3.2 Mellin Kind Statistics

Nicolas developed a framework for analysis of random variables defined on $\mathbb{R}^+$, which he called *second kind statistics*. The results presented in this thesis extend the framework to random matrices defined on Ω_+, and renames it as *Mellin kind statistics*, which is the term consistently used from this point. Another appropriate name used by some authors is log-statistics.

The starting point for Nicolas is the definition of the *Mellin kind characteristic function*. His ingenious trick is to replace the Fourier transform in the definition of the conventional characteristic function with the Mellin transform. The *Mellin kind moments* retrieved from the new characteristic function are seen to be moments computed on logarithmic scale. The *Mellin type cumulant generating function* is then defined as the logarithm of the Mellin type characteristic function. From this function we can retrieve *Mellin type cumulants*, that have some very appealing properties. When the target of analysis can be modelled as a product of random variates, the Mellin type cumulant will separate the contribution of the factors [Nicolas, 2002, Nicolas, 2006].

Analysis of radar data on logarithmic scale is not a new and revolutionary concept. Radar images have long been visualised on a logarithmic decibel scale to make more efficient use of the dynamic range of pixel intensities and improve the contrast. The homomorphic speckle filter uses a logarithmic transformation to transform the product model into an additive model, which is easier to handle with traditional signal processing methods [Franceschetti et al., 1995, Solbø and Eltoft, 2004]. The separability induced by the Mellin transform therefore comes as no great surprise. The ease of the mathematical derivations, and the simple expressions obtained for the Mellin kind statistics

of common radar distributions (that is, their characteristic function, moments, cumulant generating function and cumulants) is a more striking result. They underline the appropriateness of the approach and resonate with the words of Epstein: The Mellin transform is truly a natural tool for analysing products of random variables, and thus for multilook radar image data under the product model.

The mathematical details of Nicolas' univariate theory are omitted here. A thorough review is given in the paper of Chapter 6, which stresses the analogy between the derivations for the univariate case describing multilook single polarisation data and the matrix-variate case representing multilook polarimetric data. The paper also repeats the traditional *Fourier kind statistics* and highlights their intrinsic link with the additive signal model. Here is only given a repetition of the univariate Mellin transform, before we proceed with the definition of the matrix-variate Mellin kind statistics.

3.3 Univariate Mellin Transform

Let $g(x)$ be a real-valued function of the variable x. The univariate Mellin transform of $g(x)$ is given by

$$\mathcal{M}\{g(x)\}(s) = \int_0^\infty x^{s-1} g(x)\, dx = \mathcal{G}(s)\,, \tag{3.1}$$

where $s \in \mathbb{C}$ is the complex transform variable. Further let the Fourier transform be defined as

$$\mathcal{F}\{g(x)\}(\xi) = \int_{-\infty}^{+\infty} e^{-2\pi \jmath x \xi} g(x)\, dx \tag{3.2}$$

and the bilateral or two-sided Laplace transform as

$$\mathcal{L}_{\mathrm{B}}\{g(x)\}(s) = \int_{-\infty}^{+\infty} e^{-sx} g(x)\, dx\,. \tag{3.3}$$

The relations between the transforms are shown to be [Bertrand et al., 2000]

$$\mathcal{M}\{g(x)\}(\sigma + \jmath 2\pi\xi) = \mathcal{F}\{e^{-\sigma x} g(e^{-x})\}(\xi)\,, \tag{3.4}$$

$$\mathcal{M}\{g(x)\}(s) = \mathcal{L}_{\mathrm{B}}\{g\left(e^{-x}\right)\}(s) \tag{3.5}$$

by virtue of the substitution of $s = \sigma + \jmath 2\pi\xi$. The inverse relations are

$$\mathcal{F}\{g(x)\}(\xi) = \mathcal{M}\{x^{-\sigma} g(e^{-\ln y})\}(s)\,, \tag{3.6}$$

$$\mathcal{L}_{\mathrm{B}}\{g(x)\}(s) = \mathcal{M}\{g(-\ln x)\}(s)\,. \tag{3.7}$$

From Lebegue's dominated convergence theorem [Bartle, 1995], we know that the ordinary unilateral Laplace transform converges absolutely in the semi-plane $\mathcal{R}\mathrm{e}(s) > b$ for some constant b, and possibly at points on the line $\mathcal{R}\mathrm{e}(s) = b$. The intersection of the regions of convergence for the left-sided and the right-sided Laplace transform form

the region of converge for the bilateral Laplace transform. This becomes the region of holomorphy, also known as the analytic strip, and takes the form $b < \mathcal{R}\mathrm{e}(s) < c$, for some real constants b and c, possibly including the boundary lines. When transforming a PDF defined on $\mathbb{R}^+$, the unilateral (right-sided) and bilateral Laplace transforms are identical, and the analytic strip becomes the semi-plane $\mathcal{R}\mathrm{e}(s) > b$.

It can be seen directly from (3.5) that the analytic strip of the Mellin transform is the same as for the bilateral Laplace transform. The boundaries depends on the transformed function, and the strip may even extend to the whole complex plane. The inverse Mellin transform is defined by [Bertrand et al., 2000]

$$\mathcal{M}^{-1}\{\mathcal{G}(s)\} = \frac{1}{2\pi\jmath}\int_{c-\jmath\infty}^{c+\jmath\infty} x^{-s}\mathcal{G}(s)\,ds\,. \tag{3.8}$$

The notation for the integral limits implies that this is a line integral taken over a vertical line $\mathcal{R}\mathrm{e}(s) = c$ in the complex plane, which must lie within the analytic strip.

Comprehensive accounts on the univariate Mellin transform, including lists of fundamental properties, examples of important transforms and tables of basic integrals, are found in [Poularikas, 1999, Bertrand et al., 2000, Szpankowski, 2001, Debnath and Bhatta, 2007]. An extension of the transform to complex variables is given in [Kotlarski, 1965] and to multivariate variables in [Mathur and Krishna, 1977]. The extension to the matrix-variate case is the next topic.

3.4 Matrix-Variate Mellin Transform

A generalised transform for functions of real matrices, named the M-transform, was defined in [Mathai, 1978]. Mathai also referred to it as the generalised Mellin transform in [Mathai, 1981]. The extension to functions of complex matrices was presented in [Mathai, 1997]. In the following, let $g(\mathbf{X})$ be a real-valued scalar function defined on Ω_+, and let g be symmetric in the sense: $g(\mathbf{XY}) = g(\mathbf{YX}) = g(\mathbf{X}^{1/2}\mathbf{YX}^{1/2}) = g(\mathbf{Y}^{1/2}\mathbf{XY}^{1/2})$, where $\mathbf{X}, \mathbf{Y} \in \Omega_+$. The matrix square roots defined by $\mathbf{X} = \mathbf{X}^{1/2}\mathbf{X}^{1/2}$ and $\mathbf{Y} = \mathbf{Y}^{1/2}\mathbf{Y}^{1/2}$ are therefore guaranteed to exist. Whenever the integral exists, the complex matrix-variate Mellin transform of $g(\mathbf{X})$ is defined by

$$\mathcal{M}\{g(\mathbf{X})\}(s) = \int_{\Omega_+} |\mathbf{X}|^{s-d} g(\mathbf{X})\,d\mathbf{X} = G(s)\,. \tag{3.9}$$

We note that $\mathcal{M}\{g(\mathbf{X})\}(s)$ is a function of a complex scalar transform variable s, whereas $g(\mathbf{X})$ is defined on a matrix space, thus the transform is not unique and has no inverse. The symmetry requirement restricts the functions (3.9) can be applied to, but does not pose any problem for the distributions of the polarimetric covariance matrix, derived from the doubly stochastic product model. We may therefore use the transform to define Mellin kind statistics for the complex matrix-variate case.

The univariate Mellin transform has a convolution theorem and a correlation theorem. To derive the analogue relations for the matrix-variate Mellin transform, we need some definitions. Let $g(\mathbf{X})$ and $h(\mathbf{X})$ be two real-valued scalar functions defined on Ω_+, symmetric by the definition given above. The *matrix-variate Mellin convolution* is introduced by the definition

$$\begin{aligned}(g \mathbin{\hat{\star}} h)(\mathbf{X}) &= \int_{\Omega_+} |\mathbf{Y}|^{-d} g(\mathbf{Y}^{-\frac{1}{2}}\mathbf{X}\mathbf{Y}^{-\frac{1}{2}}) h(\mathbf{Y})\, d\mathbf{Y} \\ &= \int_{\Omega_+} |\mathbf{Y}|^{-d} h(\mathbf{Y}^{-\frac{1}{2}}\mathbf{X}\mathbf{Y}^{-\frac{1}{2}}) g(\mathbf{Y})\, d\mathbf{Y}\,,\end{aligned} \tag{3.10}$$

which is an associative and commutative operation. The *matrix-variate Mellin correlation* is also introduced, defined by

$$(g \mathbin{\hat{\otimes}} h)(\mathbf{X}) = \int_{\Omega_+} |\mathbf{Y}|^{d} g(\mathbf{Y}^{\frac{1}{2}}\mathbf{X}\mathbf{Y}^{\frac{1}{2}}) h(\mathbf{Y})\, d\mathbf{Y}\,. \tag{3.11}$$

This operation is neither associative nor commutative. It reduces to the univariate Mellin correlation defined in [Nicolas, 2002, Nicolas, 2006].

3.5 Fundamental Properties

It is now time to deduce some fundamental properties of the matrix-variate Mellin transform, in analogy with the fundamental properties that have been derived for the univariate transform.

Property 1 (Scaling by nonsingular matrix of constants):

$$\mathcal{M}\{g(\mathbf{A}\mathbf{X})\}(s) = |\mathbf{A}|^{-s} G(s)\,. \tag{3.12}$$

Proof 1: The transformation $\mathbf{Y} = \mathbf{A}\mathbf{X}$ has Jacobian determinant $|J(\mathbf{X} \to \mathbf{Y})| = |\mathbf{A}|^d$ [Mathai, 1997]. Thus,

$$\begin{aligned}\mathcal{M}\{g(\mathbf{A}\mathbf{X})\}(s) &= \int_{\Omega_+} |\mathbf{X}|^{s-d} g(\mathbf{A}\mathbf{X})\, d\mathbf{X} \\ &= \int_{\Omega_+} (|\mathbf{Y}|/|\mathbf{A}|)^{s-d} g(\mathbf{Y}) |\mathbf{A}|^{-d}\, d\mathbf{Y} = |\mathbf{A}|^{-s} \int_{\Omega_+} |\mathbf{Y}|^{s-d} g(\mathbf{Y})\, d\mathbf{Y}\,,\end{aligned}$$

which is equivalent to (3.12). ∎

Property 2 (Scaling by scalar constant):

$$\mathcal{M}\{g(a\mathbf{X})\}(s) = a^{-ds} G(s)\,. \tag{3.13}$$

Proof 2: This is a special case of (3.12) with $\mathbf{A} = a\mathbf{I}_d$, where $\mathbf{I}_d$ is the $d \times d$ identity matrix. Equation (3.13) follows from $|\mathbf{A}| = a^d$. ∎

Property 3 (Multiplication by $|\mathbf{X}|^a$):

$$\mathcal{M}\{|\mathbf{X}|^a g(\mathbf{X})\}(s) = G(s+a)\,. \tag{3.14}$$

Proof 3: The proof is trivial. Take the transform of $|\mathbf{X}|^a g(\mathbf{X})$:

$$\mathcal{M}\{|\mathbf{X}|^a g(\mathbf{X})\}(s) = \int_{\Omega_+} |\mathbf{X}|^{s-d} |\mathbf{X}|^a g(\mathbf{X})\, d\mathbf{X} = \int_{\Omega_+} |\mathbf{X}|^{(s+a)-d} g(\mathbf{X})\, d\mathbf{X}$$

and identify this as (3.14). ∎

Property 4 (Raising the independent variable to an integer power):

$$\mathcal{M}\{g(\mathbf{X}^a)\}(s) = \frac{1}{a^d}\, G\left(\frac{s}{a}\right)\,. \tag{3.15}$$

Proof 4: We use the transformation $\mathbf{Y} = \mathbf{X}^a = \overbrace{\mathbf{X}\cdots\mathbf{X}}^{a}$ with differential relation $d\mathbf{Y} = a^d |\mathbf{X}|^{(a-1)d}\, d\mathbf{Y}$ to show that

$$\begin{aligned}
\mathcal{M}\{g(\mathbf{X}^a)\}(s) &= \int_{\Omega_+} |\mathbf{X}|^{s-d} g(\mathbf{X}^a)\, d\mathbf{X} \\
&= \int_{\Omega_+} |\mathbf{Y}|^{\frac{s-d}{a}} g(\mathbf{Y}) \left(a^{-d} |\mathbf{Y}|^{-\frac{(a-1)d}{a}}\, d\mathbf{Y}\right) \\
&= \frac{1}{a^d} \int_{\Omega_+} |\mathbf{Y}|^{\frac{s}{a}-d}\, d\mathbf{Y} = \frac{1}{a^d}\, G\left(\frac{s}{a}\right)\,. \qquad ∎
\end{aligned}$$

Property 5 (Inverting the independent variable):

$$\mathcal{M}\{g(\mathbf{X}^{-1})\}(s) = (-1)^d\, G(-s)\,. \tag{3.16}$$

Proof 5: This is a special case of (3.15) with $a = -1$. ∎

Property 6 (Inverting and multiplying by the independent variable):

$$\mathcal{M}\{|\mathbf{X}|^{-1} g(\mathbf{X}^{-1})\}(s) = (-1)^d\, G(1-s)\,. \tag{3.17}$$

Proof 6: Use the transformation $\mathbf{Y} = \mathbf{X}^{-1}$ with differential relation $d\mathbf{Y} = (-1)^d|\mathbf{X}|^{-2d}$, which yields

$$\begin{aligned}\mathcal{M}\{|\mathbf{X}|^{-1}g(\mathbf{X}^{-1})\}(s) &= \int_{\Omega_+} |\mathbf{X}|^{s-d}|\mathbf{X}|^{-1}g(\mathbf{X}^{-1})\,d\mathbf{X} \\ &= \int_{\Omega_+} |\mathbf{Y}|^{-(s-d-1)}g(\mathbf{Y})\left((-1)^{-d}|\mathbf{Y}|^{-2d}\,d\mathbf{Y}\right) \\ &= (-1)^d \int_{\Omega_+} |\mathbf{Y}|^{1-s-d}g(\mathbf{Y})\,d\mathbf{Y} = (-1)^d\,G(1-s)\,. \qquad \blacksquare\end{aligned}$$

Property 7 (Multiplication by $\ln|\mathbf{X}|$):

$$\mathcal{M}\{\ln|\mathbf{X}|g(\mathbf{X})\}(s) = \frac{d}{ds}G(s)\,. \tag{3.18}$$

Proof 7: By using the result

$$\frac{d}{ds}|\mathbf{X}|^{s-d} = \frac{d}{ds}e^{\ln|\mathbf{X}|(s-d)} = |\mathbf{X}|^{s-d}\ln|\mathbf{X}|$$

and Leibniz's integral rule, we show that

$$\begin{aligned}\frac{d}{ds}G(s) &= \int_{\Omega_+}\left(\frac{d}{ds}|\mathbf{X}|^{s-d}\right)g(\mathbf{X})\,d\mathbf{X} \\ &= \int_{\Omega_+} |\mathbf{X}|^{s-d}\ln|\mathbf{X}|g(\mathbf{X})\,d\mathbf{X} = \mathcal{M}\{\ln|\mathbf{X}|g(\mathbf{X})\}\,. \qquad \blacksquare\end{aligned}$$

Property 8 (Multiplication by a power of $\ln|\mathbf{X}|$):

$$\mathcal{M}\{(\ln|\mathbf{X}|)^\nu g(\mathbf{X})\}(s) = \frac{d^\nu}{ds^\nu}G(s)\,. \tag{3.19}$$

Proof 8: We use the result

$$\frac{d^\nu}{ds^\nu}|\mathbf{X}|^{s-d} = |\mathbf{X}|^{s-d}(\ln|\mathbf{X}|)^\nu$$

and follow the approach of Proof 7. $\blacksquare$

Property 9 (Mellin convolution):

$$\mathcal{M}\{(g\,\hat{\star}\,h)(\mathbf{X})\}(s) = \mathcal{M}\{g(\mathbf{X})\}(s)\cdot\mathcal{M}\{h(\mathbf{X})\}(s)\,. \tag{3.20}$$

Proof 9: Substitute $\mathbf{Z} = \mathbf{XY}$ and note that $\mathbf{Z} \in \Omega_+$ follows. We further have $\mathbf{X} = \mathbf{Y}^{-1/2}\mathbf{Z}\mathbf{Y}^{-1/2}$ and $d\mathbf{X} = d\mathbf{Z}/|\mathbf{Y}|^d$ [Mathai, 1997], which leads to

$$\begin{aligned}
&\mathcal{M}\{g(\mathbf{X})\}(s) \cdot \mathcal{M}\{h(\mathbf{Y})\}(s) \\
&\quad = \int_{\Omega_+} (|\mathbf{Z}|/|\mathbf{Y}|)^{s-d} g(\mathbf{Y}^{-\frac{1}{2}}\mathbf{Z}\mathbf{Y}^{-\frac{1}{2}})|\mathbf{Y}|^{-d}\, d\mathbf{Z} \int_{\Omega_+} |\mathbf{Y}|^{s-d} h(\mathbf{Y})\, d\mathbf{Y} \\
&\quad = \int_{\Omega_+} |\mathbf{Z}|^{s-d} \left[\int_{\Omega_+} |\mathbf{Y}|^{-d} g(\mathbf{Y}^{-\frac{1}{2}}\mathbf{Z}\mathbf{Y}^{-\frac{1}{2}}) h(\mathbf{Y})\, d\mathbf{Y} \right] d\mathbf{Z} \\
&\quad = \mathcal{M}\{(g \mathbin{\hat{\star}} h)(\mathbf{Z})\}(s)\,,
\end{aligned}$$

since the term in the square brackets can be identified as the Mellin convolution from (3.10). ∎

Property 10 (Mellin correlation):

$$\mathcal{M}\{(g \mathbin{\hat{\otimes}} h)(\mathbf{X})\}(s) = \mathcal{M}\{g(\mathbf{X})\}(s) \cdot \mathcal{M}\{h(\mathbf{X})\}(2d - s)\,. \tag{3.21}$$

Proof 10: Use the substitution $\mathbf{X} = \mathbf{YZ}$ with differential relation $d\mathbf{X} = |\mathbf{Y}|^d d\mathbf{Z}$ to find

$$\begin{aligned}
&\mathcal{M}\{g(\mathbf{X})\}(s) \cdot \mathcal{M}\{h(\mathbf{Y})\}(2d - s) \\
&\quad = \int_{\Omega_+} (|\mathbf{YZ}|)^{s-d} g(\mathbf{YZ}) |\mathbf{Y}|^d\, d\mathbf{Z} \int_{\Omega_+} |\mathbf{Y}|^{d-s} h(\mathbf{Y})\, d\mathbf{Y} \\
&\quad = \int_{\Omega_+} |\mathbf{Z}|^{s-d} \left[\int_{\Omega_+} |\mathbf{Y}|^{-d} g(\mathbf{YZ}) h(\mathbf{Y})\, d\mathbf{Y} \right] d\mathbf{Z} \\
&\quad = \mathcal{M}\{(g \mathbin{\hat{\otimes}} h)(\mathbf{X})\}(s)\,,
\end{aligned}$$

after the term in the square brackets is identified as the matrix-variate Mellin correlation defined in (3.11). ∎

3.6 Matrix-variate Mellin Kind Statistics

The matrix-variate Mellin transform is defined on Ω_+, and can therefore be applied to PDFs that have the same domain. It is this property that makes it relevant to the statistical analysis of distributions for the polarimetric covariance matrix. From now on, we assume that $\mathbf{X}$ is a random matrix described by $f_{\mathbf{X}}(\mathbf{X})$, which is defined on Ω_+.

3.6.1 Mellin Kind Characteristic Function

The Mellin kind characteristic function of a complex random matrix $\mathbf{X}$ is defined as

$$\phi_{\mathbf{X}}(s) = \mathrm{E}\{|\mathbf{X}|^{s-d}\} = \mathcal{M}\{f_{\mathbf{X}}(\mathbf{X})\}(s)\,. \tag{3.22}$$

when $\mathbf{X}$ and its PDF, $p_{\mathbf{X}}(\mathbf{X})$, satisfy all requirements of (3.9). Note that the moments of the determinant $|\mathbf{X}|$ can be retrieved from the Mellin kind characteristic function as $\mathrm{E}\{|\mathbf{X}|^{\nu}\} = \phi_{\mathbf{X}}(\nu + d)$.

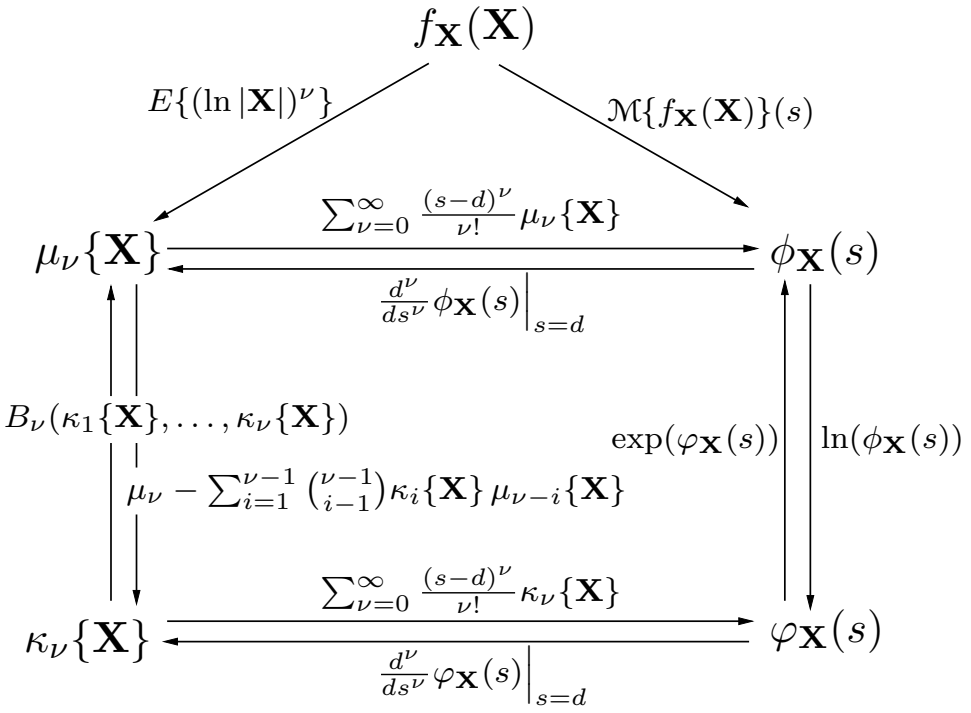

Figure 3.1: Relations between the components of matrix-variate Mellin kind statistics. The function $B_{\nu}(\cdot)$ is the complete Bell polynomial, defined in Paper 2 (Chapter 6).

3.6.2 Mellin Kind Matrix Moments

The νth-order Mellin kind moment of $\mathbf{X}$ is

$$\mu_{\nu}\{\mathbf{X}\} = \frac{d^{\nu}}{ds^{\nu}} \phi_{\mathbf{X}}(s)\Big|_{s=d} . \tag{3.23}$$

If all Mellin kind matrix moments exist, the Mellin kind characteristic function can be written as the power series expansion

$$\begin{aligned} \phi_{\mathbf{X}}(s) &= \int_{\Omega_+} e^{(s-d)\ln|\mathbf{X}|} f_{\mathbf{X}}(\mathbf{X})\, d\mathbf{X} \\ &= \sum_{\nu=0}^{\infty} \frac{(s-d)^{\nu}}{\nu!} \mu_{\nu}\{\mathbf{X}\} \end{aligned} \tag{3.24}$$

in terms of the $\mu_{\nu}\{\mathbf{X}\}$. The derivation of (3.24) reveals that

$$\mu_{\nu}\{\mathbf{X}\} = \mathrm{E}\{(\ln|\mathbf{X}|)^{\nu}\} = \int_{\Omega_+} (\ln|\mathbf{X}|)^{\nu} f_{\mathbf{X}}(\mathbf{X})\, d\mathbf{X} \tag{3.25}$$

which justifies the denotation of $\mu_{\nu}\{\mathbf{X}\}$ as a matrix log-moment.

3.6.3 Mellin Kind Cumulant Generating Function

The Mellin kind cumulant generating function of $\mathbf{X}$ is defined as

$$\varphi_{\mathbf{X}}(s) = \ln \phi_{\mathbf{X}}(s)\,. \tag{3.26}$$

3.6.4 Mellin Kind Matrix Cumulants

The νth-order Mellin kind cumulant of $\mathbf{X}$ is

$$\kappa_{\nu}\{\mathbf{X}\} = \frac{d^{\nu}}{ds^{\nu}}\varphi_{\mathbf{X}}(s)\bigg|_{s=d}\,. \tag{3.27}$$

When all Mellin kind matrix moments exist, the Mellin kind cumulant generating function can be expanded as

$$\varphi_{\mathbf{X}}(s) = \ln \phi_{\mathbf{X}}(s) = \sum_{\nu=0}^{\infty} \frac{(s-d)^{\nu}}{\nu!}\kappa_{\nu}\{\mathbf{X}\} \tag{3.28}$$

in terms of the $\kappa_{\nu}\{\mathbf{X}\}$, that are also called matrix log-cumulants.

The relations between the the functions and moments defined under the matrix-variate Mellin kind statistics framework are shown in Figure 3.1.

3.7 Multilook Polarimetric Product Model

This section shows how the matrix-variate Mellin kind statistics are applied to the doubly stochastic product model for multivariate polarimetric data. We recollect the model as

$$\mathbf{C} = \mathbf{T} \cdot \widetilde{\mathbf{W}}\,, \tag{3.29}$$

where $\mathbf{C}$ is the polarimetric covariance matrix, decomposed in terms of the random matrices $\widetilde{\mathbf{W}} \sim s\mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma})$ and $\mathbf{T}$, respectively modelling fully developed speckle and texture. The PDF of $\mathbf{T}$, $f_{\mathbf{T}}(\mathbf{T}; \boldsymbol{\theta})$, is left unspecified for the moment. We here assume that $\mathbf{T}$ has equal diagonal entries, $\mathbf{T} = \mathrm{diag}(T, \cdots, T) = T \cdot \mathbf{I}_d$, thereby restricting the texture model to a single random variable representing all polarimetric channels. The advantage of writing the scalar texture variable as a matrix will become apparent.

3.7.1 Application of the Matrix-Variate Mellin Convolution

The distribution of $\mathbf{C}$ can now be written as

$$f_{\mathbf{C}}(\mathbf{C}; L, \mathbf{\Sigma}, \boldsymbol{\theta}) = \int_{\Omega_+} f_{\mathbf{C}|\mathbf{T}}(\mathbf{C}|\mathbf{T}; L, \mathbf{\Sigma}) f_{\mathbf{T}}(\mathbf{T}; \boldsymbol{\theta})\, d\mathbf{C}\,, \tag{3.30}$$

parametrised by the equivalent number of looks L, scale matrix $\mathbf{\Sigma}$, and the texture parameter vector $\boldsymbol{\theta}$. The integral is identified as the matrix-variate Mellin convolution, defined in (3.10). From Property 9 of the matrix-variate Mellin transform, it follows that

$$\mathcal{M}\{f_{\mathbf{C}}(\mathbf{C}; L, \mathbf{\Sigma}, \boldsymbol{\theta})\} = \mathcal{M}\{f_{\widetilde{W}}(\widetilde{\mathbf{W}}; L, \mathbf{\Sigma})\} \cdot \mathcal{M}\{f_{\mathbf{T}}(\mathbf{T}; \boldsymbol{\theta})\}\,. \tag{3.31}$$

The implication in terms of the matrix-variate Mellin statistics defined in the previous section is

$$\phi_{\mathbf{C}}(s; L, \mathbf{\Sigma}, \boldsymbol{\theta}) = \phi_{\widetilde{\mathbf{W}}}(s; L, \mathbf{\Sigma}) \cdot \phi_{\mathbf{T}}(s; \boldsymbol{\theta})\,, \tag{3.32}$$

$$\varphi_{\mathbf{C}}(s; L, \mathbf{\Sigma}, \boldsymbol{\theta}) = \varphi_{\widetilde{\mathbf{W}}}(s; L, \mathbf{\Sigma}) + \varphi_{\mathbf{T}}(s; \boldsymbol{\theta})\,, \tag{3.33}$$

$$\kappa_\nu\{\mathbf{C}; L, \mathbf{\Sigma}, \boldsymbol{\theta}\} = \kappa_\nu\{\widetilde{\mathbf{W}}; L, \mathbf{\Sigma}\} + \kappa_\nu\{\mathbf{T}; \boldsymbol{\theta}\}\,. \tag{3.34}$$

These respective equations present the Mellin kind characteristic function, the Mellin kind cumulant generating function and the Mellin kind cumulants for the multivariate polarimetric product model. The Mellin kind moments can be retrieved from the cumulants by the relation

$$\kappa_\nu\{\cdot\} = \mu_\nu\{\cdot\} - \sum_{i=1}^{\nu-1} \binom{\nu-1}{i-1} \kappa_i\{\cdot\}\mu_{\nu-i}\{\cdot\}\,, \tag{3.35}$$

which is valid both in the univariate and matrix-variate case, and for moments and cumulants on linear or logarithmic scale.

3.7.2 Mellin Kind Statistics for the Multilook Polarimetric Product Model

The texture variable was written as a matrix so that the matrix-variate convolution property could be used directly. The Mellin kind statistics of $\mathbf{T}$ must still be resolved, and they have been in the paper of Chapter 6. It is found that

$$\phi_{\mathbf{T}}(s; \boldsymbol{\theta}) = \phi_T(d(s-d)+1; \boldsymbol{\theta})\,, \tag{3.36}$$

where $\phi_T(s)$ is the univariate Mellin kind characteristic function of the scalar variable T, derived in [Nicolas, 2002, Nicolas, 2006] as

$$\phi_T(s) = \int_0^\infty T^{s-1} f_T(T; \boldsymbol{\theta})\, dT = \mathrm{E}\{T^{s-1}; \boldsymbol{\theta}\}\,. \tag{3.37}$$

The cumulant generating function for $\mathbf{T}$ follows readily as

$$\varphi_{\mathbf{T}}(s; \boldsymbol{\theta}) = \varphi_T(d(s-d)+1; \boldsymbol{\theta})\,, \tag{3.38}$$

with $\varphi_T(s; \boldsymbol{\theta}) = \ln \phi_T(s; \boldsymbol{\theta})$. The Mellin kind cumulants are retrieved as

$$\kappa_\nu\{\mathbf{T}\} = d^\nu \kappa_\nu\{T\}\,, \tag{3.39}$$

where the $\kappa_\nu\{T;\boldsymbol{\theta}\}$ are the νth-order univariate Mellin kind cumulants defined in [Nicolas, 2002, Nicolas, 2006]. We thus arrive at the following Mellin kind statistics for the multivariate polarimetric product model:

$$\phi_{\mathbf{C}}(s;L,\boldsymbol{\Sigma},\boldsymbol{\theta}) = \phi_{\widetilde{\mathbf{W}}}(s;L,\boldsymbol{\Sigma})\cdot\phi_T(d(s-d)+1;\boldsymbol{\theta})\,, \tag{3.40}$$

$$\varphi_{\mathbf{C}}(s;L,\boldsymbol{\Sigma},\boldsymbol{\theta}) = \varphi_{\widetilde{\mathbf{W}}}(s;L,\boldsymbol{\Sigma}) + \varphi_T(d(s-d)+1;\boldsymbol{\theta})\,, \tag{3.41}$$

$$\kappa_\nu\{\mathbf{C};L,\boldsymbol{\Sigma},\boldsymbol{\theta}\} = \kappa_\nu\{\widetilde{\mathbf{W}};L,\boldsymbol{\Sigma}\} + d^\nu\kappa_\nu\{T;\boldsymbol{\theta}\}\,. \tag{3.42}$$

This is a general formulation, which requires input for specific choices of the texture variable distribution, $f_T(T;\boldsymbol{\theta})$. A set of candidate distributions for the univariate texture is reviewed in the next chapter. Explicit expressions for the Mellin kind statistics of the speckle distribution, $f_{\widetilde{W}}(\widetilde{W};L,\boldsymbol{\Sigma})$ are also given, such that Equations (3.40)–(3.42) can be evaluated.

Chapter 4

Probability Distributions

This chapter presents specific distributions for multilook polarimetric radar data, and evaluate their Mellin kind statistics. The mathematical expressions contain many special functions, that are defined initially. We then review a number of univariate distributions that are useful models for the texture random variable. Expressions are given for their PDF, Mellin kind characteristic function, and Mellin kind cumulants (log-cumulants). These have been derived previously in [Nicolas, 2006]. The scaled complex Wishart distribution used to model speckle is then presented, together with its Mellin kind statistics, that have been derived in Paper 2 of Chapter 6. Finally, we present the compound distributions modelling the texture modulated speckle of the polarimetric covariance matrix. The Mellin kind statistics of the texture distributions and the speckle distribution are combined into novel expressions for the multilook polarimetric case.

4.1 Special Functions

The expression of the PDFs, Mellin kind characteristic functions, moments and cumulants presented in this chapter require a number of special functions. These are defined below.

Gamma Function: The gamma function, often credited explicitly to Euler by name, is a generalisation of the factorial function to non-integer numbers. It is defined for $z \in \mathbb{C}$ with $\mathcal{R}\mathrm{e}\{z\} > 0$ by [Weisstein, 2010c]

$$\Gamma(z) = \int_0^\infty u^{z-1} e^{-u}\, du \tag{4.1}$$

and satisfies the relation

$$\Gamma(z+1) = z\Gamma(z)\,. \tag{4.2}$$

Multivariate Gamma Function of the Complex Kind: The multivariate gamma function of the complex kind is a generalisation of the gamma function. It is defined for $s \in \mathbb{C}$ by [Goodman, 1963]

$$\Gamma_d(s) = \int_{\Omega_+} \exp\left(-\operatorname{tr}(\mathbf{Z})\right) |\mathbf{Z}|^{s-d}\, d\mathbf{Z} = \pi^{d(d-1)/2} \prod_{i=0}^{d-1} \Gamma(s-i) \tag{4.3}$$

where Ω_+ is the cone of positive definite complex Hermitian matrices, $\operatorname{tr}(\cdot)$ is the trace operator, and the matrix $\mathbf{Z} \in \Omega_+$ has dimensions $d \times d$. We also note that $\Gamma_d(z)$ is the matrix-variate Mellin transform of $\exp(\operatorname{tr}(\mathbf{Z}))$.

Polygamma Function: The polygamma function of order ν is the $\nu+1$th derivative of the logarithm of the gamma function. It is defined by [Weisstein, 2010e]

$$\begin{aligned} \psi^{(\nu)}(z) &= (-1)^{(\nu+1)} \int_0^\infty \frac{u^\nu e^{-zu}}{1-e^{-u}}\, du \\ &= \frac{d^\nu}{dz^\nu}\psi^{(0)} = \frac{d^{\nu+1}}{dz^{\nu+1}} \ln \Gamma(z) \end{aligned} \tag{4.4}$$

where $\psi^{(0)}(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function. A recurrence relation is given by

$$\psi^{(\nu)}(z+1) = \psi^{(\nu)}(z) + (-1)^\nu \nu!\, z^{-(\nu+1)}\,. \tag{4.5}$$

Multivariate Polygamma Function of the Complex Kind: The multivariate polygamma function of the complex kind is introduced in this thesis as an intuitive and simplifying generalisation of the ordinary polygamma function (see the previous paragraph). It is defined as

$$\psi_d^{(\nu)}(z) = \sum_{i=0}^{d-1} \psi^{(\nu)}(z-i) = \frac{d^{\nu+1}}{dz^{\nu+1}} \ln \Gamma_d(z)\,. \tag{4.6}$$

The multivariate digamma function satisfies (see derivations in Paper 2, Chapter 6)

$$\psi_d^{(0)}(z) = \frac{\Gamma_d'(z)}{\Gamma_d(z)}\,. \tag{4.7}$$

Beta Function: The beta function, also called the Euler integral of the first kind, is defined for $\{x, y\} \in \mathbb{C}$ with $\Re\{x\} > 0$ and $\Re\{y\} > 0$ by [Weisstein, 2010a]

$$B(x,y) = \int_0^1 u^{x-1}(1-x)^{y-1}\, du = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}\,. \tag{4.8}$$

Modified Bessel function of the Second Kind: The modified Bessel function of the second kind, also known as the Bessel K function, is defined for order $\nu \in \mathbb{R}$ and argument $z \in \mathbb{C}$ by [Weisstein, 2010d]

$$K_\nu(z) = \frac{\Gamma(\nu + 1/2)(2z)^\nu}{\sqrt{\pi}} \int_0^\infty \frac{\cos u}{(u^2 + z^2)^{\nu+1/2}}\, du\,. \tag{4.9}$$

Confluent Hypergeometric Function of the Second Kind: The confluent hypergeometric function of the second kind, also known as the Kummer U function, is defined for arguments $\{a, b, z\} \in \mathbb{C}$ with $\mathcal{R}\mathrm{e}\{a\}, \mathcal{R}\mathrm{e}\{z\} > 0$ by [Weisstein, 2010b]

$$U(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zu} u^{a-1} (1 + u)^{b-a-1}\, du\,. \tag{4.10}$$

4.2 Texture Distributions

According to the discussion of texture in Section 2.3.5, a *texture distribution* describes a real random variable T, which is strictly positive and unit mean ($T \in \mathbb{R}^+,\ \mathrm{E}\{T\} = 1$). We here present some candidate distributions and expressions derived from them with full parametrisation. That is, they contain a location parameter μ, which allows for a mean value $\mu \neq 1$. This makes it easier to recognise the distribution from the literature, while the normalised expressions used in the context of the product model are easily obtained by the substitution of $\mu = 1$.

The texture distributions presented in the following were all treated in [Nicolas, 2006], where the PDF, Mellin kind characteristic function, νth-order moment and νth-order log-cumulant were given for each of them. The parametrisation used here is slightly different from the one Nicolas uses, in insisting that μ should be identical to the mean for all distributions. Departures from Nicolas' parametrisation are remarked for the distributions concerned. The advantage of this approach is that all expressions for the normalised distribution are obtained simply by setting $\mu = 1$. The shape parameters α and λ remain unchanged. On the other hand, the parametrisation of Nicolas is more elegant when showing how the composite distributions (i.e., the Fisher-Snedecor, beta, and inverse beta distribution) are composed by Mellin correlation, convolution, and inverse convolution of gamma distributions and inverse gamma distributions (see [Nicolas, 2006] for details).

4.2.1 Gamma Distribution

A gamma distributed random variable is denoted $X \sim \gamma(\mu, \alpha)$ with location parameter $\mu > 0$ and shape parameter $\alpha > 0$. A gamma distributed texture variable is denoted $T \sim \bar{\gamma}(\alpha) = \gamma(1, \alpha)$, due to the normalisation of $\mu = 1$.

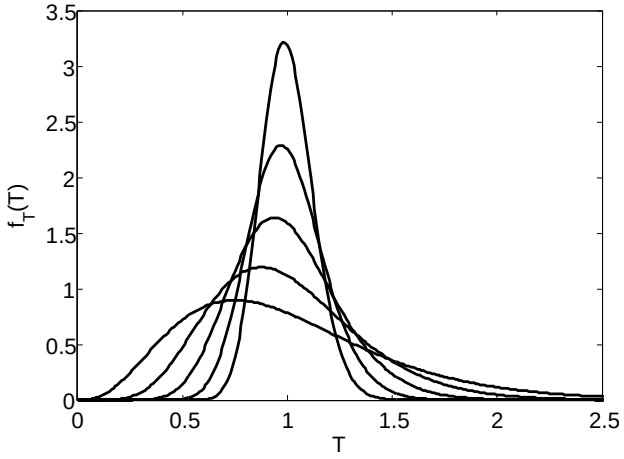

Figure 4.1: Probability density function $f_T(T;\alpha)$ of the unit mean gamma distributed texture variable T shown for shape parameters $\alpha = \{4, 8, 16, 32, 64\}$. The peakedness increases with α.

Probability density function:

$$f_X(x;\mu,\alpha) = \frac{1}{\Gamma(\alpha)} \left(\frac{\alpha}{\mu}\right) \left(\frac{\alpha x}{\mu}\right)^{\alpha-1} e^{-\frac{\alpha x}{\mu}}; \quad x \geq 0\,. \tag{4.11}$$

Mellin kind characteristic function:

$$\phi_X(s;\mu,\alpha) = \left(\frac{\mu}{\alpha}\right)^{s-1} \frac{\Gamma(\alpha+s-1)}{\Gamma(\alpha)}\,. \tag{4.12}$$

Moments:

$$m_\nu\{X;\mu,\alpha\} = \left(\frac{\mu}{\alpha}\right)^{\nu} \frac{\Gamma(\alpha+\nu)}{\Gamma(\alpha)}\,. \tag{4.13}$$

Log-cumulants:

$$\kappa_\nu\{X;\mu,\alpha\} = \begin{cases} \psi^{(0)}(\alpha) + \ln\left(\frac{\mu}{\alpha}\right) & \text{for } \nu = 1 \\ \psi^{(\nu-1)}(\alpha) & \text{for } \nu > 1 \end{cases}. \tag{4.14}$$

4.2.2 Inverse Gamma Distribution

An inverse gamma distributed random variable is denoted $X \sim \gamma^{-1}(\mu,\lambda)$ with location parameter $\mu > 0$ and shape parameter $\lambda > 0$. An inverse gamma distributed texture variable is denoted $T \sim \bar{\gamma}^{-1}(\lambda) = \gamma^{-1}(1,\lambda)$.

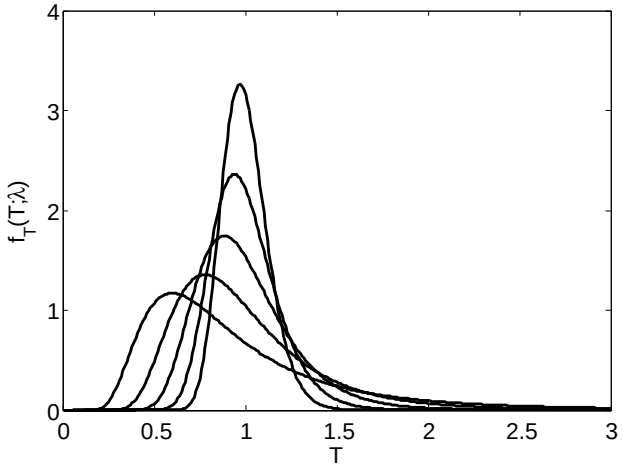

Figure 4.2: Probability density function $f_T(T;\lambda)$ of the unit mean inverse gamma distributed texture variable T shown for shape parameters $\lambda = \{4, 8, 16, 32, 64\}$. The peakedness increases with λ.

Probability density function:

$$f_X(x;\mu,\lambda) = \frac{1}{\Gamma(\lambda)}\frac{1}{(\lambda-1)\mu}\left(\frac{(\lambda-1)\mu}{x}\right)^{\lambda+1} e^{-\frac{(\lambda-1)\mu}{x}}; \quad x>0\,. \tag{4.15}$$

Mellin kind characteristic function:

$$\phi_X(s;\mu,\lambda) = \big((\lambda-1)\mu\big)^{s-1}\frac{\Gamma(\lambda+1-s)}{\Gamma(\lambda)}\,. \tag{4.16}$$

Moments:

$$m_\nu\{X;\mu,\lambda\} = \big((\lambda-1)\mu\big)^{\nu}\frac{\Gamma(\lambda-\nu)}{\Gamma(\lambda)}\,. \tag{4.17}$$

Log-cumulants:

$$\kappa_\nu\{X;\mu,\lambda\} = \begin{cases} -\psi^{(0)}(\lambda) + \ln((\lambda-1)\mu) & \text{for } \nu = 1 \\ (-1)^\nu \psi^{(\nu-1)}(\lambda) & \text{for } \nu > 1 \end{cases}. \tag{4.18}$$

The parametrisation of [Nicolas, 2006] is obtained by substituting $\mu = \left(\frac{\lambda}{\lambda-1}\right)\mu'$.

4.2.3 Fisher-Snedecor Distribution

A Fisher-Snedecor distributed random variable is denoted $X \sim \mathcal{F}(\mu,\alpha,\lambda)$ with location parameter $\mu > 0$ and shape parameters $\alpha > 0$ and $\lambda > 0$. A Fisher-Snedecor distributed texture variable is denoted $T \sim \bar{\mathcal{F}}(\alpha,\lambda) = \mathcal{F}(1,\alpha,\lambda)$.

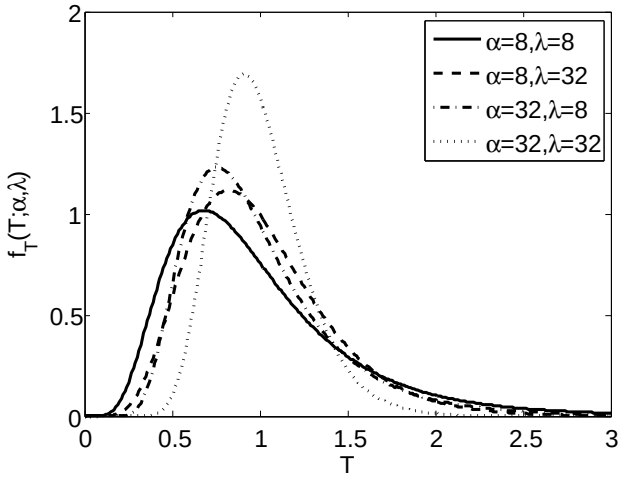

Figure 4.3: Probability density function $f_T(T;\alpha,\lambda)$ of the unit mean Fisher-Snedecor distributed texture variable T shown for different pairs of shape parameters (α,λ).

Probability density function:

$$f_X(x;\mu,\alpha,\lambda) = \frac{1}{B(\alpha,\lambda)}\frac{\alpha}{(\lambda-1)\mu}\frac{\left(\frac{\alpha x}{(\lambda-1)\mu}\right)^{\alpha-1}}{\left(\frac{\alpha x}{(\lambda-1)\mu}+1\right)^{\alpha+\lambda}}\,; \quad x \geq 0\,. \tag{4.19}$$

Mellin kind characteristic function:

$$\phi_X(s;\mu,\alpha,\lambda) = \left(\frac{(\lambda-1)\mu}{\alpha}\right)^{s-1}\frac{\Gamma(\alpha+s-1)}{\Gamma(\alpha)}\frac{\Gamma(\lambda+1-s)}{\Gamma(\lambda)}\,. \tag{4.20}$$

Moments:

$$m_\nu\{X;\mu,\alpha,\lambda\} = \left(\frac{(\lambda-1)\mu}{\alpha}\right)^{\nu}\frac{\Gamma(\alpha+\nu)}{\Gamma(\alpha)}\frac{\Gamma(\lambda-\nu)}{\Gamma(\lambda)}\,. \tag{4.21}$$

Log-cumulants:

$$\kappa_\nu\{X;\mu,\alpha,\lambda\} = \begin{cases}\psi^{(0)}(\alpha)-\psi^{(0)}(\lambda)+\ln\left(\frac{(\lambda-1)\mu}{\alpha}\right) & \text{for } \nu=1\\ \psi^{(\nu-1)}(\alpha)+(-1)^\nu\psi^{(\nu-1)}(\lambda) & \text{for } \nu>1\end{cases}\,. \tag{4.22}$$

The parametrisation of [Nicolas, 2006] is obtained by substituting $\mu = \left(\frac{\lambda}{\lambda-1}\right)\mu'$. The Fisher-Snedecor distribution is also called Fisher distribution. It can be seen as a generalised F distribution with an additional location parameter.

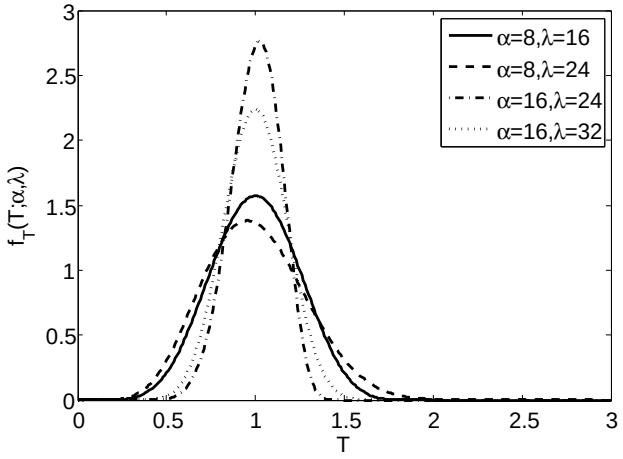

Figure 4.4: Probability density function $f_T(T;\alpha,\lambda)$ of the unit mean beta distributed texture variable T shown for different pairs of shape parameters (α,λ).

4.2.4 Beta Distribution

A beta distributed random variable is denoted $X \sim \beta(\mu,\alpha,\lambda)$ with location parameter $\mu > 0$ and shape parameters $\alpha > 0$ and $\lambda > 0$. A beta distributed texture variable is denoted $T \sim \bar{\beta}(\alpha,\lambda) = \beta(1,\alpha,\lambda)$.

Probability density function:

$$f_X(x;\mu,\alpha,\lambda) = \frac{1}{B(\alpha,\lambda-\alpha)}\frac{\alpha}{\lambda\mu}\left(\frac{\alpha x}{\lambda\mu}\right)^{\alpha-1}\left(1-\frac{\alpha x}{\lambda\mu}\right)^{\lambda-\alpha-1} \quad ; \quad 0 \le x \le \frac{\lambda\mu}{\alpha}\,. \tag{4.23}$$

Mellin kind characteristic function:

$$\phi_X(s;\mu,\alpha,\lambda) = \left(\frac{\lambda\mu}{\alpha}\right)^{s-1}\frac{\Gamma(\alpha+s-1)}{\Gamma(\alpha)}\frac{\Gamma(\lambda)}{\Gamma(\lambda+s-1)}\,. \tag{4.24}$$

Moments:

$$m_\nu\{X;\mu,\alpha,\lambda\} = \left(\frac{\lambda\mu}{\alpha}\right)^{\nu}\frac{\Gamma(\alpha+\nu)}{\Gamma(\alpha)}\frac{\Gamma(\lambda)}{\Gamma(\lambda+\nu)}\,. \tag{4.25}$$

Log-cumulants:

$$\kappa_\nu\{X;\mu,\alpha,\lambda\} = \begin{cases}\psi^{(0)}(\alpha) - \psi^{(0)}(\lambda) + \ln\left(\frac{\lambda\mu}{\alpha}\right) & \text{for } \nu = 1\\ \psi^{(\nu-1)}(\alpha) - \psi^{(\nu-1)}(\lambda) & \text{for } \nu > 1\end{cases}\,. \tag{4.26}$$

The beta distribution is often defined with two shape parameters only. The definition given here is a generalisation of the two parameter version, with an additional location parameter.

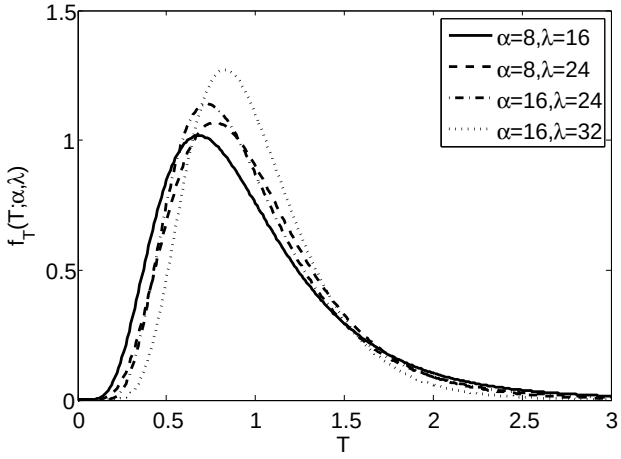

Figure 4.5: Probability density function $f_T(T;\alpha,\lambda)$ of the unit mean inverse beta distributed texture variable T shown for different pairs of shape parameters (α,λ).

4.2.5 Inverse Beta Distribution

An inverse beta distributed random variable is denoted $X \sim \beta^{-1}(\mu,\alpha,\lambda)$ with location parameter $\mu > 0$ and shape parameters $\alpha > 0$ and $\lambda > 0$. An inverse beta distributed texture variable is denoted $T \sim \bar{\beta}^{-1}(\alpha,\lambda) = \beta^{-1}(1,\alpha,\lambda)$.

Probability density function:

$$f_X(x;\mu,\alpha,\lambda) = \frac{1}{B(\alpha,\lambda-\alpha)}\frac{(\lambda-1)}{(\alpha-1)\mu}\frac{\left(\frac{(\lambda-1)x}{(\alpha-1)\mu}-1\right)^{\lambda-\alpha-1}}{\left(\frac{(\lambda-1)x}{(\alpha-1)\mu}\right)^{\lambda}}\,; \quad x \geq \frac{(\alpha-1)\mu}{\lambda-1}\,,\ \lambda \geq \alpha\,. \tag{4.27}$$

Mellin kind characteristic function:

$$\phi_X(s;\mu,\alpha,\lambda) = \left(\frac{(\alpha-1)\mu}{\lambda-1}\right)^{s-1}\frac{\Gamma(\alpha+1-s)}{\Gamma(\alpha)}\frac{\Gamma(\lambda)}{\Gamma(\lambda+1-s)}\,. \tag{4.28}$$

Moments:

$$m_\nu\{X;\mu,\alpha,\lambda\} = \left(\frac{(\alpha-1)\mu}{\lambda-1}\right)^{\nu}\frac{\Gamma(\alpha-\nu)}{\Gamma(\alpha)}\frac{\Gamma(\lambda)}{\Gamma(\lambda-\nu)}\,. \tag{4.29}$$

Log-cumulants:

$$\kappa_\nu\{X;\mu,\alpha,\lambda\} = \begin{cases} -\psi^{(0)}(\alpha)+\psi^{(0)}(\lambda)+\ln\left(\frac{(\alpha-1)\mu}{(\lambda-1)}\right) & \text{for } \nu=1 \\ (-1)^\nu\psi^{(\nu-1)}(\alpha)+(-1)^{\nu-1}\psi^{(\nu-1)}(\lambda) & \text{for } \nu>1 \end{cases}\,. \tag{4.30}$$

The parametrisation of [Nicolas, 2006] is obtained by substituting $\mu = \left(\frac{\alpha-1}{\alpha}\right)\left(\frac{\lambda}{\lambda-1}\right)\mu'$. The inverse beta distribution is also known as the beta distribution of the second kind, or the beta prime distribution. It is often defined with two shape parameters only. The definition given here is a generalisation of the two parameter version, with an additional location parameter.

4.3 Speckle Distribution

The term *speckle distribution* is here used to denote a distribution which models the randomness of the radar signal due to the interference phenomenon only. That is, it describes pure speckle. When it is compounded with a texture distribution, we obtain under the doubly stochastic product model a distribution which describes texture modulated speckle. The Weibull and log-normal distributions for single polarisation amplitude and intensity data [Oliver and Quegan, 2004] are examples of distributions that describe texture modulated speckle without explicitly modelling the texture through a texture variable. Matrix-variate counterparts of these distributions have not been derived, as far as the author is aware of. There are no known distributions that describe partially developed speckle in the multilook polarimetric case either. We are left with the scaled complex Wishart distribution, which models pure and fully developed speckle, but will also present the matrix-variate Mellin kind statistics of the true complex Wishart distribution.

4.3.1 Complex Wishart Distribution

The complex extension of the Wishart distribution, defined on Ω_+, was first reported in [Goodman, 1963]. The distribution describes $\mathbf{W} = L\widetilde{\mathbf{W}}$ and the complex Wishart distributed matrix is denoted $\mathbf{W} \sim \mathcal{W}_d^{\mathrm{C}}(L, \mathbf{\Sigma})$, where the equivalent number of looks, L, is a shape parameter and $\mathbf{\Sigma} = \mathrm{E}\{\mathbf{W}\}/L$ is a scale matrix. The nonsingular distribution is obtained for $L \geq d$.

Probability density function:

$$f_{\mathbf{W}}(\mathbf{W}; L, \mathbf{\Sigma}) = \frac{|\mathbf{C}|^{L-d}}{\Gamma_d(L)|\mathbf{\Sigma}|^L} \operatorname{etr}(-\mathbf{\Sigma}^{-1}\mathbf{C}) \,. \tag{4.31}$$

Mellin kind characteristic function:

$$\phi_{\mathbf{W}}(s; L, \mathbf{\Sigma}) = \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)}|\mathbf{\Sigma}|^{s-d} \,. \tag{4.32}$$

Mellin kind cumulant generating function:

$$\varphi_{\mathbf{W}}(s; L, \mathbf{\Sigma}) = \ln\Gamma_d(L+s-d) - \ln\Gamma_d(L) + (s-d)\ln|\mathbf{\Sigma}| \,. \tag{4.33}$$

Matrix log-cumulants:

$$\kappa_\nu\{W; L, \boldsymbol{\Sigma}\} = \begin{cases} \psi_d^{(0)}(L) + \ln|\boldsymbol{\Sigma}| & \text{for } \nu = 1\,, \\ \psi_d^{(\nu-1)}(L) & \text{for } \nu > 1\,. \end{cases} \tag{4.34}$$

The derivation of the Mellin kind statistics is shown in Paper 2 of Chapter 6.

4.3.2 Scaled Complex Wishart Distribution

The observable covariance matrix in pure and fully developed speckle is $\widetilde{\mathbf{W}}$, which follows the scaled complex Wishart distribution. This is denoted $\widetilde{\mathbf{W}} \sim s\mathcal{W}_d^{\mathbb{C}}(L, \boldsymbol{\Sigma})$. The PDF is derived from (4.31) as $f_{\widetilde{\mathbf{W}}}(\widetilde{\mathbf{W}}) = f_{\mathbf{W}}(L\widetilde{\mathbf{W}})|J_{\mathbf{W}\to\widetilde{\mathbf{W}}}|$ by using the transformation $\widetilde{\mathbf{W}} = \mathbf{W}/L$ with Jacobian determinant $|J_{\mathbf{W}\to\widetilde{W}}| = L^{d^2}$.

Probability density function:

$$f_{\mathbf{C}}(\mathbf{C}; L, \boldsymbol{\Sigma}) = \frac{L^{Ld}}{\Gamma_d(L)} \frac{|\mathbf{C}|^{L-d}}{|\boldsymbol{\Sigma}|^L} \operatorname{etr}(-L\boldsymbol{\Sigma}^{-1}\mathbf{C})\,. \tag{4.35}$$

Mellin kind characteristic function:

$$\phi_{\mathbf{W}}(s; L, \boldsymbol{\Sigma}) = \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} \left(\frac{|\boldsymbol{\Sigma}|}{L^d}\right)^{s-d}\,. \tag{4.36}$$

Mellin kind cumulant generating function:

$$\varphi_{\mathbf{W}}(s; L, \boldsymbol{\Sigma}) = \ln\Gamma_d(L+s-d) - \ln\Gamma_d(L) + (s-d)(\ln|\boldsymbol{\Sigma}| - d\ln L)\,. \tag{4.37}$$

Matrix log-cumulants:

$$\kappa_\nu\{W; L, \boldsymbol{\Sigma}\} = \begin{cases} \psi_d^{(0)}(L) + \ln|\boldsymbol{\Sigma}| - d\ln L & \text{for } \nu = 1\,, \\ \psi_d^{(\nu-1)}(L) & \text{for } \nu > 1\,. \end{cases} \tag{4.38}$$

The derivation of the Mellin kind statistics is shown in Paper 2 of Chapter 6.

4.4 Compound Matrix Distributions

Compound matrix distributions that model texture modulated speckle are obtained from the Mellin convolution in (2.26). Five distributions are reported, that combine the scaled complex Wishart distribution with each of the texture distribution presented in Section 4.2 through the multilook polarimetric product model. For all of them, the PDF is listed together with the matrix-variate Mellin kind characteristic function, cumulant generating function and cumulants. The Mellin kind statistics are derived by evaluating Equations (3.40)–(3.42) with the expressions given in Sections 4.2 and 4.3.2.

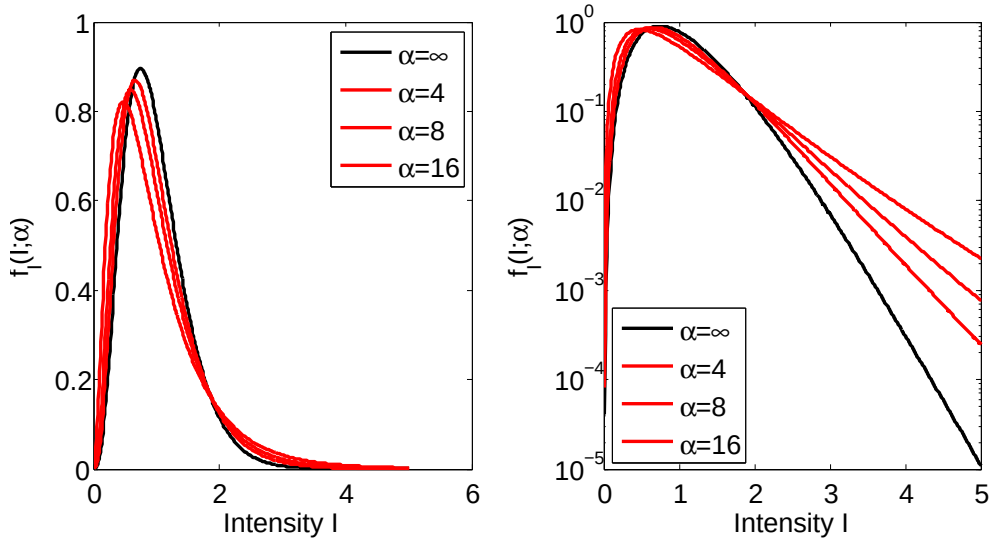

Figure 4.6: PDF $f_I(I;\alpha)$ of the $\mathcal{K}$ distributed multilook intensity I shown for shape parameters $\alpha = \{4, 6, 8, \infty\}$ at linear (left) and logarithmic (right) scale. The asymptotic case $\alpha = \infty$ is equivalent to gamma distributed intensity, corresponding to a Wishart distibuted polarimetric covariance matrix $\mathbf{C}$. The other parameters are $L=4$ and $\langle I \rangle = 1$.

4.4.1 Matrix-Variate $\mathcal{K}$ Distribution

The matrix-variate $\mathcal{K}$ distribution was derived in [Lee et al., 1994]. The name originates from the Bessel K function, which appears in the PDF expression. It results from a Mellin convolution of the scaled complex Wishart distribution with the normalised (unit mean) gamma distribution. A $\mathcal{K}$ distributed covariance matrix is denoted $\mathbf{C} \sim \mathcal{K}(L, \mathbf{\Sigma}, \alpha)$, where the parameters are inherited from the speckle and texture distribution.

Probability density function:

$$f_{\mathbf{C}}(\mathbf{C}; L, \mathbf{\Sigma}, \alpha) = \frac{|\mathbf{C}|^{L-d}}{|\mathbf{\Sigma}|^{L}} \frac{2(L\alpha)^{\frac{\alpha+Ld}{2}}}{\Gamma_d(L)\Gamma(\alpha)} \big(\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\big)^{\frac{\alpha-Ld}{2}} K_{\alpha-Ld}\big(2\sqrt{L\alpha \operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})}\big) . \quad (4.39)$$

Mellin kind characteristic function:

$$\phi_{\mathbf{C}}(s; L, \mathbf{\Sigma}, \alpha) = \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} \frac{\Gamma(\alpha + d(s-d))}{\Gamma(\alpha)} \left(\frac{|\mathbf{\Sigma}|}{(\alpha L)^d}\right)^{s-d} . \quad (4.40)$$

Mellin kind cumulant generating function:

$$\varphi_{\mathbf{C}}(s; L, \mathbf{\Sigma}, \alpha) = \ln \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} + \ln \frac{\Gamma(\alpha + d(s-d))}{\Gamma(\alpha)} + (s-d)(\ln|\mathbf{\Sigma}| - d\ln(\alpha L)) . \quad (4.41)$$

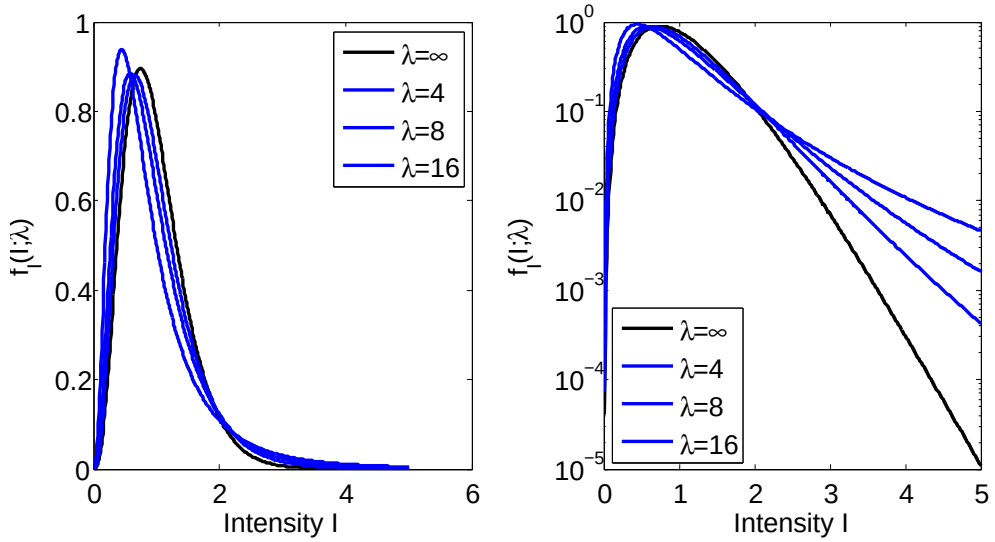

Figure 4.7: PDF $f_I(I;\lambda)$ of the $\mathcal{G}^0$ distributed multilook intensity I shown for shape parameters $\lambda = \{4, 6, 8, \infty\}$ at linear (left) and logarithmic (right) scale. The asymptotic case $\lambda = \infty$ is equivalent to gamma distributed intensity, corresponding to a Wishart distibuted polarimetric covariance matrix $\mathbf{C}$. The other parameters are $L=4$ and $\langle I \rangle = 1$.

Matrix log-cumulants:

$$\kappa_\nu\{\mathbf{C}; L, \mathbf{\Sigma}, \alpha\} = \begin{cases} \psi_d^{(0)}(L) + \ln|\mathbf{\Sigma}| + d\Big(\psi^{(0)}(\alpha) - \ln(\alpha L)\Big) & \text{for } \nu = 1\,, \\ \psi_d^{(\nu-1)}(L) + d^\nu \psi^{(\nu-1)}(\alpha) & \text{for } \nu > 1\,. \end{cases} \tag{4.42}$$

Matrix-variate distributions are difficult to visualise, and the solution is to plot instead the the marginal density of the single polarisation intensities found as diagonal elements in the polarimetric covariance matrix. Figure 4.6 shows examples of univariate $\mathcal{K}$ distributions of intensity for different values of the texture parameter α.

4.4.2 Matrix-Variate $\mathcal{G}^0$ Distribution

The matrix-variate $\mathcal{G}^0$ distribution was derived in [Freitas et al., 2005]. It results from a Mellin convolution of the scaled complex Wishart distribution with the normalised inverse gamma distribution. A $\mathcal{G}^0$ distributed covariance matrix is denoted $\mathbf{C} \sim \mathcal{G}^0(L, \mathbf{\Sigma}, \lambda)$, where the parameters are inherited from the speckle and texture distribution.

Probability density function:

$$f_{\mathbf{C}}(\mathbf{C}; L, \mathbf{\Sigma}) = \frac{L^{Ld}}{\Gamma_d(L)} \frac{|\mathbf{C}|^{L-d}}{|\mathbf{\Sigma}|^L} \frac{\Gamma(Ld+\lambda)(\lambda-1)^\lambda}{\Gamma(\lambda)} \left(L\,\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}) + \lambda - 1\right)^{-\lambda-Ld}\,. \tag{4.43}$$

Mellin kind characteristic function:

$$\phi_{\mathbf{C}}(s; L, \boldsymbol{\Sigma}, \lambda) = \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} \frac{\Gamma(\lambda - d(s-d))}{\Gamma(\lambda)} \left(|\boldsymbol{\Sigma}| \left(\frac{\lambda-1}{L} \right)^d \right)^{s-d} . \tag{4.44}$$

Mellin kind cumulant generating function:

$$\begin{aligned} \varphi_{\mathbf{C}}(s; L, \boldsymbol{\Sigma}, \lambda) = \ln \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} + \ln \frac{\Gamma(\lambda - d(s-d))}{\Gamma(\lambda)} \\ + (s-d) \Big(\ln |\boldsymbol{\Sigma}| + d(\ln(\lambda-1) - \ln L) \Big) . \end{aligned} \tag{4.45}$$

Matrix log-cumulants:

$$\kappa_\nu\{\mathbf{C}; L, \boldsymbol{\Sigma}, \lambda\} = \begin{cases} \psi_d^{(0)}(L) + \ln |\boldsymbol{\Sigma}| + d \Big(\ln \left(\frac{\lambda-1}{L} \right) - \psi^{(0)}(\lambda) \Big) & \text{for } \nu = 1 , \\ \psi_d^{(\nu-1)}(L) + (-d)^\nu \psi^{(\nu-1)}(\lambda) & \text{for } \nu > 1 . \end{cases} \tag{4.46}$$

Figure 4.7 shows examples of univariate $\mathcal{G}^0$ distributions of intensity for different values of the texture parameter λ.

4.4.3 Matrix-Variate $\mathcal{U}$ Distribution

The matrix-variate $\mathcal{U}$ distribution was derived in [Bombrun and Beaulieu, 2008]. The name originates from the Kummer U function, which appears in the PDF expression. It results from a Mellin convolution of the scaled complex Wishart distribution with the normalised Fisher-Snedecor distribution. A $\mathcal{U}$ distributed covariance matrix is denoted $\mathbf{C} \sim \mathcal{U}(L, \boldsymbol{\Sigma}, \alpha, \lambda)$, where the parameters are inherited from the speckle and texture distribution.

Probability density function:

$$\begin{aligned} f_{\mathbf{C}}(\mathbf{C}; L, \boldsymbol{\Sigma}, \alpha, \lambda) = \frac{L^{Ld}}{\Gamma_d(L)} \frac{|\mathbf{C}|^{L-d}}{|\boldsymbol{\Sigma}|^L} \frac{\Gamma(\alpha+\lambda)\Gamma(Ld+\lambda)}{\Gamma(\alpha)\Gamma(\lambda)} \left(\frac{\alpha}{\lambda-1} \right) \\ \times U \left(Ld+\lambda,\ Ld-\alpha+1,\ L \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C}) \frac{\alpha}{(\lambda-1)} \right) . \end{aligned} \tag{4.47}$$

Mellin kind characteristic function:

$$\begin{aligned} \phi_{\mathbf{C}}(s; L, \boldsymbol{\Sigma}, \alpha, \lambda) = \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} \frac{\Gamma(\alpha + d(s-d))}{\Gamma(\alpha)} \frac{\Gamma(\lambda - d(s-d))}{\Gamma(\lambda)} \\ \times \left(|\boldsymbol{\Sigma}| \left(\frac{\lambda-1}{\alpha L} \right)^d \right)^{s-d} . \end{aligned} \tag{4.48}$$

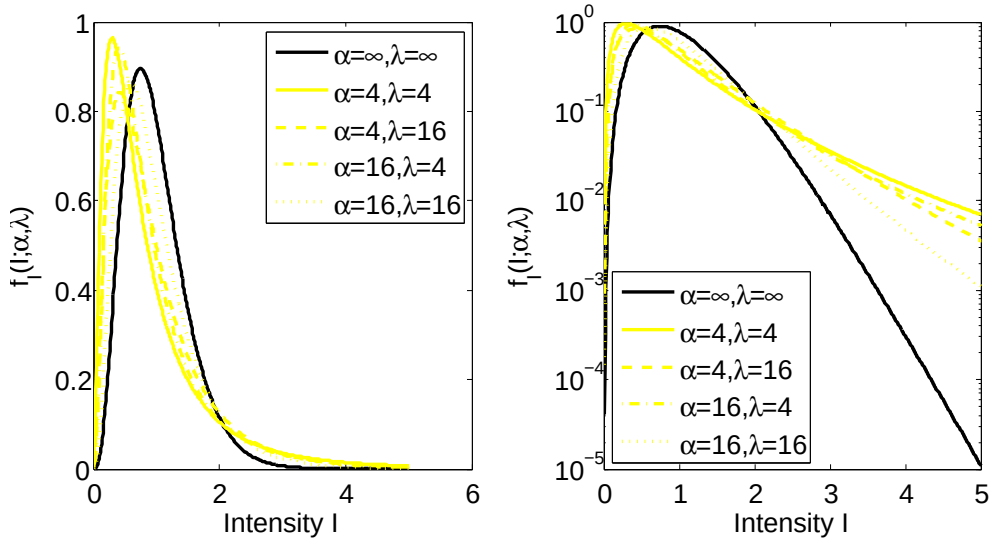

Figure 4.8: PDF $f_I(I;\lambda)$ of the $\mathcal{U}$ distributed multilook intensity I shown for different pairs of shape parameters, (α,λ), at linear (left) and logarithmic (right) scale. The asymptotic case ($\alpha = \infty$,$\lambda = \infty$) is equivalent to gamma distributed I, corresponding to a scaled Wishart distributed $\mathbf{C}$. The asymptotic cases $\alpha \to \infty$ and $\lambda \to \infty$ correspond to $\mathcal{K}$ and $\mathcal{G}^0$ distributed I and $\mathbf{C}$, respectively. The other parameters are $L = 4$ and $\langle I \rangle = 1$.

Mellin kind cumulant generating function:

$$\begin{aligned}\varphi_{\mathbf{C}}(s;L,\boldsymbol{\Sigma},\alpha,\lambda) = \ln\frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} + \ln\frac{\Gamma(\alpha+d(s-d))}{\Gamma(\alpha)} + \ln\frac{\Gamma(\lambda-d(s-d))}{\Gamma(\lambda)} \\ + (s-d)\Big(\ln|\boldsymbol{\Sigma}| + d(\ln(\lambda-1) - \ln\alpha - \ln L)\Big)\,.\end{aligned} \tag{4.49}$$

Matrix log-cumulants:

$$\begin{aligned}\kappa_1\{\mathbf{C};L,\boldsymbol{\Sigma},\alpha,\lambda\} &= \psi_d^{(0)}(L) + \ln|\boldsymbol{\Sigma}| + d\left(\psi^{(0)}(\alpha) - \psi^{(0)}(\lambda) + \ln\left(\frac{\lambda-1}{\alpha L}\right)\right)\,, \\ \kappa_{\nu>1}\{\mathbf{C};L,\boldsymbol{\Sigma},\alpha,\lambda\} &= \psi_d^{(\nu-1)}(L) + d^{\nu}\Big(\psi^{(\nu-1)}(\alpha) + (-1)^{\nu}\psi^{(\nu-1)}(\lambda)\Big)\,.\end{aligned} \tag{4.50}$$

Figure 4.8 shows examples of univariate $\mathcal{U}$ distributions of intensity for different pairs of the texture parameters α and λ.

4.4.4 Matrix-Variate $\mathcal{W}$ Distribution

The PDF of the matrix-variate $\mathcal{W}$ distribution has not yet been derived, but it is found that it will contain a special function known as the Whittaker $\mathcal{W}$ function, hence the proposed name. The distribution results from a Mellin convolution of the scaled complex

Wishart distribution with the normalised beta distribution, and the Mellin statistics are easily deduced from the existing results. A $\mathcal{W}$ distributed covariance matrix is denoted $\mathbf{C} \sim \mathcal{W}(L, \mathbf{\Sigma}, \alpha, \lambda)$, where the parameters are inherited from the speckle and texture distribution.

Mellin kind characteristic function:

$$\phi_{\mathbf{C}}(s; L, \mathbf{\Sigma}, \alpha, \lambda) = \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} \frac{\Gamma(\alpha + d(s-d))}{\Gamma(\alpha)} \frac{\Gamma(\lambda)}{\Gamma(\lambda + d(s-d))} \left(|\mathbf{\Sigma}| \left(\frac{\lambda}{\alpha L} \right)^d \right)^{s-d}. \tag{4.51}$$

Mellin kind cumulant generating function:

$$\begin{aligned} \varphi_{\mathbf{C}}(s; L, \mathbf{\Sigma}, \alpha, \lambda) = \ln \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} + \ln \frac{\Gamma(\alpha + d(s-d))}{\Gamma(\alpha)} + \ln \frac{\Gamma(\lambda)}{\Gamma(\lambda + d(s-d))} \\ + (s-d) \Big(\ln |\mathbf{\Sigma}| + d(\ln \lambda - \ln \alpha - \ln L) \Big). \end{aligned} \tag{4.52}$$

Matrix log-cumulants:

$$\begin{aligned} \kappa_1\{\mathbf{C}; L, \mathbf{\Sigma}, \alpha, \lambda\} &= \psi_d^{(0)}(L) + \ln |\mathbf{\Sigma}| + d \left(\psi^{(0)}(\alpha) - \psi^{(0)}(\lambda) + \ln \left(\frac{\lambda}{\alpha L} \right) \right) \\ \kappa_{\nu>1}\{\mathbf{C}; L, \mathbf{\Sigma}, \alpha, \lambda\} &= \psi_d^{(\nu-1)}(L) + d^{\nu} \Big(\psi^{(\nu-1)}(\alpha) - \psi^{(\nu-1)}(\lambda) \Big). \end{aligned} \tag{4.53}$$

4.4.5 Matrix-Variate $\mathcal{M}$ Distribution

The PDF of the matrix-variate $\mathcal{M}$ distribution has not yet been derived, but it is found that it will contain a special function known as the Whittaker $\mathcal{M}$ function, hence the proposed name. The distribution results from a Mellin convolution of the scaled complex Wishart distribution with the normalised inverse beta distribution, and the Mellin statistics are easily deduced from the existing results. A $\mathcal{M}$ distributed covariance matrix is denoted $\mathbf{C} \sim \mathcal{M}(L, \mathbf{\Sigma}, \alpha, \lambda)$, where the parameters are inherited from the speckle and texture distribution.

Mellin kind characteristic function:

$$\phi_{\mathbf{C}}(s; L, \mathbf{\Sigma}, \alpha, \lambda) = \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} \frac{\Gamma(\alpha - d(s-d))}{\Gamma(\alpha)} \frac{\Gamma(\lambda)}{\Gamma(\lambda - d(s-d))} \left(|\mathbf{\Sigma}| \left(\frac{(\alpha-1)}{(\lambda-1)L} \right)^d \right)^{s-d}. \tag{4.54}$$

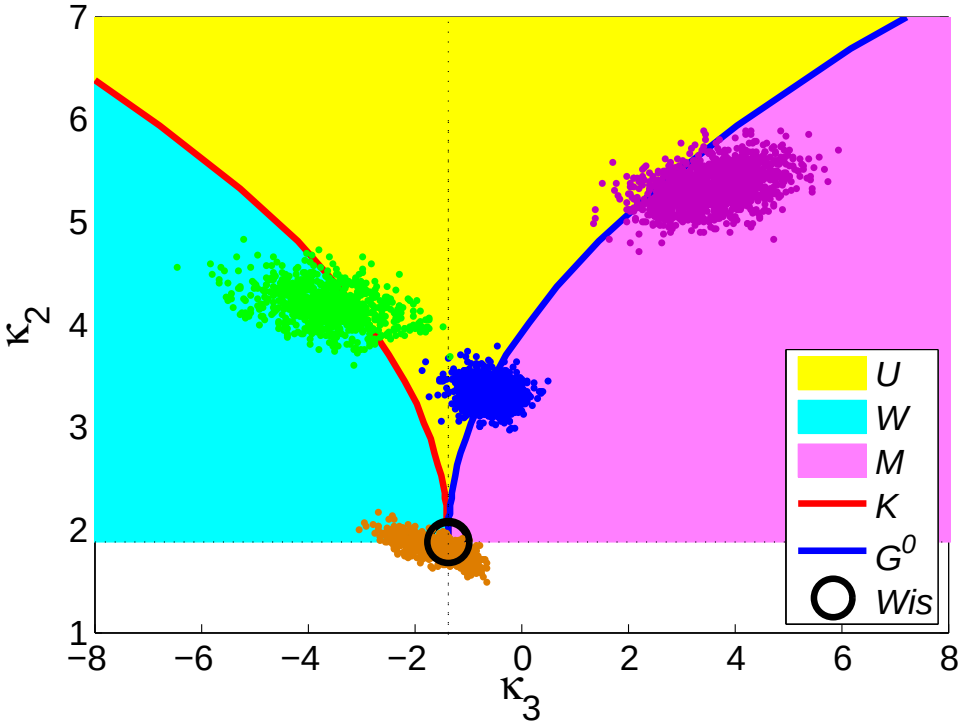

Figure 4.9: The matrix log-cumulant diagram

Mellin kind cumulant generating function:

$$\varphi_{\mathbf{C}}(s; L, \boldsymbol{\Sigma}, \alpha, \lambda) = \ln \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} + \ln \frac{\Gamma(\alpha - d(s-d))}{\Gamma(\alpha)} + \ln \frac{\Gamma(\lambda)}{\Gamma(\lambda - d(s-d))} + (s-d)\Big(\ln|\boldsymbol{\Sigma}| + d(\ln(\alpha-1) - \ln(\lambda-1) - \ln L)\Big). \tag{4.55}$$

Matrix log-cumulants:

$$\begin{aligned} \kappa_1\{\mathbf{C}; L, \boldsymbol{\Sigma}, \alpha, \lambda\} &= \psi_d^{(0)}(L) + \ln|\boldsymbol{\Sigma}| + d\left(\psi^{(0)}(\lambda) - \psi^{(0)}(\alpha) + \ln\left(\frac{(\alpha-1)}{(\lambda-1)L}\right)\right) \\ \kappa_{\nu>1}\{\mathbf{C}; L, \boldsymbol{\Sigma}, \alpha, \lambda\} &= \psi_d^{(\nu-1)}(L) + (-d)^{\nu}\Big(\psi^{(\nu-1)}(\alpha) - \psi^{(\nu-1)}(\lambda)\Big). \end{aligned} \tag{4.56}$$

4.5 The Matrix Log-Cumulant Diagram

We now introduce the matrix log-cumulant diagram, which is shown in Figure 4.9. This is a visualisation tool which allows us to both characterise the different matrix distributions and compare data to the distribution models. The diagram promotes intuition about the applications of the Mellin kind statistics framework, and is used in Papers 2

and 3 (Chapters 6 and 7) to provide geometrical interpretations of the proposed estimation procedures and goodness-of-fit tests.

In the diagram we plot: (i) manifolds spanned by the theoretical matrix log-cumulants that can be attained under given models, and (ii) points that represent the empirical sample matrix log-cumulants computed from data samples. The dimension of a given manifold equals the number of parameters in the texture distribution associated with that particular model. The Wishart distribution has no texture parameters, and is represented by a black point (zero-dimensional manifold). The matrix-variate $\mathcal{K}$ and $\mathcal{G}^0$ distributions have one texture parameter, and are represented by the red and blue curves (one-dimensional manifolds), respectively. The $\mathcal{U}$, $\mathcal{W}$ and $\mathcal{M}$ distributions have two texture parameters, and are visualised in the diagram as the respective yellow, cyan and magenta surfaces.

The diagram is spanned by the second and third-order matrix log-cumulants, and is a direct extension of the log-cumulants proposed in [Nicolas, 2002,Nicolas, 2006] for the univariate case. It was shown in (4.38) that under the multilook polarimetric product model, matrix log-cumulants of order $\nu \geq 2$ are independent of the speckle distribution scale matrix $\boldsymbol{\Sigma}$. They depend only on L, which is considered a constant, and the texture parameters. By plotting the third-order matrix log-cumulant against the second-order matrix log-cumulant (a convention introduced by Nicolas), we obtain a diagram which shows the solitary impact of the texture parameters upon the models. Thus, it promotes insight about how we can select between the different compound matrix distributions (including the Wishart distribution). We can also assess the fit between data and models by looking the distances between them in matrix log-cumulant space. Finally, the diagram visualises how texture parameters are estimated from data by projecting sample matrix log-cumulants onto the manifolds of population matrix log-cumulants.

The clusters of sample matrix log-cumulants plotted in Figure 4.9 represent targets that have been selected from different images acquired by the airborne AIRSAR L-band sensor previously operated by NASA/JPL. The samples represent forest (green cluster), sea (blue cluster), urban area (purple cluster) and cropland (brown cluster), and indicate which model is suitable in each case. The multiple points in each cluster were obtained by bootstrap sampling the covariance matrix samples. This is a way of showing the variance of the sample matrix log-cumulants. Note also that the matrix distributions presented in this chapter cover the whole matrix log-cumulant plane, remembering that the lower semi-plane represents texture distributions with negative variances.

Chapter 5

Paper 1:
Estimation of the Equivalent Number of Looks in Polarimetric Synthetic Aperture Radar Imagery

Estimation of the Equivalent Number of Looks in Polarimetric Synthetic Aperture Radar Imagery

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

***Abstract*—This paper addresses estimation of the equivalent number of looks (ENL), an important parameter in statistical modelling of multilook synthetic aperture radar (SAR) images. Two new ENL estimators are discovered by looking at certain moments of the multilook polarimetric covariance matrix, which is commonly used to represent multilook polarimetric SAR data, and assuming that the covariance matrix is complex Wishart distributed. Firstly, a second-order trace moment provides a polarimetric extension of the ENL definition, and also a matrix-variate version of the conventional ENL estimator. The second estimator is obtained from the log-determinant matrix moment, and is also shown to be the maximum likelihood estimator under the Wishart model. It proves to have much lower variance than any other known ENL estimator, whether applied to single polarisation or polarimetric SAR data. Moreover, this estimator is less affected by texture, and thus provides more accurate results than other estimators, should the assumption of Gaussian statistics for the complex scattering coefficients be violated. These are the first known estimators to use the full covariance matrix as input, rather than individual intensity channels, and therefore to utilise all the statistical information available. We finally demonstrate how an ENL estimate can be computed automatically from the empirical density of small sample estimates calculated over a whole scene. We show that this method is more robust than procedures where the estimate is calculated in a manually selected region of interest.**

***Index Terms*—Radar polarimetry, synthetic aperture radar, parameter estimation, moment methods, unsupervised learning**

I. Introduction

THE equivalent (or effective) number of looks (ENL) is a parameter of multilook synthetic aperture radar (SAR) images, which describes the degree of averaging applied to the SAR measurements during data formation and postprocessing. Multilooking is performed in order to mitigate the noiselike effect of interference, known as speckle, which is characteristic of all coherent imaging systems. In this process, correlated measurements are averaged, which complicates statistical modelling of the resulting multilook data. The pragmatic solution is to model the output as an average of independent measurements, and to replace the actual number of correlated samples by an equivalent number of independent ones, that is, the ENL. The ENL estimate is the parameter value that produces a best match between empirical moments of the correlated data and theoretical moments of the data model, which assumes independency. The ENL is generally a noninteger number.

The processing task normally referred to as multilooking is performed in the frequency domain. It is part of the range/azimuth compression leading up to a focused SAR image [1], [2]. Multiple measurements are obtained by splitting the synthetic aperture Doppler bandwidth into a number of subbands, each giving rise to a separate image referred to as a look. All looks are averaged in the power domain to produce multilook data. However, multilooking can also be done as postprocessing, that is, after a well focused image is generated. This method requires that the data are available in single-look complex (SLC) format. Averaging is then performed in the spatial domain. In addition to reducing speckle, both approaches to multilooking reduce image resolution, and hence the amount of data, an effect which is sometimes desired to ease the computational burden.

Being a distribution parameter, the ENL has influence on the accuracy of the information extracted by methods based upon statistical modelling of multilook SAR data. For instance, the ENL is necessary input to important classification and

This is a slightly modified version of the paper published in IEEE Transactions on Geoscience and Remote Sensing, Vol. 47, No. 11, Novermber 2009.

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

change detection algorithms for PolSAR data. The discriminant function of the popular Wishart classifier [3], [4] avoids dependency upon the ENL by the restrictive assumption of equiprobable classes. For nontrivial choices of prior probability, Bayesian classifiers based on the Wishart distribution [5] or more sophisticated data models [6], [7] require an estimate of the ENL. So does the change detection algorithm derived from the generalised likelihood ratio of two unknown Wishart distributed matrices [8].

The ENL is commonly estimated by identifying homogeneous regions in an image, where the speckle is fully developed and contribution of texture is negligible, meaning that the radar cross section is assumed to be constant. These conditions assure that the distribution of the scattering coefficients can be assumed complex Gaussian [1]. Under this statistical model, the ENL can be estimated from simple image statistics. A reliable ENL estimate can be obtained for a given sensor and fixed data processing scheme by manually selecting appropriate calibration targets, and such a value is sometimes provided as part of the image metadata. However, a processing chain with selectable algorithms and processing parametres will clearly benefit from having a robust and automatic estimation method, but such methods are difficult to design due to the required identification of homogeneous regions. Underestimation of the ENL occurs in the presence of texture and other sources of inhomogeneity.

The ENL and the conventional ENL estimator have been defined for the case of single polarisation SAR, as described in [1], [2]. For PolSAR data, the ENL has traditionally been estimated separately for each polarimetric channel, and then averaged, as in [6], [9]. In the following, we will develop a general theory for fully polarimetric SAR data, for which ENL estimation from single polarisation images becomes a special case. The objective of this work is twofold: We want to extend the theory of ENL estimation to the polarimetric case, where estimates are derived explicitly from matrix-variate statistics. We next want to design a fully automatic estimation procedure that requires no parameter selection or manual intervention, such as selection of homogeneous regions where image statistics are to be calculated.

The paper is structured as follows: Sec. II introduces SAR polarimetry, with different data formats and their distribution models. Sec. III presents the traditional definition of the ENL and reviews the literature of known estimators. In Sec. IV we present certain moment expressions for the Wishart distribution, and use them to derive new ENL estimators for PolSAR data. The contents of Sec. V are related to performance evaluation. We derive a lower bound to the variance of the ENL estimator, closely related to the Cramér-Rao bound, and further present a statistical model, which will be used to assess robustness to texture. The discussion of robustness is particularly relevant to unsupervised estimation, which is the topic of Sec. VI. We here propose an estimation procedure that is fully automatic. Sec. VII presents results of experiments with synthetic and real data. In Sec. VIII we give our conclusions.

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. For simplicity, we have not distinguished between random variables and instances of random variables, as such can be ascertained through context.

II. Statistical Modelling of PolSAR Data

The full-polarimetric SAR instrument separately transmits orthogonally polarised microwaves pulses, and measures orthogonal components of the received signal. For each pixel, the measurements result in a matrix of scattering coefficients. These are complex-valued, dimensionless numbers that describe the transformation of the transmitted (incoming) electromagnetic (EM) field to the received (backscattered) EM field for all combinations of transmit and receive polarisation.

The transformation can be expressed as

$$\begin{bmatrix} E_h^r \\ E_v^r \end{bmatrix} = \frac{e^{jkr}}{r} \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \begin{bmatrix} E_h^t \\ E_v^t \end{bmatrix} \quad (1)$$

where k denotes wavenumber and r is the distance between radar and target. The subscript of the EM field components E_i^j denotes horizontal (h) or vertical (v) polarisation, which is the most common set of orthogonal polarisations, while the superscript indicates transmitted (t) or received (r) wave. The scattering coefficients S_{ij} are subscripted with the associated receive and transmit polarisation, in that

order. Together, they form the scattering matrix, denoted $\mathbf{S} = [S_{ij}]$.

The scattering matrix can be reduced to one of the vectors

$$\mathbf{s} = \begin{bmatrix} S_{hh} \\ (S_{hv} + S_{vh})/\sqrt{2} \\ S_{vv} \end{bmatrix} \tag{2}$$

or

$$\mathbf{k} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh} + S_{vv} \\ S_{hh} - S_{vv} \\ S_{hv} + S_{vh} \end{bmatrix} . \tag{3}$$

The lexicographic scattering vector, denoted $\mathbf{s}$, is the vectorised version of $\mathbf{S}$ after the cross-polarisation terms S_{hv} and S_{vh} have been averaged, assuming reciprocity of the target. The scaling with a factor $\sqrt{2}$ is done to preserve total power of the signal. The Pauli basis scattering vector, denoted $\mathbf{k}$, is a linear transformation of $\mathbf{s}$, which provides physical interpretations of its elements in terms of basic scattering mechanisms [4].

A. Gaussian Model

It is commonly assumed that the scattering vector elements are jointly circular complex Gaussian. This is strictly justified only for homogeneous regions of the image, characterised by fully developed speckle and no texture. The notion of texture describes spatial variation in the backscatter that is due to target variability, that is, fluctuations in the radar cross section. The Gaussian model only encompasses variability due to speckle.

The matrix $\mathbf{S}$ and the vectors $\mathbf{s}$ and $\mathbf{k}$ are SLC format representations of PolSAR data. Multilook PolSAR data is commonly represented by

$$\mathbf{C}_\mathbf{s} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{s}_i \mathbf{s}_i^H \text{ or } \mathbf{C}_\mathbf{k} = \frac{1}{L}\sum_{i=1}^{L} \mathbf{k}_i \mathbf{k}_i^H \tag{4}$$

known as the sample covariance matrix and coherency matrix, respectively. They are formed as the mean Hermitian outer product of the single-look scattering vectors $\{\mathbf{s}_i\}_{i=1}^L$ and $\{\mathbf{k}_i\}_{i=1}^L$, respectively, where L is the nominal number of looks. The superscript H means complex conjugate transpose. Assume that $\mathbf{s}$ (or $\mathbf{k}$) is zero mean and circular complex multivariate Gaussian, denoted $\mathbf{s} \sim \mathcal{N}_d^{\mathbb{C}}(\mathbf{0}, \mathbf{\Sigma}_\mathbf{s})$, where $\mathbf{0}$ is a column vector of zeros, d is the dimension of $\mathbf{s}$, and $\mathbf{\Sigma}_\mathbf{s} = \mathrm{E}\{\mathbf{s}\mathbf{s}^H\}$ is the covariance matrix of $\mathbf{s}$. The probability density function (pdf) of $\mathbf{s}$ is thus

$$p_\mathbf{s}(\mathbf{s}; \mathbf{\Sigma}_\mathbf{s}) = \frac{1}{\pi^d |\mathbf{\Sigma}_\mathbf{s}|} \exp(-\mathbf{s}^H \mathbf{\Sigma}_\mathbf{s}^{-1} \mathbf{s}) \tag{5}$$

where $|\cdot|$ is the determinant operator. It follows that if $L \geq d$ and the $\mathbf{s}_i$ (or $\mathbf{k}_i$) in (4) are independent, then the scaled covariance matrix, defined as $\mathbf{Z} = L\mathbf{C}_\mathbf{s}$ (or $\mathbf{Z} = L\mathbf{C}_\mathbf{k}$), follows the nonsingular complex Wishart distribution [10]:

$$p_\mathbf{Z}(\mathbf{Z}; L, \mathbf{\Sigma}) = \frac{|\mathbf{Z}|^{L-d}}{|\mathbf{\Sigma}|^L \Gamma_d(L)} \exp\left(-\operatorname{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{Z}\right)\right) \tag{6}$$

where $\operatorname{tr}(\cdot)$ is the trace operator and $\mathbf{\Sigma} = \mathrm{E}\{\mathbf{Z}\}/L = \mathrm{E}\{\mathbf{C}_\mathbf{s}\}$. We write this as $\mathbf{Z} \sim \mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma})$. The normalisation constant $\Gamma_d(L)$ is the multivariate Gamma function, defined as

$$\Gamma_d(L) = \pi^{d(d-1)/2} \prod_{i=0}^{d-1} \Gamma(L - i) \tag{7}$$

where $\Gamma(L)$ is the standard Euler gamma function.

B. Product Model

The randomness of a SAR measurement is mainly attributed to two unrelated factors, namely speckle and texture. The latter represents the natural spatial variation of the radar cross section, which is generally not perfectly homogeneous for pixels that are thematically mapped as one class. Whereas the Gaussian model only accounts for speckle, several statistical models exist that also incorporate texture, either by assuming statistics that imply a non-Gaussian scattering vector, or explicitly modelling texture as a separate random variable (rv). The latter case leads to a doubly stochastic model with a compounded distribution.

The well known product model, reviewed e.g. in [1], [11], has been shown to be both mathematically tractable and successful for modelling and prediction purposes. In the polarimetric version [12], it decomposes the scattering vector $\mathbf{z}$ (defined on a lexicographic or Pauli basis) as a product of two independent stochastic processes with individual distributions:

$$\mathbf{z} = \sqrt{\gamma}\mathbf{w} . \tag{8}$$

The first process, $\mathbf{w} \sim \mathcal{N}_d^{\mathbb{C}}(\mathbf{0}, \mathbf{\Sigma}_\mathbf{w})$, models speckle. The second process generates texture, represented by the scalar rv γ, under the assumption that the

texture is independent of polarisation. The multiplicative property of the model is preserved as data is transformed from single-look format to multilook format, provided the fluctuations in the radar cross section occur on a scale that is larger than or equal to the multilook cell. In the multilook covariance matrix domain the product model becomes

$$\mathbf{Z} = \gamma_L \mathbf{W} \tag{9}$$

where we have defined $\mathbf{W} = \sum_{i=1}^{L} \mathbf{w}_i \mathbf{w}_i^H \sim \mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma}_{\mathbf{w}})$ and $\mathbf{Z} = \sum_{i=1}^{L} \mathbf{z}_i \mathbf{z}_i^H$. The pdf of $\mathbf{Z}$ depends on the multilook texture rv γ_L, which is related, but not identical to γ.

The multilook polarimetric product model leading up to Eq. (9) is extensively reviewed in [13], where the family of generalised inverse Gaussian distributions is proposed as a model for γ, and implicitly also for γ_L. Selecting amongst several applicable members of this family, we shall assume that γ is gamma distributed, denoted $\gamma \sim \boldsymbol{\gamma}(\mu, \alpha)$, with unit mean ($\mu = \mathrm{E}\{\gamma\} = 1$) and shape parameter $\alpha = \mu^2 / \operatorname{Var}\{\gamma\} = 1 / \operatorname{Var}\{\gamma\}$. The pdf of $\gamma \sim \boldsymbol{\gamma}(1, \alpha)$ is

$$p_\gamma(\gamma; \alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \gamma^{\alpha-1} e^{-\alpha\gamma} \,. \tag{10}$$

Based upon the product model with $\gamma \sim \boldsymbol{\gamma}(1, \alpha)$, a family of distributions can be derived for the complex scattering coefficient, multilook detected amplitude, multilook intensity, and their polarimetric counterparts, referred to in common as K-distributions. The K-distribution for the polarimetric scattering vector $\mathbf{z}$ was derived in [12]. A K-distribution for the multilook polarimetric covariance matrix $\mathbf{Z}$ was first presented in [9]:

$$\begin{aligned} &p_{\mathbf{Z}}(\mathbf{Z}; L, \mathbf{\Sigma}, \alpha_L) \\ &\quad = \frac{2|\mathbf{Z}|^{L-d} \alpha_L^{\frac{\alpha_L + Ld}{2}}}{\Gamma_d(L) |\mathbf{\Sigma}|^L \Gamma(\alpha_L)} \big(\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{Z}) \big)^{\frac{\alpha_L - Ld}{2}} \\ &\qquad \times K_{\alpha_L - Ld} \left(2\sqrt{\alpha_L \operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{Z})} \right) . \end{aligned} \tag{11}$$

Here, $K_\nu(\cdot)$ is the modified Bessel function of the second kind with order ν. Further, α_L is a distribution parameter of $\gamma_L \sim \Gamma(1, \alpha_L)$, which also becomes a parameter of $p_{\mathbf{Z}}(\mathbf{Z})$. It was shown in [7] that

$$\alpha_L = \frac{Ld + 1}{d + 1} \alpha \tag{12}$$

assures consistency between the models of Eqs. (8) and (9) with respect to certain moment relations.

For interpretation purposes, we note that $\gamma_L \to 1$ and the multilook polarimetric K-distribution in Eq. (11) converges in distribution to the complex Wishart distribution in Eq. (6) as $\alpha_L \to \infty$. Thus, high values of α_L imply little texture, whereas low values refer to significant texture and non-Gaussianity.

In the following sections, we use the Wishart distribution $p_{\mathbf{Z}}(\mathbf{Z}; \mathrm{L_e}, \mathbf{\Sigma})$ as the underlying model when deriving ENL estimators. The multilook polarimetric K-distribution $p_{\mathbf{Z}}(\mathbf{Z}; \mathrm{L_e}, \mathbf{\Sigma}, \alpha_L)$ is used to investigate how deviation from the Wishart model, in terms of texture, affects the performance of the proposed estimators. It is possible to derive an ENL estimator from the multilook polarimetric K-distribution, but this introduces α_L as an additional nuisance parameter to be estimated. Therefore, we will not pursue this approach.

C. Modelling Correlated Data

In the derivation of the distributions in Eqs. (6) and (11) it was assumed that the single-look scattering vectors used to form the multilook polarimetric covariance matrices are independent. This assumption does not hold, as discussed in Sec. I. An exact analytic expression for the pdf of $\mathbf{Z}$ that accounts for correlation of the $\mathbf{z}$ samples has, to the best of our knowledge, not been obtained, and the derivation is regarded as intractable (see e.g. [14]). The practical solution for distribution modelling of correlated data has been to maintain the functional form of Eqs. (6) and (11), but to replace the number of correlated looks, L, with an equivalent number of uncorrelated looks, L_e, that makes certain moment relations (to be defined in Eqs. (15) and (17)) of the theoretical model consistent with empirical moments.

III. Known Estimators

A. Coefficient of Variation Estimator

The traditional approach to ENL estimation for single polarisation SAR data has been to manually select a homogeneous image region, where the assumptions of fully developed speckle and no texture assure that the scattering coefficient is circular complex Gaussian. A single polarisation multilook intensity I, which is found as a diagonal entry of $\mathbf{Z}$, will then be distributed as $\boldsymbol{\gamma}(\sigma, L)$:

$$p_I(I; \sigma, L) = \frac{1}{\Gamma(L)} \left(\frac{L}{\sigma} \right)^L I^{L-1} e^{-LI/\sigma} \tag{13}$$

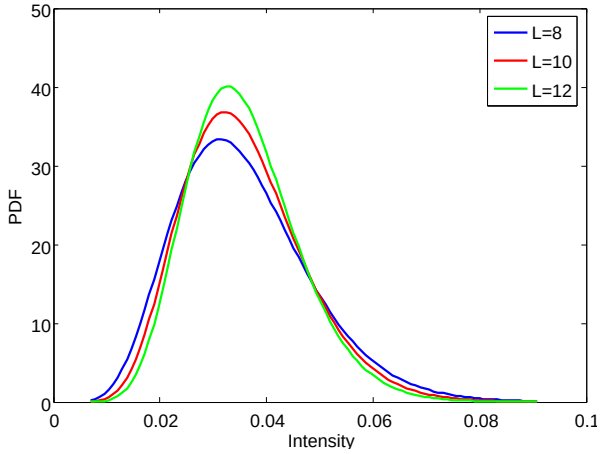

Fig. 1. Example of gamma distribution $\gamma(\sigma, L)$ parametrised with mean intensity $\sigma = 0.0358$ and number of looks $L = \{8, 10, 12\}$.

with the mean intensity σ and the number of looks L as parameters of the gamma distribution. Fig. 1 illustrates how the intensity distribution is affected by a varying number of looks.

The k-th order moment of I is given by [1]

$$\mathrm{E}\{I^k\} = \frac{\Gamma(L+k)}{\Gamma(L)} \left(\frac{\sigma}{L}\right)^k \tag{14}$$

assuming uncorrelated data. We specifically find that $\mathrm{E}\{I\} = \sigma$ and $\mathrm{Var}\{I\} = \sigma^2/L$, thus $\mathrm{E}\{I\}^2/\mathrm{Var}\{I\} = L$. This does not hold for correlated data, but in this case L can then be replaced by the ENL, defined as

$$\mathrm{L_e} = \frac{E\{I\}^2}{\mathrm{Var}\{I\}} \,. \tag{15}$$

The right hand side of Eq. (15) defines the entity known as the coefficient of variation (CV). The traditional ENL estimator, which arises as

$$\hat{L}_e^{(CV)} = \frac{\langle I \rangle^2}{\langle I^2 \rangle - \langle I \rangle^2} \tag{16}$$

is therefore named the CV estimator. Here, $\langle \cdot \rangle$ denotes sample average. This estimator has a simple form and is easy to apply, which may explain the very limited interest in ENL estimation found in the literature, both for single polarisation and polarimetric SAR data. We have not discovered any known methods that are tailored for PolSAR data, in the sense that they process the full covariance or coherency matrices in (4), thereby utilising all available statistical information. Methods designed for mono-polarised SAR are used to handle both cases, as we describe below.

B. Fractional Moment-Based Estimator

Though Eq. (15) is commonly referred to as the definition of the ENL [1], [2], there are other ways to solve for L from statistics of the given model, that may also be used to determine $\mathrm{L_e}$. An alternative estimator was suggested in [6], based upon the same distribution model, but using a fractional moment (FM) of the multilook intensity.

From (14) we have

$$\mathrm{E}\{I^{1/2}\} = \frac{\Gamma(L+\frac{1}{2})}{\Gamma(L)} \sqrt{\frac{\sigma}{L}} \,. \tag{17}$$

Replacing $\mathrm{E}\{I^{1/2}\}$ and σ with the estimates $\langle I^{1/2} \rangle$ and $\langle I \rangle$, we obtain the equation

$$f\big(\hat{L}_e^{(FM)}\big) = \frac{\Gamma\big(\hat{L}_e^{(FM)} + \frac{1}{2}\big)}{\Gamma\big(\hat{L}_e^{(FM)}\big)\sqrt{\hat{L}_e^{(FM)}}} \sqrt{\langle I \rangle} - \langle \sqrt{I} \rangle = 0 \tag{18}$$

which must be solved numerically for the fractional moment estimate, denoted $\hat{L}_e^{(\mathrm{FM})}$. The existance of a root of $f\big(\hat{L}_e^{(FM)}\big)$ is proved in Appendix A.

Frery et al. [6] used this method on polarimetric SAR data, estimating the ENL separately for each polarisation, and then averaging the results. We note that Eq. (18) differs from the definition found in [6], which contains some errors.

C. Other Approaches

The earliest publications we have found that are dedicated to ENL estimation, are two papers by Lee et al. that propose to estimate the pair $\mathrm{E}\{I\}^2$ and $\sigma_I^2 = \mathrm{Var}\{I\}$ in small windows over the whole image. One may then infer L_e from the scatter plot of paired $(\mathrm{E}\{I\}, \sigma_I)$ estimates, using the Hough transform [15] or an angular sweep method [16] to determine the best fitted line, whose inverse slope squared is the $\mathrm{L_e}$ estimate. A refined method is proposed by Foucher et al. [17], which is based on nonparametric estimation of the distribution of $\sigma_I/E\{I\}$ values using orthogonal Laguerre functions, and estimation of $\mathrm{L_e}$ from the mode of the distribution. The focus of all these papers are on unsupervised estimation omitting selection of a region of interest.

Further approches include the papers of Gierull et al. [14] and Joughin et al. [18], where the authors derive ENL estimators for mono-polarised SAR data

from the distribution of interferometric phase. This is a more specialised application, which requires multiple baseline data, and is therefore outside our scope. Moreover, a general estimator can be applied also to interferometric data. We finally mention the ENL estimators proposed by El Zaart et al. [19]. They are derived from the gamma distribution using maximum likelihood theory, in a manner similar to how we will subsequently derive matrix-variate methods for polarimetric data from the Wishart distribution. A shortened version of this paper was presented in [20].

IV. New Estimators

So far, we have not been able to find any ENL estimators in the literature that use the full sample covariance or coherency matrix, or any other matrix-variate statistic, as input. We have therefore tried to derive moment based estimators founded on the Wishart distribution.

A. Trace Moment-Based Estimator

Assume that the random matrix $\mathbf{Z}$ is positive semidefinite and complex Wishart distributed with L degrees of freedom and scale matrix $\mathbf{\Sigma} = \mathrm{E}\{\mathbf{Z}\}/L$. The degrees of freedom are equivalent to the number of looks, and the Wishart law is denoted $\mathbf{Z} \sim \mathcal{W}_C(L, \mathbf{\Sigma})$. The following moments of $\mathbf{Z}$ are derived in [21]:

$$\mathrm{E}\{\mathrm{tr}(\mathbf{ZZ})\} = L^2\,\mathrm{tr}(\mathbf{\Sigma\Sigma}) + L\,\mathrm{tr}(\mathbf{\Sigma})^2\,. \tag{19}$$

$$\mathrm{E}\{\mathrm{tr}(\mathbf{Z})^2\} = L^2\,\mathrm{tr}(\mathbf{\Sigma})^2 + L\,\mathrm{tr}(\mathbf{\Sigma\Sigma})\,. \tag{20}$$

These expressions lead to respective estimators for L_e:

$$\hat{L}_e^{(1)} = \frac{\mathrm{tr}(\mathbf{\Sigma})^2}{\langle \mathrm{tr}(\mathbf{CC})\rangle - \mathrm{tr}(\mathbf{\Sigma\Sigma})} \tag{21}$$

$$\hat{L}_e^{(2)} = \frac{\mathrm{tr}(\mathbf{\Sigma\Sigma})}{\langle \mathrm{tr}(\mathbf{C})^2\rangle - \mathrm{tr}(\mathbf{\Sigma})^2} \tag{22}$$

now expressed in terms of $\mathbf{C} = \mathbf{Z}/L$, which is the supplied PolSAR data format. Out of these two estimators, we prefer the former, i.e., the estimator that originates from the second-order trace moment in (19). This is because it uses all the elements of $\mathbf{C}$ and thus all polarimetric information through $\mathrm{tr}(\mathbf{CC})$, whereas the latter uses $\mathrm{tr}(\mathbf{C})$, which only contains the intensities on the diagonal. The variance of $\hat{L}_e^{(1)}$ is also observed experimentally to be superior to that of $\hat{L}_e^{(2)}$(22). We further note that in the single polarisation case, both Eqs. (21) and (22) reduce to Eq. (16). We have thus found two matrix-variate extensions of the conventional ENL estimator, and denote the preferred estimator in Eq. (21) by $\hat{L}_e^{(\mathrm{TM})}$, where *TM* is short for trace moment.

B. Log-Determinant Moment-Based Estimator

We next turn to some other moment relations involving the determinant of a complex Wishart matrix, and the logarithm thereof. For the normalised determinant of a complex Wishart matrix, we have [22]

$$\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|} \sim \prod_{i=0}^{d-1} \frac{1}{2}\chi^2_{2(L-i)} \tag{23}$$

where d is the dimension of $\mathbf{Z}$ (or, equivalently, the number of polarimetric channels). That is, $|\mathbf{Z}|/|\mathbf{\Sigma}|$ is distributed like a product of chi-square distributed variables, scaled by the factor $1/2$, and with different degrees of freedom, as denoted by the subscript of χ^2_i. The moments of $|\mathbf{Z}|/|\mathbf{\Sigma}|$ were found in [23] for real Wishart matrices. In the complex case, we follow the same procedure to obtain

$$\mathrm{E}\left\{\left(\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|}\right)^r\right\} = \prod_{i=0}^{d-1} \frac{\Gamma(L-i+r)}{\Gamma(L-i)}\,. \tag{24}$$

To find the moments of $\ln(|\mathbf{Z}|/|\mathbf{\Sigma}|)$, we note that the moment generating function of $\ln(|\mathbf{Z}|/|\mathbf{\Sigma}|)$ is defined as

$$M_{\ln\left(\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|}\right)}(r) = \mathrm{E}\left\{\exp\left(r\ln\left(\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|}\right)\right)\right\} \tag{25}$$

which is identical to the left hand side of the expression in Eq. (24). It follows that

$$\begin{aligned}\mathrm{E}&\left\{\left[\ln\left(\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|}\right)\right]^k\right\} \\ &= \left[\frac{d^k}{dr^k}\prod_{i=0}^{d-1}\frac{\Gamma(L-i+r)}{\Gamma(L-i)}\right]\Bigg|_{r=0}\,.\end{aligned} \tag{26}$$

The first-order moment becomes

$$\mathrm{E}\left\{\ln\left(\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|}\right)\right\} = \sum_{i=0}^{d-1}\Psi^{(0)}(L-i) \tag{27}$$

where $\Psi^{(0)}(L) = \Gamma'(L)/\Gamma(L)$ is known as the digamma function, which is one of Euler's polygamma functions, defined as

$$\begin{aligned}\Psi^{(m)}(z) &= \frac{d^{m+1}}{dL^{m+1}} \ln \Gamma(L) \\ &= (-1)^m \int_0^\infty \frac{t^m e^{-zt}}{1-e^{-t}}\, dt\ .\end{aligned} \tag{28}$$

The derivation of Eq. (27) is shown in Appendix B. Since data is supplied as covariance matrices in the format $\mathbf{C} = \mathbf{Z}/L$, we use $\ln|\mathbf{Z}| = \ln|\mathbf{C}| + d\ln L$ to write

$$\mathrm{E}\left\{\ln|\mathbf{C}|\right\} = \ln|\mathbf{\Sigma}| + \sum_{i=0}^{d-1} \Psi^{(0)}(L-i) - d\ln L\ . \tag{29}$$

This equation defines our new estimator. The estimate, denoted $\hat{L}_e^{(ML)}$ for reasons explained below, is the root of

$$\begin{aligned} g\big(\hat{L}_e^{(ML)}\big) &= \langle \ln|\mathbf{C}|\rangle - \ln|\langle \mathbf{C}\rangle| \\ &\quad - \sum_{i=0}^{d-1} \Psi^{(0)}\big(\hat{L}_e^{(ML)} - i\big) + d\ln \hat{L}_e^{(ML)} = 0 \end{aligned} \tag{30}$$

where the mathematical expectation $\mathrm{E}\{\ln|\mathbf{C}|\}$ has been replaced by the empirical mean $\langle\ln|\mathbf{C}|\rangle$ and $\mathbf{\Sigma}$ by $\langle\mathbf{C}\rangle$. Eq. (30) must be solved numerically in the same fashion as the estimator defined by (18). The existance of a unique root of $g(L)$ is proved in Appendix A.

From the complex Wishart distribution in Eq. (6), it is easy to verify that

$$\begin{aligned} &\frac{\partial}{\partial L} \ln p_{\mathbf{Z}}(\mathbf{Z}; L, \mathbf{\Sigma}) \\ &\qquad = \ln|\mathbf{Z}| - \ln|\mathbf{\Sigma}| - \frac{\partial}{\partial L}\ln\Gamma_d(L) \\ &\qquad = \ln\left(\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|}\right) - \sum_{i=0}^{d-1}\Psi^{(0)}(L-i)\,. \end{aligned} \tag{31}$$

By comparison of Eq. (31) with Eq. (27), it is revealed that the solution of Eq. (30) is the maximum likelihood (ML) estimate of $\mathrm{L_e}$. It is thus asympotically unbiased, efficient, and Gaussian [24].

We finally remark that efficient implementation of the sum of polygamma functions is aided by the recurrence relation:

$$\Psi^{(m)}(z+1) = \Psi^{(m)}(z) + (-1)^m m!\, z^{-(m+1)}\ . \tag{32}$$

V. Performance Evaluation

The obvious way of evaluating estimator performance is by looking at statistical properties such as bias and (co)variance. Let $\boldsymbol{\theta} = [\mathrm{L_e}, \Sigma_{11}, \Sigma_{21}, \ldots, \Sigma_{dd}]^T = [\mathrm{L_e}, \mathrm{vec}(\mathbf{\Sigma})^T]^T$ be the complex-valued parameter vector of the Wishart model, with the vectorisation (column stacking) operator denoted as $\mathrm{vec}(\cdot)$, and let $\hat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta}$. The length of $\boldsymbol{\theta}$ is $k = d^2 + 1$.

When estimating $\mathrm{L_e}$, the entries of the covariance matrix, denoted $\Sigma_{ij}, i, j \in \{1, \ldots, d\}$, become nuisance parameters whose uncertainty degrade the estimate of $\mathrm{L_e}$. The estimators defined by Eqs. (18), (21), and (30) are too complicated to find analytic expressions for neither the distribution, the bias vector, nor the covariance matrix of $\hat{\boldsymbol{\theta}}$. The bias vector and the covariance matrix are defined as

$$\mathbf{b}(\hat{\boldsymbol{\theta}}) = \mathrm{E}\{\hat{\boldsymbol{\theta}}\} - \boldsymbol{\theta} \tag{33}$$

and

$$\mathrm{Cov}\{\hat{\boldsymbol{\theta}}\} = \mathrm{E}\{(\hat{\boldsymbol{\theta}} - \mathrm{E}\{\hat{\boldsymbol{\theta}}\})(\hat{\boldsymbol{\theta}} - \mathrm{E}\{\hat{\boldsymbol{\theta}}\})^H\} \tag{34}$$

respectively. However, we can evaluate both bias and covariance empirically, for instance using bootstrap methods. We are also able to establish a lower bound on the variance of $\mathrm{L_e}$.

A. A Bound on the Variance of ENL Estimators

Assume that we have a set $\mathscr{Z} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_N\}$ of N independent and complex Wishart distributed sample covariance matrices. The log-likelihood function of $\mathscr{Z}$ is

$$\begin{aligned} \mathcal{L}(\mathscr{Z}; \mathrm{L_e}, \mathbf{\Sigma}) &= \ln \prod_{i=1}^{N} p_{\mathbf{Z}}(\mathbf{Z}_i; \mathrm{L_e}, \mathbf{\Sigma}) \\ &= \sum_{i=1}^{N} \ln p_{\mathbf{Z}}(\mathbf{Z}_i; \mathrm{L_e}, \mathbf{\Sigma}) \end{aligned} \tag{35}$$

with $p_{\mathbf{Z}}(\mathbf{Z}; L, \mathbf{\Sigma})$ given by Eq. (6). The Cramér-Rao bound (CRB) establishes a lower bound on the covariance of the stochastic $\hat{\boldsymbol{\theta}}$. For the complex parameter vector $\boldsymbol{\theta}$, the CRB is defined as [25]

$$\begin{aligned} &\mathrm{Cov}\{\hat{\boldsymbol{\theta}}\} \\ &\succeq \frac{\partial}{\partial \boldsymbol{\theta}^T}\big(\boldsymbol{\theta} + \mathbf{b}(\hat{\boldsymbol{\theta}})\big)\, \mathbf{J}^{-1} \left(\frac{\partial}{\partial \boldsymbol{\theta}^T}\big(\boldsymbol{\theta} + \mathbf{b}(\hat{\boldsymbol{\theta}})\big)\right)^H \\ &= \frac{\partial\, \mathrm{E}\{\hat{\boldsymbol{\theta}}\}}{\partial \boldsymbol{\theta}^T}\, \mathbf{J}^{-1} \left(\frac{\partial\, \mathrm{E}\{\hat{\boldsymbol{\theta}}\}}{\partial \boldsymbol{\theta}^T}\right)^H \end{aligned} \tag{36}$$

where the matrix expression $\mathbf{A} \succeq \mathbf{B}$ denotes that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. Further, $\mathbf{J}$ denotes the Fisher information matrix (FIM), given by

$$\begin{aligned}\mathbf{J} &= \mathrm{E}\left\{\left(\frac{\partial \mathcal{L}(\mathscr{Z})}{\partial \boldsymbol{\theta}^T}\right)^H \frac{\partial \mathcal{L}(\mathscr{Z})}{\partial \boldsymbol{\theta}^T}\right\} \\ &= -\mathrm{E}\left\{\frac{\partial}{\partial \boldsymbol{\theta}^*}\left(\frac{\partial}{\partial \boldsymbol{\theta}}\mathcal{L}(\mathscr{Z})\right)^T\right\}\end{aligned} \tag{37}$$

for the complex case, where superscript $*$ denotes complex conjugation. The parameters of $\mathcal{L}(\mathscr{Z})$ are suppressed for brevity. The first equality of Eq. (37) is proven in [25], and the proof of the second is straight-forward by analogy with the real case [24].

If the estimator $\hat{\boldsymbol{\theta}}$ is unbiased, then Eq. (36) becomes the familiar $\mathrm{Cov}\{\hat{\boldsymbol{\theta}}\} \succeq \mathbf{J}^{-1}$. However, the estimators that we study are biased, and since we are not able to evaluate the term $\partial\,\mathrm{E}\{\hat{\boldsymbol{\theta}}\}/\partial\boldsymbol{\theta}^T$ in Eq. (36), the true CRB cannot be determined analytically. Still, by noting that

$$\frac{\partial\,\mathrm{E}\{\hat{\boldsymbol{\theta}}\}}{\partial \boldsymbol{\theta}^T} = \frac{\partial\big(\boldsymbol{\theta} + \mathbf{b}(\hat{\boldsymbol{\theta}})\big)}{\partial \boldsymbol{\theta}^T} = \mathbf{I}_k + \frac{\partial \mathbf{b}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^T} \tag{38}$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix, Eq. (36) is rewritten as

$$\begin{aligned}&\mathrm{Cov}\{\hat{\boldsymbol{\theta}}\} \\ &\quad\succeq \mathbf{J}^{-1} + \frac{\partial \mathbf{b}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^T}\mathbf{J}^{-1} + \mathbf{J}^{-1}\left(\frac{\partial \mathbf{b}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^T}\right)^H \\ &\qquad + \frac{\partial \mathbf{b}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^T}\mathbf{J}^{-1}\left(\frac{\partial \mathbf{b}(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^T}\right)^H \\ &\quad= \mathbf{J}^{-1} + (\mathbf{K} + \mathbf{K}^H) + \mathbf{K}\mathbf{J}\mathbf{K}^H\,.\end{aligned} \tag{39}$$

We have here defined $\mathbf{K} = (\partial \mathbf{b}(\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}^T)\mathbf{J}^{-1}$. It is easily shown that the term $\mathbf{K}\mathbf{J}\mathbf{K}^H$ on the right hand side is positive semidefinite, and thus contributes to a tighter bound on $\mathrm{Cov}\{\hat{\boldsymbol{\theta}}\}$. However, this cannot be proven for $(\mathbf{K} + \mathbf{K}^H)$, and the relationship between the inverse FIM, $\mathbf{J}^{-1}$, and the true CRB remains undefined. Still, $\mathbf{J}^{-1}$ is the best indication we can obtain of a performance bound, and we shall refer to it as the unbiased CRB (UCRB).

The inverse FIM is given by

$$\mathbf{J}^{-1} = \frac{1}{N}\begin{bmatrix} \displaystyle\sum_{i=0}^{d-1} \Psi^{(1)}(\mathrm{L_e} - i) & \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^T \\ \mathrm{vec}(\boldsymbol{\Sigma}^{-1})^* & \mathrm{L_e}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \end{bmatrix}^{-1} \tag{40}$$

where $\otimes$ denotes the Kronecker product. The derivation is shown in Appendix C. The bound on the variance of the ENL estimator thus becomes

$$\mathrm{Var}\{\hat{L}_e\} \geq J_{11}^{-1} \tag{41}$$

where J_{ij}^{-1} denotes element (i, j) of $\mathbf{J}^{-1}$. Eq. (41) must be evaluated numerically, but we see that the variance bound depends on the true $\mathrm{L_e}$ and $\boldsymbol{\Sigma}$, and that the rate of convergence is $1/N$.

B. Robustness to Texture

The concept of texture in SAR images was discussed in Sec. II-B. The product model was also introduced as a scheme to develop statistical models that accommodate texture, and thereby provide more flexible and accurate descriptions of PolSAR data than the Wishart distribution, which has been assumed in the derivation of all estimators so far. The multilook polarimetric K-distribution in (11) was presented as a concrete candidate for modelling of texture modulated covariance matrix data.

Models that include and quantify texture become relevant when we want to investigate the influence of texture on the ENL estimation performance. The textural variability of the target will add to the randomness inflicted by the measurement process through speckle. Consider multilook intensity data for simplicity: It is evident that the presence of texture will increase $\mathrm{Var}\{I\}$, when compared to the variance produced by speckle alone. This leads to underestimation of $\mathrm{L_e}$, as seen from (15).

To assess the effect of texture on different ENL estimators, we would ideally evaluate the mean and bias of the candidate estimators under a distribution that includes texture. Due to the complexity of both the estimators and the aspiring distributions, this is not possible. We must therefore resort to generating textured data, and use them to evaluate statistics of the estimators experimentally. The multilook polarimetric K-distribution is used for this purpose.

We note that, assuming the texture in all polarimetric channels can be modelled by a scalar random variable, the phase difference, amplitude ratio, and intensity ratio are all insensitive to texture. This was pointed out by Lee et al. [9], who used the fact to estimate the ENL, without explicitly stating how. The invariance of the amplitude ratio and also the phase to texture can be used to design robust ENL estimators. This approach has not been examined,

due to the complicated distribution of the phase difference, amplitude ratio, and intensity ratio [26], [27].

VI. Unsupervised Estimation

Some attempts have already been made to design a fully automatic estimation algorithm that avoids manual selection of a region of interest [15]– [17], as reviewed in Sec. III. Since all these methods are related to the traditional CV estimator in Eq. (16), and use only one polarisation at the time, we here propose a new unsupervised estimator based upon the polarimetric ML estimator defined by Eq. (30).

For an arbitrary SAR scene, there is no guarantee that we can find an image subset with fully developed speckle and no texture. If such a region exists, it may not contain enough samples to ensure that empirical moments can be calculated with sufficient accuracy. This motivates a different approach, where moments are calculated and the estimator evaluated in small windows over the whole image. The ENL is then inferred from the distribution of small sample estimates. However, this method has a number of inherent problems that need to be considered. Some of the windows will contain a mixture of pixels from different classes, and some will contain texture. Both of these conditions lead to underestimation of the ENL. We may also encounter areas where the contribution of coherent scatterers makes the zero mean assumption on the scattering coefficients invalid. The nonzero mean will increase the average intensity, and thus leads to overestimation of the ENL. Finally, when small sample sizes are used, the bias of all the estimators studied is significant. This is demonstrated in the experiments.

The method used in [15], [16] is to produce a scatter plot of $\widehat{E\{I\}}$ versus $\widehat{\sigma}_I$ values estimated over a whole scene. The idea is that values computed under no texture and fully developed speckle will dominate the population of estimates. Hence, they will stand out as a linear feature, such that the ENL can be inferred from the slope. Instead of performing line extraction in a two-dimensional space of empirical moments, we follow the approach of [17] and compute a single statistic, namely the ENL itself, hence producing a one-dimensional distribution of small sample ENL estimates. We use the same reasoning, hoping that a large enough proportion of the estimation windows satisfy the statistical assumptions. In this case, the overall distribution of estimates should be dominated by estimates computed from truly Wishart distributed samples, and the mode value can be used as an estimate of the ENL.

A. Nonparametric Estimation

The distribution of the ENL estimates will depend strongly upon the properties of the given image, that is, the homogeneity of the scene, the extent of the homogeneous regions, the amount of texture within the classes, and the presence of homogeneous regions that exhibit coherent scattering. Due to the unpredictable shape and possible multimodality of the distribution, we must resort to nonparametric estimation, and propose to use a kernel density estimator (KDE) implemented with the Epanechnikov kernel function [28], [29].

The KDE yields the following distribution estimate:

$$\hat{p}(\hat{\mathscr{L}}_e) = \frac{1}{nh}\sum_{i=1}^{n} K_h(\hat{\mathscr{L}}_e - \hat{\ell}_e(i)) \qquad (42)$$

where $\hat{\mathscr{L}}_e$ is the stochastic small sample ENL estimator, $\{\hat{\ell}_e(i)\}_{i=1}^n$ is a set of n instances produced by this estimator in separate windows, $K_h(\cdot)$ is the kernel function, and h is the kernel bandwidth that determines the degree of smoothing. The Epanechnikov kernel is defined as

$$K_h(x) = \frac{3}{4}\left(1 - \left(\frac{x}{h}\right)^2\right)\mathbb{1}_{\{|\frac{x}{h}|<1\}} \qquad (43)$$

where the indicator function $\mathbb{1}_{\{\Omega\}}$ denotes 1 when condition Ω holds, and 0 when it does not. The KDE is chosen because it is simple and has a convergence rate of $n^{-4/5}$, as compared to n^{-1}, which is common for parametric estimators.

The kernel bandwidth has a strong impact on the magnitude of the estimated distribution, but not so much on the sample mode. Since the aim is to extract the mode value and use it as an ENL estimate, determination of a near optimal bandwidth is not critical. We have therefore assumed that simple bandwidth selection rules from the literature (see e.g. [28], [29]) are sufficient. The Epanechnikov kernel is optimal with respect to the asympotic mean integrated squared error (AMISE) of the KDE. Equally important, it provides a fast implementation due to its finite support.

Let the final estimate extracted as the mode of the kernel density estimate be denoted $\hat{L}_e$.

B. Bias Correction

It will be shown in Sec. VII that one notable side effect of using small sample estimates of the ENL, is that they contain significant bias. This bias transfers directly to the value inferred from the distribution of estimates. An illustration is given in the result section. It is possible to estimate the bias by means of jackknife resampling, and the bias estimate can be used to obtain a corrected ENL estimate.

Jackknifing [30], [31] is a resampling technique that can be used to estimate the bias and variance in an estimator. If the original sample contains m observations, the jackknife procedure consists of recomputing the estimator m times, leaving out one observation from the full sample at a time. This produces m jackknife replications, $\{\hat{\ell}_e(i,j)\}_{j=1}^{m}$, for a given small sample estimate, $\hat{\ell}_e(i)$. The bias estimate based on sample window i is computed as

$$\hat{b}\big(\hat{\mathscr{L}}_e, i\big) = (m-1)\big(\hat{\ell}_e(i,\cdot) - \hat{\ell}_e(i)\big) \tag{44}$$

where $\hat{\ell}_e(i,\cdot)$ is the mean of the m jackknife replications, defined as

$$\hat{\ell}_e(i,\cdot) = \frac{1}{m}\sum_{j=1}^{m} \hat{\ell}_e(i,j)\,. \tag{45}$$

Bias estimation introduces considerable overhead to the algorithm, if we choose to compute a jackknife estimate $\hat{b}\big(\hat{\mathscr{L}}_e, i\big)$ for each of the n small sample windows in the image. We propose to process only a user specified number (or a percentage of the total number) of samples, selecting those that correspond to the estimates $\hat{\ell}_e(i)$ that are closest to the mode value, as these are most likely to comply with the statistical assumptions. This yields a collection of small sample bias estimates. The final estimate, $\hat{b}(\hat{L}_e)$, could have been obtained in the same manner as $\hat{L}_e$, i.e., by nonparametric density estimation and extraction of the mode. Instead, we suggest for simplicity to use the median value, which has proven experimentally to be consistently close to the mode value. The bias corrected ENL estimate thus becomes

$$\begin{aligned}\hat{L}'_e &= \hat{L}_e - \hat{b}(\hat{L}_e) \\ &= \arg\max_{\hat{\mathscr{L}}_e}\{\hat{p}(\hat{\mathscr{L}}_e)\} - \mathrm{med}\{\hat{b}\big(\hat{\mathscr{L}}_e, i\big)\}\end{aligned} \tag{46}$$

where $\mathrm{med}\{\cdot\}$ is the median operator.

One problem with the bias correction procedure is that the bias estimator itself has a bias. If the number of observations, m, becomes too small, then the correction is inaccurate. This must be taken into consideration when selecting the sample size m. The result section will indicate for which values of m a bias correction is needed and for which values a reliable correction can be obtained.

VII. Results

In the experiments we used synthetic and real data to compare the following algorithms:

1) Coefficient of varation (CV) estimator [(16)]
2) Fractional moment (FM) estimator [(18)]
3) Trace moment (TM) estimator [(21)]
4) Maximum likelihood (ML) estimator [(30)]

The CV is the conventional estimator, while the FM estimator [6] is a lesser known alternative from the literature. The TM estimator is our polarimetric generalisation of the CV estimator. The ML estimator is the proposed estimator based on first log-determinant moment of the multilook polarimetric covariance (or coherency) matrix.

A. Synthetic Data

We first tested the estimators on random generated data from a single class. The synthetic data set consisted of $N = 1{,}000{,}000$ coherency matrix samples drawn from a complex, circular, and zero mean Wishart distribution. The distribution was parametrised by a scale matrix $\mathbf{\Sigma}$ that had been computed by averaging a homogeneous region in the NASA/JPL AIRSAR L-band image of Flevoland, the Netherlands, and thus represented a realistic model of natural vegetation. The number of looks was set to $L = 10$.

1) Statistical Properties: From the population of $N = 1,000,000$ Wishart samples, we drew $M_b = 10,000$ bootstrap samples of variable size N_b, and then used the bootstrap estimator [30], [31], [32] to estimate the bias and variance of the ENL estimators. The upper panel of Fig. 2 displays the estimated bias versus sample size N_b, and ranks the ML estimator as the best, followed by the TM estimator, and then the FM estimator, which is slightly better than the CV estimator. The order of performance is the same for variance versus sample size N_b, as shown in the middle panel.

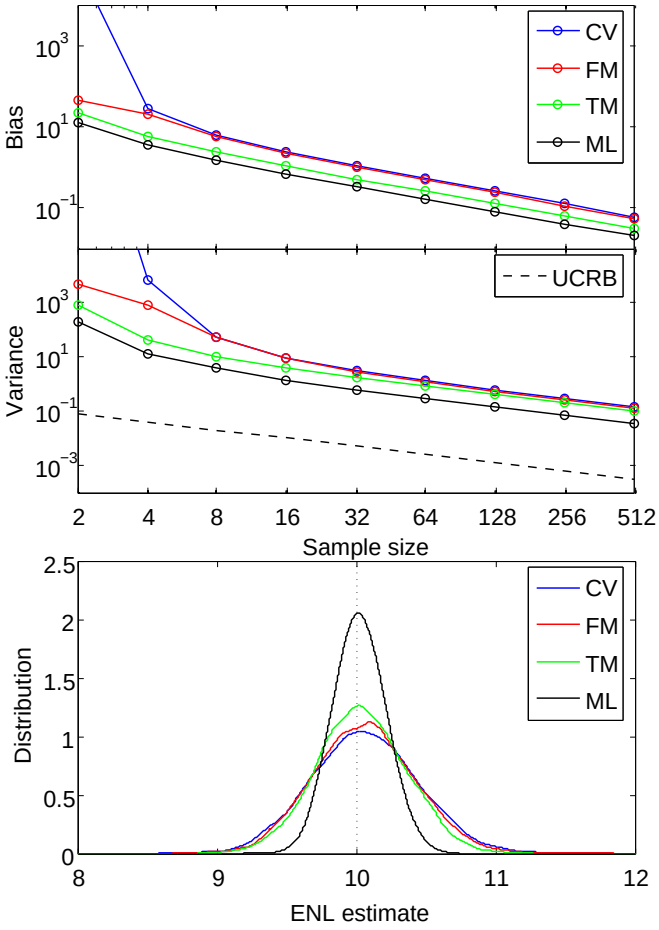

Fig. 2. Estimator bias (top) and variance (middle) as a function of sample size N_b, and the distribution of ENL estimates for $N_b = 512$ (bottom). Results shown for the CV, FM, TM, and ML estimator. The variance plot includes the unbiased Cramér-Rao Bound (UCRB). True $L = 10$ shown as dotted line.

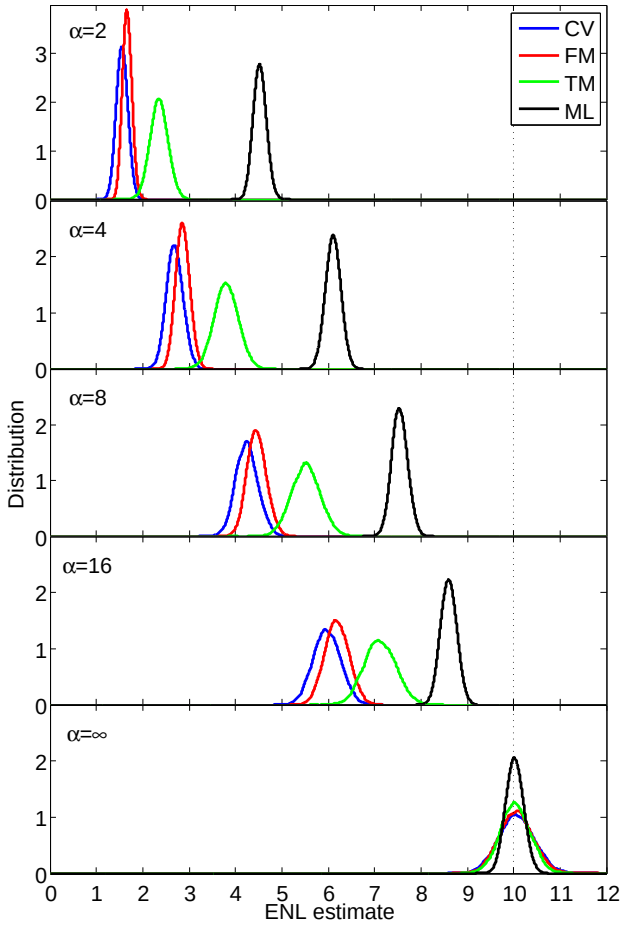

Fig. 3. Distribution estimates for the CV, FM, TM, and ML estimator, calculated from single class multilook polarimetric K-distributed data random generated with a fixed $\mathbf{\Sigma}$ and $\alpha = \{2, 4, 8, 16, \infty\}$.

The lower panel shows the distribution of ENL estimates for a fixed sample size of $N_b = 512$. The distribution was computed with a KDE estimator with Epanechnikov kernel and kernel bandwidth $h = 0.1$. We see that all estimators produce distributions that are centered approximately around the true number of looks, $L = 10$, as the random generated data had no correlation. We note that a considerable improvement in terms of reduced variance is visible for the ML estimator. Its variance is well above the UCRB (see the middle panel), but we have observed experimentally that much of the gap can be attributed to the nuisance parameters in $\mathbf{\Sigma}$.

2) Robustness to Texture: The experiments were repeated for multilook polarimetric K-distributed data with different degrees of texture, which increases with decreasing values of the distribution parameter α, as discussed in Sec. II-B. Data were generated with parameter values ranging from $\alpha = 2$, which corresponds to a strongly heterogeneous environment such as an urban area, to $\alpha = 16$, which may characterise vegetation such as forest or certain crops. The limiting case, $\alpha = \infty$, which is equivalent to no texture and Wishart distributed data, was also included.

Fig. 3 shows the distribution of the ENL estimators for different values of α with $L = 10$ (dotted line). The figure illustrates that the mode and mean of the distributions depend strongly on α, and that the estimate is severely distorted by texture. The ML estimator is least affected, followed by the TM estimator, with the FM estimator, and then the CV estimator as the inferior.

B. Real Data

1) Unsupervised Estimation: After having established the statistical properties of the ENL esti-

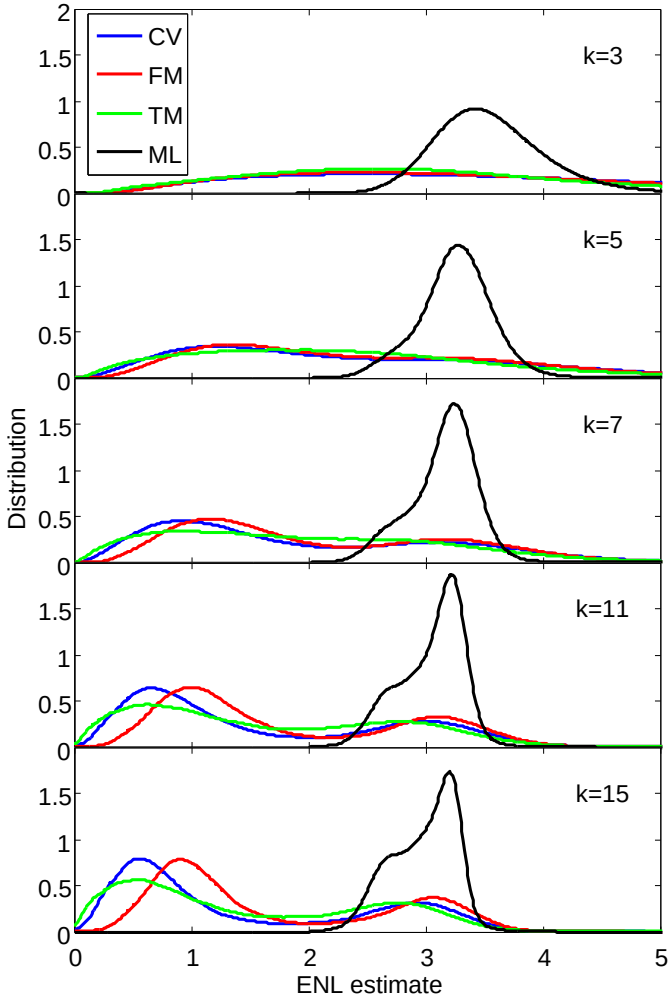

Fig. 4. Distribution estimates for the CV, FM, TM, and ML estimator calculated from the AIRSAR image of Flevoland. No speckle filter applied. ENL estimated for window sizes of $k = \{3, 5, 7, 11, 15\}$.

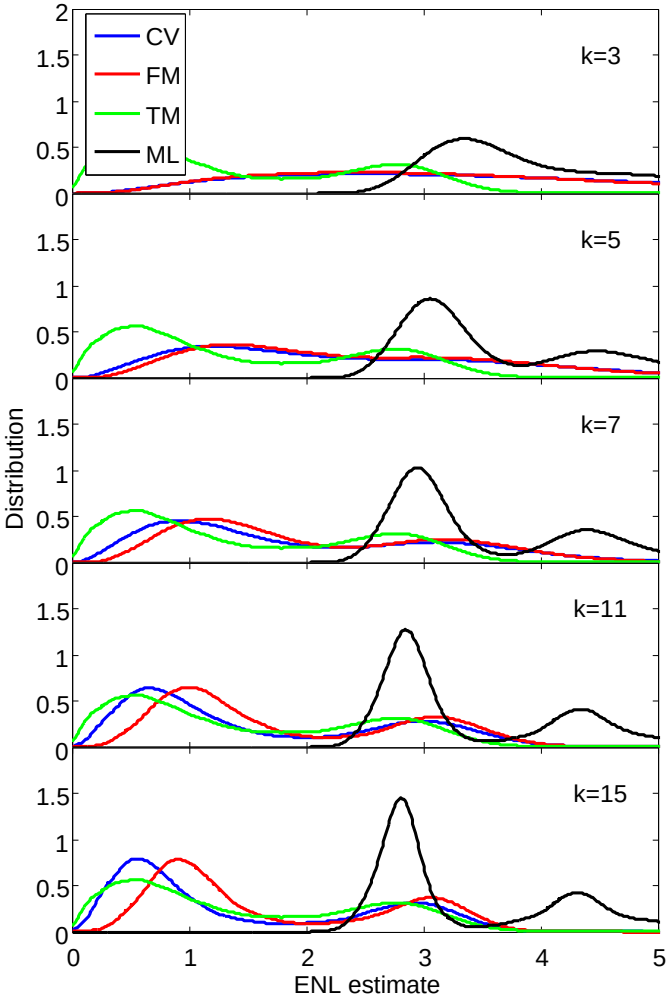

Fig. 5. Distribution estimates for the CV, FM, TM, and ML estimator calculated from the AIRSAR image of San Francisco. No speckle filter applied. ENL estimated for window sizes of $k = \{3, 5, 7, 11, 15\}$.

mators with synthetic data, we turned to real data for a realistic assessment of their applicability to unsupervised estimation. We chose to use two data sets acquired by the airborne NASA/JPL AIRSAR L-band instrument: one image of an agricultural area in Flevoland, The Netherlands, from 1989, and one image of the San Fransisco Bay area in California, USA, from 1988. Both data sets contain four-looked coherency matrices, with a pixel resolution of about $10\,\mathrm{m} \times 10\,\mathrm{m}$.

The landscape of the Flevoland image consists mainly of homogeneous fields, and also some forest areas, straight roads, and farm houses. The San Francisco Bay image contains mostly sea and urban areas, in addition to some parks and hills covered by vegetation. There are few homogeneous areas of considerable size, except for the ocean. One would therefore expect that it is relatively simpler to estimate the ENL from the Flevoland image.

Each image was processed by computing the estimators in a sliding window of size $k \times k$ pixels, covering the whole image. The window size was varied from $k = 3$ to $k = 15$. No speckle filter was applied initially. The distribution of each estimator was estimated from the collection of local estimates. We used a KDE with Epanechnikov kernel function and a kernel bandwidth of $h = 0.1$. The results are shown in Fig. 4 for the Flevoland image and Fig. 5 for the San Francisco image. A modified Lee filter [33] with window size $\ell = 7$ was then applied to the images to reduce the level of speckle. The results were similar, and are therefore only presented for the Flevoland image. The estimate distributions, shown in Fig. 6, were obtained with

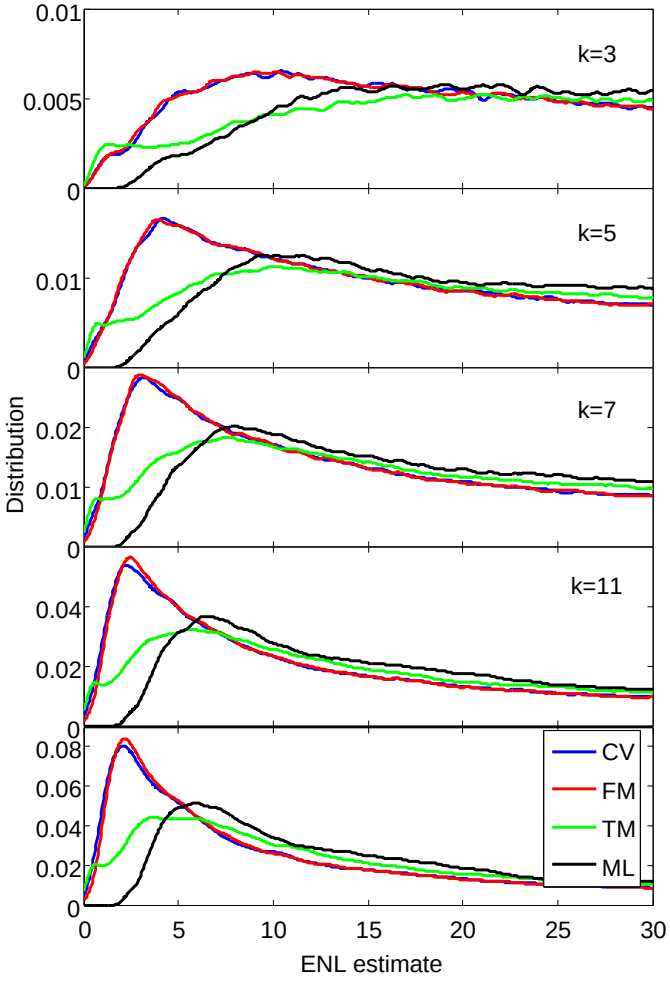

Fig. 6. Distribution estimates for the CV, TM, TM, and ML estimator calculated from the AIRSAR image of Flevoland. Modified Lee filter with window size 7×7 applied. ENL estimated for window sizes of $k = \{3, 5, 7, 11, 15\}$.

kernel bandwidth $h = 0.5$. None of the estimates were bias corrected at this stage.

2) Effect of window size: From the panels of Figs. 4 and 5, we can study the evolution of the distribution of ENL estimates as the window size increases. Denote by H_0 the hypothesis that the estimation sample is drawn from a homogeneous area with fully developed speckle and no texture, i.e., the statistical conditions assumed for ideal ENL estimation. Let H_1 be the complementary hypothesis, which indicates presence of multiple classes, texture, or coherent scattering. The overall distribution can then be modelled as a mixture:

$$f(\hat{L}_e) = \alpha_0 f_0(\hat{L}_e) + \alpha_1 f_1(\hat{L}_e)\,. \qquad (47)$$

The first mixture component, $f_0(\hat{L}_e)$, consists of estimates calculated under H_0, which occurs with relative frequency α_0. It is the desired component, and should ideally be sufficiently dominant to produce an identifiable mode close to the true ENL. The other component results from estimates produced under H_1. This component modifies the shape of the overall distribution and, depending on the magnitude of its relative frequency, $\alpha_1 = 1 - \alpha_0$, it may even give rise to additional modes.

Two expected effects can be seen as the number of samples within the estimation window increases: Firstly, the variance becomes lower and the modes narrower. This is most clearly observed for the ML estimator, which has a well-defined mode for all window sizes. Secondly, the probability of having mixed classes within the estimation window increases, and consequently, so does the proportion of underestimated values. This is seen as a growing negative skewness, and the tendency towards a bimodal distribution for all estimators. It also partially explains the shift of the mode value towards a lower ENL with increasing k, even though the bias of the ENL estimator also contributes to the observed effect.

Fig. 7 is a map of the locally estimated ENL values, obtained with the ML estimator and $k = 7$ for the Flevoland data set. It confirms that the mode in Fig. 4, centered around 3.2, corresponds to values that are estimated within homogeneous crop fields, while the mode emerging with increasing k around 2.7 relates to values estimated at class boundaries. In the same manner, Fig. 8 demonstrates for the San Francisco image that the main mode of the ML estimator with $k = 7$, located around 3.0, corresponds to values estimated over land. The second mode at higher values is discussed in the sequel. The fact that the mode value is more sensitive to k for the San Francisco data set, indicates that this image has less homogeneous regions on the scale of the estimation window.

From the investigations with synthetic data, it is obvious that the presence of texture will add to the underestimation, but this effect is not affected by the window size k. Following the discussion above, it seems reasonable to use the smallest window size possible to suppress the mixed class effect, while at the same time maintaining low enough variance and bias to obtain the required accuracy and precision in determination of the mode value.

3) Effect of Coherent Scatterers: Fig. 5 displays the influence of the window size, as discussed

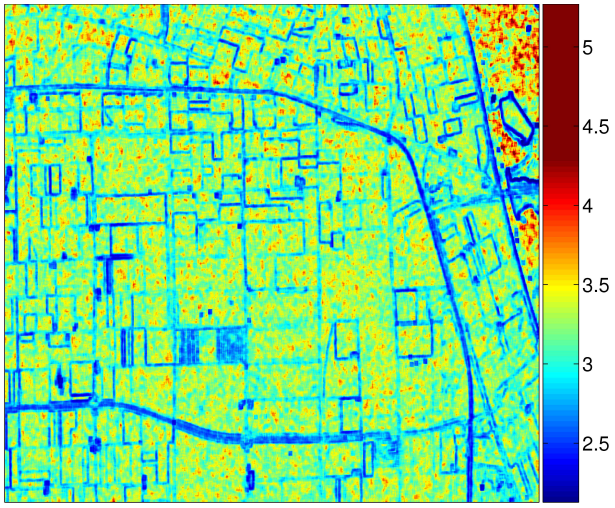

Fig. 7. Local ENL estimates obtained with the ML estimator and window size $k=7$ for the AIRSAR image of Flevoland. No speckle filter applied.

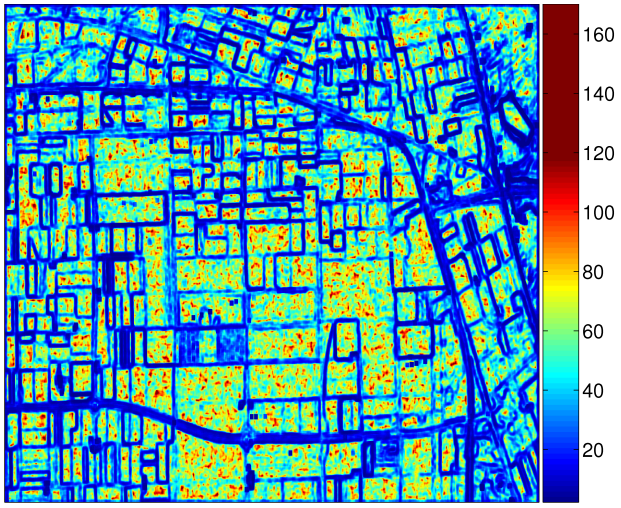

Fig. 9. Local ENL estimates obtained with the ML estimator and window size $k=7$ for the AIRSAR image of Flevoland. Modified Lee filter with window size $\ell=7$ applied.

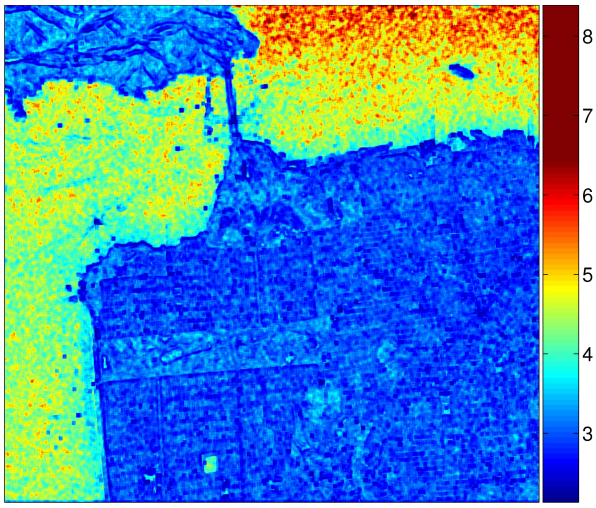

Fig. 8. Local ENL estimates obtained with the ML estimator and window size $k=7$ for the AIRSAR image of San Francisco. No speckle filter applied.

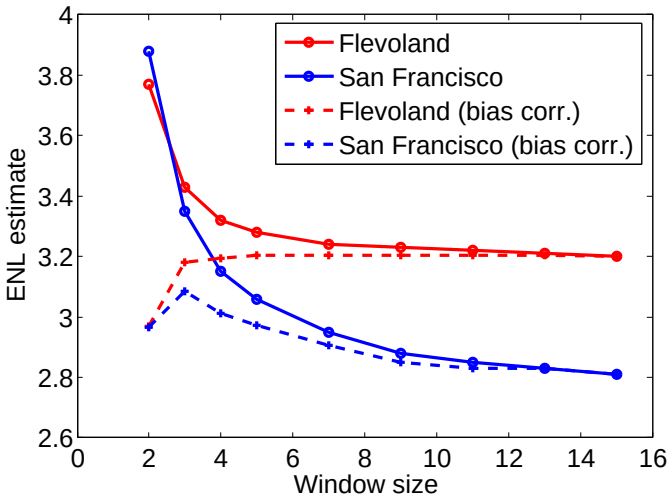

Fig. 10. ENL estimates obtained with the ML estimator as a function of window size k for the AIRSAR images of Flevoland and San Francisco, with and without bias correction. No speckle filter applied.

above, but also reveals another source of disturbance that only seems to affect the ML estimator. For increasing k we see the emergence of a second mode, which is located between 4 and 5, i.e., at values higher than the true number of looks ($L=4$). In Fig. 8, this cluster of estimates is observed to be spatially located over ocean, and the highest ENL estimates are obtained in the top right corner of the image, where the incidence angle reduces to five degrees. We believe that overestimation occurs because specular reflection from the water surface contributes a strong coherent component, which is consistent within local neighbourhoods. This makes the zero mean assumption on the scattering coefficients invalid.

The given explanation is mathematically consistent, although we have no firm evidence. The same phenomenon is observed for the Flevoland data set; Fig. 7 shows that the highest ENL estimates are found over water, i.e., in the triangular area in the top right corner. However, this image contains too little water surface for the overestimation effect to be clearly visible in the distribution of estimates (Fig. 4). Other regions, such as the large urban areas in the San Francisco image, are also expected to contain significant coherent scattering, but these are

too heterogeneous to produce overestimated ENL values.

4) Effect of Speckle Filtering: The unsupervised estimation procedure was also tested on speckle filtered data. The simplest speckle filter, a boxcar filter, smooths all pixels equally by averaging over a fixed size window, and thus acts like a spatial domain multilook operator. Hence, the effect on the ENL estimate is a simple scaling. More sophisticated filters perform adaptive smoothing. They take local variability in the image into account, in order to preserve details like edges and points. As a consequence, the ENL will no longer be a constant value, but a spatially varying number.

We applied a modified Lee filter [33], because of its widespread use. It is also simple enough that it allows us to quantify the amount of averaging it performs. The modified Lee filter is basically a linear minimum mean squared error (LMMSE) filter, whose output is a weighted sum of the centre pixel data value on the one hand and the average of a fixed size smoothing region selected from the filter window on the other. The weight is determined from the homogeneity of the smoothing region. Hence, the maximum smoothing factor is equal to the number of pixel of the smoothing region, denoted N_w, and the minimum is none. From the specification of the modified Lee filter, we have $N_w = \ell(\ell+1)/2$ when the full window size is $\ell \times \ell$, and the dynamic ENL after adaptive speckle filtering will lie in the range between L_e, the original ENL value, and $N_w \cdot L_e$. We see that a common window size of $\ell = 7$ yields $N_w = 28$, which illustrates that speckle filtering transforms a single-valued ENL into a wide range of values.

Fig. 6 shows the estimation results obtained on the Flevoland image processed with a modified Lee filter with $\ell = 7$. A mode becomes visible with increasing window size, but it occurs at very low ENL values. Fig. 9 is a map of the local estimates produced with the ML estimator. It illustrates that the mode emerging at $5 < \hat{L}_e < 10$ corresponds to estimates obtained over class boundaries. It can therefore not be related to the true ENL. The desired mode that appeared in Fig. 4 has vanished, as the distribution has been stretched due to the variable degree of smoothing. The areas that produced ENL estimates around the mode value of Fig. 4, now produce estimates in an interval ranging from 40 to 100. The same observations were made for the San Francisco image.

We acknowledge that other adaptive speckle filters will lead to different distributions of the ENL. Nevertheless, our observations strongly suggests that unsupervised ENL estimation is impossible for dynamically filtered data. This does not imply that our method has failed, but rather that the Wishart model, and in particular the parametrisation with a single-valued ENL, is inappropriate. The implications for statistical modelling should be addressed by future research.

5) Effect of estimator bias: The effect of the estimator bias is demonstrated in Fig. 10. The plot shows the mode value extracted by means of the KDE as a function of window size k. The respective estimates, $\hat{L}_e$ and $\hat{L}'_e$, obtained before and after bias correction are shown for both the Flevoland and the San Francisco data set. We observe for the Flevoland data that the bias corrected estimate is relatively constant from $k = 3$ and onwards. This indicates that the window size has no influence on the estimate after bias have been removed. The low value of the bias corrected estimate for $k = 2$ suggests that the bias is overestimated for low values of k. This naturally concerns both data sets. The ENL estimated from the San Francisco data shows a decreasing trend with k, also after bias correction. We interpret this as an effect of mixed classes, which increases with window size.

6) Estimation Results: The estimation results in Fig. 10 suggest that the data sets have different ENL values. This is not, however, supported by the knowledge that both images are produced with the same data processor, and that they have very similar ground resolution. The difference could stem from differences in acquisition parameters, but we believe it is more likely due to a differing amount of texture found in the respective images, and particularly in the areas where the estimates contributing to the main mode in the pdf estimates originate. The San Francisco image has very little homogeneous areas, and much of the estimates around the mode are collected from urban area and hilly terrain. With reference to the discussion of texture influence, illustrated by Fig. 3, this could well explain the lower ENL values extracted from the San Francisco data set. It is possible that also the ENL level estimated from the Flevoland data is lowered with respect to the true value by the presence of texture, but by a smaller amount.

It seems clear that the ability of our method to obtain an estimate that is close to the true ENL depends entirely on how susceptible the data set is to estimation. However, the alternative to our unsupervised procedure is manual selection of a sufficiently large region with approximately constant radar cross section, which is not possible for the San Francisco image. Such regions can be found in the Flevoland image, but the resulting estimate still varies, depending on the exact positioning of the estimation window within seemingly homogeneous areas. Another discussion goes to whether one should really aim at the true ENL value, or rather a value that provides a better model for the data by implicitly incorporating some of the texture not accounted for by the Wishart model. The ENL is not a physical entity, but a parameter of the less than perfect statistical model, which could justify a more pragmatic approach. If we choose to accept an ENL estimate that assimilates texture, then our unsupervised procedure that collects small sample estimates from the whole scene is appropriate, since the result is representative for the whole image.

Earlier studies of the ENL for four-look AIRSAR data have concluded that the data have characteristics close to that of three-look [34]. By matching distributions of phase and amplitude ratio that are assumed to be insensitive to texture, Lee et al. estimated the ENL for an AIRSAR C-band image of Howland Forest, USA to a value of 3.3 [9]. This is compatible with the results displayed in Fig. 10, but we still need to decide on a window size in order to obtain a value to compare with. The discussion on the window size effect related to mixed classes prescribes the use of the smallest window size possible. On the other hand, consideration of estimator bias forces us to increase the window size slightly. We believe that $k = 5$ is a good compromise, which should be applicable to various data sets. This window size yields bias corrected ENL estimates of 3.21 and 2.97 for the Flevoland and San Francisco data sets, respectively.

C. Computational Complexity

We finally present some results on the computational complexity of the tested algorithms. All algorithms are implemented in C language and optimised for speed. The performance measure is central processing unit (CPU) time, as measured by the Matlab *profile* function, on a 2.0 GHz Intel Pentium M processor. Fig. 11 displays mean CPU time required per estimate calculation as a function of sample size.

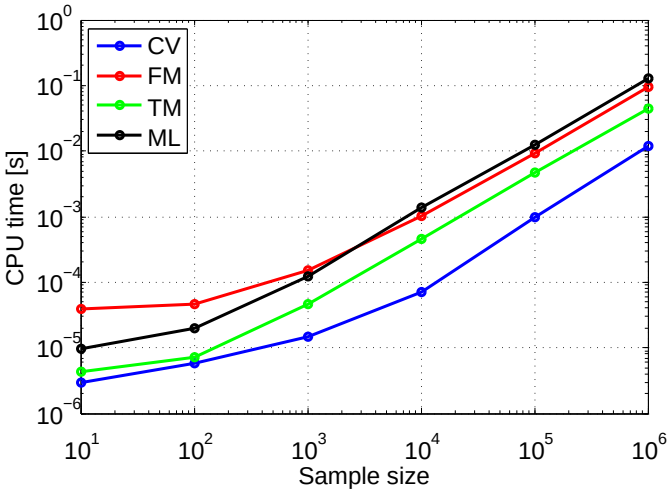

Fig. 11. Computational complexity of the CV, FM, TM, and ML estimator measured in CPU time per estimate calculation as function of sample size.

The figure shows that the CV estimator has the lowest computational cost, followed by the TM estimator. This is expected, since these are the mathematically simplest functions, with analytical solutions. The CV estimator is typically in the order of five to fifteen times faster than the ML estimator, depending on the sample size. The ranking of the inferior FM and ML estimators also depends on sample size, which can be explained. Both of these estimators are solved numerically and must be seeded with an initial value. For small sample sizes, the estimate is more likely to lie far off the seed value, which is typically chosen as the nominal number of looks. When this happens, the numerical scheme needs more time to converge. The FM estimator has higher variance than the ML estimator, as seen in Fig. 2, and is therefore more affected. As the sample size size gets higher, the variance becomes lower. Thus, the convergence time becomes shorter and less important, and algorithm speed depends more on the complexity of the mathematical functions. The digamma function in the ML estimator makes it slightly slower than the FM estimator, which can be implemented with the log-of-gamma function.

VIII. Conclusion

We have proposed two new estimators for the equivalent number of looks (ENL) that are adapted for polarimetric SAR (PolSAR) data. The expressions are derived by examining moment expressions of the multilook polarimetric covariance matrix (or, equivalently, the coherency matrix). The first estimator is found by rearranging the second order trace moment of the covariance matrix, and is thus called the trace moment (TM) estimator. The expression also provides a matrix-variate generalisation of the traditional definition of the ENL, established in the theory of single polarisation SAR. The second estimator is found from the log-determinant moment of the covariance matrix, and is also observed to be the maximum likelihood (ML) estimator based on the Wishart model for multilook PolSAR data. It is therefore coined the ML estimator. The proposed estimators are, as far as we know, the first ones to process the full multilook polarimetric covariance matrix, thus utilising all the available statistical information of PolSAR data. They readily reduce to estimators for single polarisation SAR data as a one-dimensional special case.

The new estimators have been compared with two estimators from the literature. The first is the traditional coefficient of variation (CV) estimator. The second, which we have called the fractional moment (FM) estimator, is the best method in the sparse literature on ENL estimation for PolSAR data. Both are based on moments of single polarisation intensities. Assessment of the statistical properties of all estimators shows that the TM estimator represents improvement with respect to the previously known methods, but the ML estimator is by far the superior one. We therefore launch it as the preferred estimator, not only for PolSAR data, but for SAR data in general. We have compared the bias and the variance of the estimators in experiments. A bound on the variance of an ENL estimator has also been derived, which is closely related to the Cramér-Rao bound. In addition to achieving the lowest bias and variance, the ML estimator is also shown to be less affected by texture, when the assumption of constant radar cross section does not hold for the input data sample.

We have finally examined the applicability of the ML estimator to unsupervised estimation, which obsoletes the manual selection of a region characterised by the appropriate statistics assumed in the definition of the ENL. An unsupervised estimation procedure is described. It is further shown through experiments that the low variance property of the ML estimator is the key feature that enables extraction of a reliable ENL estimate from the distribution of small sample estimates that have been calculated over the whole image without regards to the appropriateness of local statistics. Possible sources of error are discussed in detail, and practical solutions to issues such as bias reduction and selection of processing parameters have been proposed. The fully automatic unsupervised procedure offers a robust alternative to manual procedures, and represents a potensial improvement to an operational processing chain.

Appendix A: Proof of Estimator Convergence

We here analyse the convergence properties of the estimators that must be solved numerically because they have no analytic solution.

The fractional moment (FM) estimator is defined as the root of the polynomial $f(L)$, as given in Eq. (18). It can be shown that $f(L)$ is a monotonically increasing function of L. To prove that $f(L)$ has a root, we shall study the limiting values of $f(L)$ as $L \to \infty$ and $L \to 0$.

From [35], we have

$$\frac{\Gamma(L+1/2)}{\Gamma(L)\sqrt{L}} = 1 - \frac{1}{8L} + \frac{1}{128L^2} + \frac{5}{1024L^3} - \dots \,. \tag{48}$$

Thus,

$$\lim_{L\to\infty} f(L) = \lim_{L\to\infty} \frac{\Gamma(L+1/2)}{\Gamma(L)\sqrt{L}}\sqrt{\langle I\rangle} - \langle\sqrt{I}\rangle = \sqrt{\langle I\rangle} - \langle\sqrt{I}\rangle \geq 0\,. \tag{49}$$

The limiting value is the difference between the root mean square and the arithmetic mean of the detected amplitude, $\sqrt{I}$, which is always nonnegative by the known inequality for these entities.

Next step is to determine the limit of Eq. (48) as $L \to 0$. A standard power series expansion of $\Gamma(L)$ shows that

$$\Gamma(L) \propto 1/L - \gamma_{\text{EM}} + O(L) \tag{50}$$

where γ_{EM} is the Euler-Mascheroni constant and $O(\cdot)$ is Landau notation to denote order. It follows that

$$\Gamma(L)\sqrt{L} \propto \frac{1}{\sqrt{L}} - \gamma_{\text{EM}}\sqrt{L} + O(L^{3/2}) . \tag{51}$$

Thus, the numerator $\Gamma(L+1/2) \to \Gamma(1/2) = \sqrt{\pi}$ and the denominator $\Gamma(L)\sqrt{L} \to \infty$ as $L \to 0$. The limit becomes

$$\lim_{L \to \mathbf{0}} \frac{\Gamma(L+1/2)}{\Gamma(L)\sqrt{L}} = 0 \tag{52}$$

which proves that

$$\lim_{L \to 0} f(L) = -\langle \sqrt{I} \rangle . \tag{53}$$

The limit of $f(L)$ is negative as $L \to 0$ and $f(L)$ is a monotonically increasing function. Hence, it can be proved that there exists exactly one root of $f(L)$ in the interval $0 < L < \infty$ if and only if inequality occurs in Eq. (49), i.e., the limit as $L \to \infty$ must be strictly positive.

For a sample size of $N = 1$, $f(L)$ has no root, since in this case, $\sqrt{\langle I \rangle} = \langle \sqrt{I} \rangle$, and $f(L) < 0$ with probability equal to one for finite L. However, for $N \geq 2$, the limit is positive unless all samples have the same value. Thus, as long as the samples are nonidentical, a root exists and the estimator converges, provided it is implemented with a reliable root-finding algorithm. We have used the bisection method.

The proof for the maximum likelihood (ML) estimator follows the same path. The ML estimator is defined as the root of $g(L)$, as defined in Eq. (30). It can be shown that $g(L)$ is a monotonically decreasing function of L in the interval $d-1 < L < \infty$. Next observe that the digamma function can be expanded as

$$\Psi(L) \propto \ln(L) - \frac{1}{2L} - \frac{1}{12L^2}\left(1 + O\left(\frac{1}{L^2}\right)\right) \tag{54}$$

which is used to show that

$$\begin{aligned} \lim_{L\to\infty} g(L) &= \lim_{L\to\infty} \langle \ln|\mathbf{C}| \rangle - \ln\langle |\mathbf{C}| \rangle \\ &\quad - \sum_{i=0}^{d-1} \left(\ln\left(\frac{L-i}{L}\right) - \frac{1}{2(L-i)} \right. \\ &\quad \left. - \frac{1}{12(L-i)^2}\left[1 + O\left(\frac{1}{(L-i)^2}\right)\right]\right) \\ &= \langle \ln|\mathbf{C}| \rangle - \ln\langle |\mathbf{C}| \rangle \leq 0 . \end{aligned} \tag{55}$$

The inequality on the bottom line is easily proved by means of Jensen's inequality on finite form. It is also readily shown that

$$\lim_{L \to d-1} g(L) = \infty . \tag{56}$$

The limit as $L \to d-1$ is positive and $g(L)$ is a monotonically decreasing function. Thus, the existance of a root of $g(L)$ requires that the inequality in Eq. (55) is strictly negative. Equality occurs in Eq. (55) if and only if there is no variation in the sample, with $N = 1$ as a special case. Otherwise, for $N \geq 2$, a unique root of $g(L)$ exists in the interval $d-1 < L < \infty$ and the estimator converges. We note that the lower limit of this interval, introduced by the discontinuity of $g(L)$ at $d-1$, restricts the allowed range of the ML estimate. However, this is not a conceptual problem, since estimates $\hat{L}_e < d$ are in conflict with the condition for the Wishart distribution to be nonsingular.

APPENDIX B: DERIVATION OF LOG-DETERMINANT MOMENTS

In this appendix we derive low-order moments of $\ln(|\mathbf{Z}|/|\mathbf{\Sigma}|)$. By combining Eqs. (24) and (25), the moment generating function of $\ln(|\mathbf{Z}|/|\mathbf{\Sigma}|)$ was found to be

$$M_{\ln\left(\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|}\right)}(r) = \prod_{i=0}^{d-1} \frac{\Gamma(L-i+r)}{\Gamma(L-i)} . \tag{57}$$

The first-order moment thus becomes

$$\begin{aligned} \mathrm{E}\left\{\ln\left(\frac{|\mathbf{Z}|}{|\mathbf{\Sigma}|}\right)\right\} &= \left[\frac{d}{dr}\prod_{i=0}^{d-1}\frac{\Gamma(L-i+r)}{\Gamma(L-i)}\right]\Bigg|_{r=0} \\ &= \frac{\left[\sum_{i=0}^{d-1}\left(\Gamma'(L-i+r)\prod_{\substack{j=0\\ j\neq i}}^{d-1}\Gamma(L-j+r)\right)\right]\Bigg|_{r=0}}{\prod_{j=0}^{d-1}\Gamma(L-j)} \\ &= \frac{\left[\sum_{i=0}^{d-1}\Psi^{(0)}(L-i+r)\prod_{j=0}^{d-1}\Gamma(L-j+r)\right]\Bigg|_{r=0}}{\prod_{j=0}^{d-1}\Gamma(L-j)} \\ &= \sum_{i=0}^{d-1}\Psi^{(0)}(L-i) . \end{aligned} \tag{58}$$

To arrive at this result we have used the product rule of differentiation repeatedly and utilised the relation $\Gamma'(L) = \Gamma(L)\Psi^{(0)}(L)$, where $\Gamma'(L)$ denotes the derivative of $\Gamma(L)$.

In the same manner, the second-order moment is derived as

$$\mathrm{E}\left\{\ln^2\left(\frac{|\mathbf{Z}|}{|\boldsymbol{\Sigma}|}\right)\right\} = \sum_{i=0}^{d-1}\Psi^{(1)}(L-i) + \left(\sum_{i=0}^{d-1}\Psi^{(0)}(L-i)\right)^2 . \quad (59)$$

By combining the first-order and second-order moment, it is discovered that

$$\begin{aligned}\mathrm{Var}\left\{\ln|\mathbf{Z}|\right\} &= \mathrm{Var}\left\{\ln\left(\frac{|\mathbf{Z}|}{|\boldsymbol{\Sigma}|}\right)\right\} \\ &= \sum_{i=0}^{d-1}\Psi^{(1)}(L-i)\,.\end{aligned} \quad (60)$$

This expression can also be used to estimate L, but the performance is inferior to the estimator derived from (58), as the second-order moment is more difficult to estimate than the first-order moment.

Appendix C: Derivation of Fisher Information Matrix

In this appendix we derive the Fisher information matrix (FIM) of the complex parameter vector $\boldsymbol{\theta} = [\mathrm{L_e}, \mathrm{vec}(\boldsymbol{\Sigma})^T]^T$. The log-likelihood function of a size N complex Wishart distributed sample $\mathscr{Z} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_N\}$ was given in (35). It can be expanded to

$$\begin{aligned}&\mathcal{L}(\mathscr{Z}; \mathrm{L_e}, \boldsymbol{\Sigma}) \\ &\quad= \sum_{i=1}^{N}\left((n-p)\ln|\mathbf{Z}_i| - n\ln|\boldsymbol{\Sigma}|\right. \\ &\qquad\left. - \ln\Gamma_d(\mathrm{L_e}) - \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Z}_i)\right).\end{aligned} \quad (61)$$

The partial derivatives of $\mathcal{L}(\mathscr{Z})$ (with parameters suppressed) with respect to $\mathrm{L_e}$ follow readily as

$$\frac{\partial\mathcal{L}(\mathscr{Z})}{\partial\,\mathrm{L_e}} = N\left(\langle\ln|\mathbf{Z}|\rangle - \ln|\boldsymbol{\Sigma}| - \sum_{i=0}^{d-1}\Psi^{(0)}(\mathrm{L_e}-i)\right) \quad (62)$$

and

$$\frac{\partial^2\mathcal{L}(\mathscr{Z})}{\partial\,\mathrm{L_e}^2} = -N\sum_{i=0}^{d-1}\Psi^{(1)}(\mathrm{L_e}-i)\,. \quad (63)$$

The first partial derivative with respect to $\boldsymbol{\Sigma}$ is found from standard rules of complex matrix calculus [36]:

$$\frac{\partial\mathcal{L}(\mathscr{Z})}{\partial\boldsymbol{\Sigma}} = -N\,\mathrm{L_e}\,\boldsymbol{\Sigma}^{-1} + N\boldsymbol{\Sigma}^{-1}\langle\mathbf{Z}\rangle\boldsymbol{\Sigma}^{-1}\,. \quad (64)$$

To obtain the second partial derivative, we need

$$\begin{aligned}\frac{\partial\boldsymbol{\Sigma}^{-1}\langle\mathbf{Z}\rangle\boldsymbol{\Sigma}^{-1}}{\partial\boldsymbol{\Sigma}} &= \frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\boldsymbol{\Sigma}}\left(\mathbf{I}_d\otimes\langle\mathbf{Z}\rangle\boldsymbol{\Sigma}^{-1}\right) \\ &\quad+ \left(\mathbf{I}_d\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial(\langle\mathbf{Z}\rangle\boldsymbol{\Sigma}^{-1})}{\partial\boldsymbol{\Sigma}} \\ &= (-\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})(\mathbf{I}_d\otimes\langle\mathbf{Z}\rangle\boldsymbol{\Sigma}^{-1}) \\ &\quad+ (\mathbf{I}_d\otimes\boldsymbol{\Sigma}^{-1})(-\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})(\mathbf{I}_d\otimes\langle\mathbf{Z}\rangle) \\ &= -(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\langle\mathbf{Z}\rangle\boldsymbol{\Sigma}^{-1}) \\ &\quad- (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\langle\mathbf{Z}\rangle)\,.\end{aligned} \quad (65)$$

This result occurs after repeated applications of the chain rule and the product rule in Theorem 4.3 of [37]. When differentiating with respect to $\boldsymbol{\Sigma}$, it takes the form

$$\frac{\partial(\mathbf{AB})}{\partial\boldsymbol{\Sigma}} = \frac{\partial\mathbf{A}}{\partial\boldsymbol{\Sigma}}(\mathbf{I}_d\otimes\mathbf{B}) + (\mathbf{I}_d\otimes\mathbf{A})\frac{\partial\mathbf{B}}{\partial\boldsymbol{\Sigma}} \quad (66)$$

for two arbitrary complex matrices $\mathbf{A}$ and $\mathbf{B}$ with compatible dimensions. We have also used $\partial\boldsymbol{\Sigma}^{-1}/\partial\boldsymbol{\Sigma} = -\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}$ [36]. It follows that

$$\begin{aligned}\frac{\partial^2\mathcal{L}(\mathscr{Z})}{\partial\boldsymbol{\Sigma}^2} &= N\,\mathrm{L_e}\,\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1} \\ &\quad- N(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\langle\mathbf{Z}\rangle\boldsymbol{\Sigma}^{-1}) - N(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\langle\mathbf{Z}\rangle)\,.\end{aligned} \quad (67)$$

From Eqs. (62) and (64) we finally obtain

$$\begin{aligned}\frac{\partial}{\partial\,\mathrm{L_e}}\left(\frac{\partial\mathcal{L}(\mathscr{Z})}{\partial\boldsymbol{\Sigma}}\right) &= \frac{\partial}{\partial\boldsymbol{\Sigma}}\left(\frac{\partial\mathcal{L}(\mathscr{Z})}{\partial\,\mathrm{L_e}}\right) \\ &= -N\boldsymbol{\Sigma}^{-1}\end{aligned} \quad (68)$$

and are now equipped for the derivation of $\mathbf{J}$, the FIM of $\boldsymbol{\theta}$.

By elaborating on (37), $\mathbf{J}$ can be expressed as

$$\mathbf{J} = \begin{bmatrix}\mathbf{J}_{11} & \mathbf{J}_{12}\\ \mathbf{J}_{21} & \mathbf{J}_{22}\end{bmatrix} \quad (69)$$

with quadrant submatrices defined as

$$\begin{aligned}\mathbf{J}_{11} &= -\,\mathrm{E}\left\{\frac{\partial^2}{\partial\,\mathrm{L_e}^2}\mathcal{L}(\mathscr{Z})\right\} \\ &= N\sum_{i=0}^{d-1}\Psi^{(1)}(\mathrm{L_e}-i)\end{aligned} \quad (70)$$

$$\mathbf{J}_{12} = -\,\mathrm{E}\left\{\frac{\partial}{\partial\,\mathrm{L_e}}\left(\frac{\partial\mathcal{L}(\mathscr{Z})}{\partial\,\mathrm{vec}(\mathbf{\Sigma})}\right)^T\right\} = N\,\mathrm{vec}(\mathbf{\Sigma}^{-1})^T \tag{71}$$

$$\mathbf{J}_{21} = -\,\mathrm{E}\left\{\frac{\partial}{\partial\,\mathrm{vec}(\mathbf{\Sigma})^*}\left(\frac{\partial\mathcal{L}(\mathscr{Z})}{\partial\,\mathrm{L_e}}\right)\right\} = N\,\mathrm{vec}(\mathbf{\Sigma}^{-1})^* \tag{72}$$

and

$$\mathbf{J}_{22} = -\,\mathrm{E}\left\{\frac{\partial}{\partial\,\mathrm{vec}(\mathbf{\Sigma})^*}\left(\frac{\partial\mathcal{L}(\mathscr{Z})}{\partial\,\mathrm{vec}(\mathbf{\Sigma})}\right)^T\right\} = N\,\mathrm{L_e}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\,. \tag{73}$$

In the evaluation of the submatrices of $\mathbf{J}$, we have used equations (63), (67), and (68), together with the differential relation $\partial/\partial\,\mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\partial/\partial\mathbf{A})$, which is valid for an arbitrary complex matrix $\mathbf{A}$ [36]. We thus have

$$\mathbf{J} = N\begin{bmatrix}\displaystyle\sum_{i=0}^{d-1}\Psi^{(1)}(\mathrm{L_e}-i) & \mathrm{vec}(\mathbf{\Sigma}^{-1})^T \\ \mathrm{vec}(\mathbf{\Sigma}^{-1})^* & \mathrm{L_e}(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\end{bmatrix} \tag{74}$$

and the inverse FIM in (40), defining the bound discussed in section V-A, follows readily.

ACKNOWLEDGMENT

The authors would like to thank NASA/JPL-Caltech for making available the AIRSAR data sets used in the paper. We further thank the European Space Agency and the POLSARPRO software development team for their efforts in publishing open source software and PolSAR data samples. Finally, we wish to thank the anonymous reviewers for their valuable comments.

REFERENCES

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

[2] I. G. Cumming and F. H. Wong, *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*. Norwood, U.S.: Artech House, 2005.

[3] J.-S. Lee, M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sensing*, vol. 15, no. 11, pp. 229–231, Nov. 1994.

[4] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart distribution," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2259, Sep. 1999.

[5] P. R. Kersten, J.-S. Lee, and T. L. Ainsworth, "Unsupervised classification of polarimetric synthetic aperture radar images using fuzzy clustering and EM clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 519–527, Mar. 2005.

[6] A. C. Frery, A. H. Correia, and C. C. Freitas, "Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3098–3109, Oct. 2007.

[7] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.

[8] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 4–19, Jan. 2003.

[9] 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 Sensing Symp., IGARSS'94*, vol. 4, Pasadena, U.S., Aug. 1994, pp. 2179–2181.

[10] 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.

[11] R. Touzi, W. M. Boerner, J.-S. Lee, and E. Lüneburg, "A review of polarimetry in the context of synthetic aperture radar: Concepts and information extraction," *Can. J. Remote Sensing*, vol. 30, no. 3, pp. 380–407, 2004.

[12] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *J. Electrom. Waves Applic.*, vol. 3, no. 8, pp. 747–768, 1989.

[13] 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.

[14] C. H. Gierull and I. C. Sikaneta, "Estimating the effective number of looks in interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1733–1742, Aug. 2002.

[15] J.-S. Lee and K. Hoppel, "Noise modeling and estimation of remotely-sensed images," in *Proc. IEEE Int. Geosc. Remote Sensing Symp., IGARSS'89*, vol. 2, Vancouver, Canada, Jul. 1989, pp. 1005–1008.

[16] J.-S. Lee, K. Hoppel, and S. A. Mango, "Unsupervised estimation of speckle noise in radar images," *International Journal of Imaging Systems and Technology*, vol. 4, no. 4, pp. 298–305, 1992.

[17] S. Foucher, J.-M. Boucher, and G. B. Benie, "Maximum likelihood estimation of the number of looks in SAR images," in *International Conference on Microwaves, Radar and Wireless Communications*, vol. 2, Wroclaw, Poland, 2000, pp. 657–660.

[18] I. R. Joughin and D. P. Winebrenner, "Effective number of looks for a multilook interferometric phase distribution," in *Proc. IEEE Int. Geosc. Remote Sensing Symp., IGARSS'94*, Pasadena, U.S., Aug. 1994, pp. 2276–2278.

[19] A. El Zaart, D. Ziou, S. Wang, and Q. Jiang, "Segmentation of SAR images," *Pattern Recognition*, vol. 35, no. 3, pp. 713–724, mar 2002.

[20] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Estimation of the equivalent number of looks in polarimetric SAR imagery," in *Proc. IEEE Int. Geosc. Remote Sensing Symp., IGARSS'08*, vol. 4, Boston, U.S., Jul. 2008, pp. 487–490.

[21] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proc. Radar, Sonar, Navigation*, vol. 147, no. 4, pp. 162–168, 2000.

[22] H. H. Andersen, M. Højbjerre, D. Sørensen, and P. S. Eriksen,

Linear and Graphical Models for the Multivariate Complex Normal Distribution, ser. Lecture Notes in Statistics. New York, U.S.: Springer, 1995.

[23] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Hoboken, U.S.: John Wiley & Sons, 1982.

[24] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. New York, U.S.: Addison-Wesley, 1991.

[25] A. van den Bos, "A Cramér-Rao lower bound for complex parameters," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2859–2859, Oct. 1994.

[26] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1017–1028, Sep. 1994.

[27] I. R. Joughin, D. P. Winebrenner, and D. B. Percival, "Probability density functions for multilook polarimetric signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 562–574, May 1994.

[28] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Boca Raton, U.S.: Chapman & Hall, 1986.

[29] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, U.K.: Chapman & Hall, 1995.

[30] D. N. Politis, "Computer-intensive methods in statistical analysis," *Signal Processing Magazine*, vol. 15, no. 1, pp. 39–55, Jan. 1998.

[31] A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *Signal Processing Magazine*, vol. 15, no. 1, pp. 56–76, Jan. 1998.

[32] F. Cribari-Neto, A. C. Frery, and M. F. Silva, "Improved estimation of clutter properties in speckled imagery," *Computational Statistics and Data Analysis*, vol. 40, no. 4, pp. 801–824, Oct. 2002.

[33] J.-S. Lee, M. R. Grunes, and G. de Grandi, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2363–2373, Sep. 1999.

[34] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," in *Proc. IEEE Int. Geosc. Remote Sensing Symp., IGARSS'93*, vol. 2, Tokyo, Japan, Aug. 1993, pp. 813–816.

[35] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Answer to problem 9.60. Reading, U.S.: Addison-Wesley, 1994.

[36] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[37] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 772–781, Sep. 1978.

Chapter 6

Paper 2:
Application of the Matrix-Variate Mellin Transform to Analysis of Polarimetric Radar Images

Application of the Matrix-Variate Mellin Transform to Analysis of Polarimetric Radar Images

Stian Normann Anfinsen, *Student Member, IEEE* and Torbjørn Eltoft, *Member, IEEE*

***Abstract*—In this paper we propose to use a matrix-variate Mellin transform in the statistical analysis of multi-look polarimetric radar data. The domain of the transform integral is the cone of complex positive definite matrices, which allows for transformation of the covariance matrix distributions used to model multilook polarimetric radar data. Based on the matrix-variate Mellin transform, an alternative characteristic function is defined, from which we can retrieve a new kind of matrix log-moments and log-cumulants. It is demonstrated that the matrix log-cumulants are of great value to analysis of polarimetric radar data, and that they can be used to derive estimators for the distribution parameters with low bias and variance.**

***Index Terms*—Radar polarimetry, synthetic aperture radar, Mellin transform, matrix-variate statistics, parameter estimation, method-of-log-cumulants, doubly stochastic product model**

I. Introduction

POLARIMETRIC radar has become an important remote sensing instrument due to its ability to discriminate between different scattering mechanisms. It can therefore characterise physical properties of the target that cannot be determined from single polarisation (mono-pol) radar measurements. To fully utilise the polarimetric information captured, it is necessary to analyse the complex correlations between all polarimetric channels, incorporating all intensity and phase information. This requires relatively complicated data models, that together with the speckle phenomenon, inherent to all types of coherent imaging, make analysis of multiple polarization radar data a challenging task.

It was noted already by Epstein [1] that the Mellin transform (MT) is a natural analytical tool to use when studying the distribution of products and quotients of independent random variables (RV). Nicolas [2], [3] utilised this fact in the analysis of compounded distributions used to model synthetic aperture radar (SAR) data. He introduced a new theoretical framework by replacing the Fouriér transform (FT) with the MT in the definition of the characteristic function (CF) and cumulant generating function (CGF). This framework was originally coined second kind statistics, but we shall refer to it as *Mellin kind statistics* (MKS). From the resulting Mellin kind CF and CGF one can retrieve the statistics known as log-moments and log-cumulants.

The most important development under this framework is the *method of log-cumulants* (MoLC) for parameter estimation [2], [3], which Nicolas applied to a number of doubly stochastic distributions, as well as the positive alpha-stable distribution [4], and members of the generalised gamma distribution (GγD) family (e.g., the Weibull and log-normal distribution) [5]. The same method has earlier been applied to GγDs [6], though without relating it to the MT. The list of recently proposed SAR image analysis and image processing algorithms that employ the MoLC covers diverse applications such as statistical modelling [7], [8], [9], speckle filtering [10], [11], [12], classification [13], segmentation [14], [15], change detection [16], [17], [18], interferometric coherence estimation [19], and image compression [20].

Being aware of the impact that MKS has had on mono-pol SAR image analysis, we here extend the theory to the matrix-variate case which describes multilook polarimetric radar data. This is done by introducing a matrix-variate version of the MT [21], which is used to define a Mellin kind CF and CGF of random matrices. We then show how *matrix log-moments* and *matrix log-cumulants* can be obtained from the Mellin kind matrix CF and CGF, respectively. For all the theoretical derivations, we highlight the analogy with the univariate case developed by Nicolas, and also the classical theory where the FT is used instead of the MT.

The paper is organised as follows. In Section II

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

we describe the data delivered by mono-pol and polarimetric radars, together with the probability density functions (PDFs) commonly used to model the data. In Section III we review the classical definition of moments and cumulants and give an overview of univariate MKS, before presenting the extension to complex matrix-variate MKS including new definitions and derivations of key properties. The application to parameter estimation for multilook polarimetric radar data distributions using the *method of matrix log-cumulants* (MoMLC) is presented in Section IV, accompanied by numerical simulations that document the improvement of estimator precision and accuracy. 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

CF	characteristic function
CGF	cumulant generating function
FT	Fourier transform
GγD	generalised gamma distribution
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. RADAR DATA MODELS

A full-polarimetric imaging radar separately transmits orthogonally polarised microwave pulses, and measures orthogonal polarisations of the received signal. For each pixel, the measurements result in a matrix of scattering coefficients. These are complex-valued, dimensionless numbers that describe the transformation of the transmitted electromagnetic field to the received electromagnetic field for all combinations of transmit and receive polarisation, and assuming no atmospheric disturbance (i.e. zero Faraday rotation).

The transformation can be expressed as

$$\begin{bmatrix} \mathcal{E}_x^{\uparrow} \\ \mathcal{E}_y^{\uparrow} \end{bmatrix} = \frac{e^{\jmath kr}}{r} \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix} \begin{bmatrix} \mathcal{E}_x^{\downarrow} \\ \mathcal{E}_y^{\downarrow} \end{bmatrix} \tag{1}$$

where k denotes wavenumber and r is the distance between radar and target. The subscript of the electromagnetic field components $\mathcal{E}_i^{\downarrow}$ and $\mathcal{E}_i^{\uparrow}$, $i \in \{x, y\}$, refers to one of the orthogonal polarisations x and y. The superscript indicates transmitted/incident ($\downarrow$) or received/backscattered ($\uparrow$) wave. SAR systems normally use linear polarisations (horizontal and vertical), while using circular polarisations (left and right) is another choice. The scattering coefficients S_{ij} are subscripted with the associated receive and transmit polarisation, in that order. Together, they form the scattering matrix, denoted $\mathbf{S} = [S_{ij}] \in \mathbb{C}^{2\times 2}$.

The scattering vector may be defined as

$$\mathbf{s} = [S_{xx}\ S_{xy}\ S_{yx}\ S_{yy}]^T = \mathrm{vec}(\mathbf{S}^T) \in \mathbb{C}^d \tag{2}$$

where $(\cdot)^T$ means transposition, $\mathrm{vec}(\cdot)$ denotes vectorisation by column stacking, and $d = \dim(\mathbf{s}) = 4$ is the vector dimension. Other definitions are possible [22], since the vector can be linearly transformed to emphasise physical interpretations of the elements (i.e., Pauli basis), or the dimension can be reduced to $d = 3$ by assuming reciprocity of the target (i.e, $S_{xy} = S_{yx}$). A reduced version also results when only a subset of $\mathbf{S}$ is measured, such as for mono-pol SAR ($d = 1$) and dual polarisation SAR ($d = 2$).

Radar images are affected by an interference phenomenon which is a characteristic of all coherent imaging systems. The noise-like effect of interference, known as speckle, can be mitigated by a processing step called multilooking. Multiple measurements are obtained by splitting the Doppler bandwidth into a number of subbands, each giving rise to a separate image referred to as a look. All looks are averaged in the power domain to produce multilook data.

The matrix $\mathbf{S}$ and the vector $\mathbf{s}$ are single-look complex format representations of polarimetric radar data. The L looks extracted in the multilooking process may be represented by the set $\{\mathbf{s}_\ell\}_{\ell=1}^L$, or $\{S_\ell\}_{\ell=1}^L$ in the mono-pol case. The data formats obtained in the multilook intensity domain are

$$C = \frac{1}{L}\sum_{\ell=1}^{L} S_\ell S_\ell^*, \quad d = 1 \tag{3}$$

where $(\cdot)^*$ denotes complex conjugation, and

$$\mathbf{C} = \frac{1}{L}\sum_{\ell=1}^{L}\mathbf{s}_\ell\mathbf{s}_\ell^H\,, \quad d>1 \tag{4}$$

where $(\cdot)^H$ denotes the Hermitian (conjugate transposition) operator. We refer to $C \in \mathbb{R}^+$ as the multilook intensity and $\mathbf{C} \in \Omega_+ \subset \mathbb{C}^{d\times d}$ as the multilook polarimetric covariance matrix. Note that C is a real positive RV, whereas $\mathbf{C}$ is a random matrix defined on the cone[1] Ω_+ of positive definite complex Hermitian matrices:

$$\Omega_+ = \{\mathbf{X} : \mathbf{X} \succ \mathbf{0},\, \mathbf{X} = \mathbf{X}^H \in \mathbb{C}^{d\times d}\} \tag{5}$$

where $\mathbf{X} \succ \mathbf{0}$ means that $\mathbf{X}$ is positive definite.

A. Gaussian Model

It is commonly assumed that the scattering vector elements are jointly circular complex Gaussian. This is strictly justified only for homogeneous regions of the image characterised by fully developed speckle and no texture. The notion of texture is here defined as spatial variation in the backscatter that is due to target variability, i.e., fluctuations in the radar cross section. The Gaussian model only encompasses variability due to the stochastic interference pattern, that is, speckle.

Assume for the moment that $\mathbf{s}$ is zero mean and circular complex multivariate Gaussian, denoted $\mathbf{s} \sim \mathcal{N}_d^{\mathbb{C}}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\mathbf{0}$ is a column vector of zeros and $\boldsymbol{\Sigma} = \mathrm{E}\{\mathbf{s}\mathbf{s}^H\}$ is the covariance matrix of $\mathbf{s}$. $\mathrm{E}\{\cdot\}$ denotes the expectation value. The PDF of $\mathbf{s}$ is

$$p_{\mathbf{s}}(\mathbf{s};\boldsymbol{\Sigma}) = \frac{1}{\pi^d|\boldsymbol{\Sigma}|}\exp(-\mathbf{s}^H\boldsymbol{\Sigma}^{-1}\mathbf{s}) \tag{6}$$

where $|\cdot|$ is the determinant operator. It follows that if $L \geq d$ and the $\mathbf{s}_\ell$ are independent, then the scaled sample covariance matrix, defined as $\mathbf{Z} = L\mathbf{C}$, follows the nonsingular complex Wishart distribution [23]:

$$p_{\mathbf{Z}}(\mathbf{Z};L,\boldsymbol{\Sigma}) = \frac{|\mathbf{Z}|^{L-d}}{\Gamma_d(L)|\boldsymbol{\Sigma}|^L}\,\mathrm{etr}(-\boldsymbol{\Sigma}^{-1}\mathbf{Z}) \tag{7}$$

where $\mathrm{etr}(\cdot) = \exp(\mathrm{tr}(\cdot))$ is the exponential trace operator and $\boldsymbol{\Sigma} = \mathrm{E}\{\mathbf{C}\} = \mathrm{E}\{\mathbf{Z}\}/L$. We write this as $\mathbf{Z} \sim \mathcal{W}_d^{\mathbb{C}}(L,\boldsymbol{\Sigma})$. The normalisation constant $\Gamma_d(L)$ is the multivariate gamma function of the complex case, defined as

$$\begin{aligned}\Gamma_d(L) &= \int_{\Omega_+}|\mathbf{Z}|^{L-d}\,\mathrm{etr}(-\mathbf{Z})\,d\mathbf{Z}\\ &= \pi^{d(d-1)/2}\prod_{i=0}^{d-1}\Gamma(L-i)\end{aligned} \tag{8}$$

where $\Gamma(L)$ is the standard Euler gamma function. We further have $p_{\mathbf{C}}(\mathbf{C}) = p_{\mathbf{Z}}(L\mathbf{C})|J_{\mathbf{Z}\to\mathbf{C}}|$, where $|J_{\mathbf{Z}\to\mathbf{C}}| = L^{d^2}$ is the Jacobian determinant of the transformation $\mathbf{Z} = L\mathbf{C}$ [21]. The PDF of $\mathbf{C}$ becomes

$$p_{\mathbf{C}}(\mathbf{C};L,\boldsymbol{\Sigma}) = \frac{L^{Ld}}{\Gamma_d(L)}\frac{|\mathbf{C}|^{L-d}}{|\boldsymbol{\Sigma}|^L}\,\mathrm{etr}(-L\boldsymbol{\Sigma}^{-1}\mathbf{C})\,. \tag{9}$$

In the one-dimensional case, the complex Wishart distribution reduces to the gamma distribution with PDF:

$$p_C(c;L,\sigma) = \frac{L^L}{\Gamma(L)}\frac{c^{L-1}}{\sigma^L}\exp\left(-\frac{Lc}{\sigma}\right) \tag{10}$$

where $\sigma = \mathrm{E}\{C\}$. This is denoted $C \sim \gamma(L,\sigma)$.

For the Gaussian model, we denote the scaled covariance matrix $\mathbf{Z}$ by $\mathbf{W}$, and $\mathbf{C}$ by $\tilde{\mathbf{W}}$. We also refer to the PDF in (7) as $p_{\mathbf{W}}(\mathbf{W})$, to emphasise that it is a complex Wishart distribution.

B. Product Model

As described above, the randomness of a radar image measurement is commonly attributed to two unrelated factors, namely speckle and texture. The latter represents the natural spatial variation of the radar cross section, which is generally not perfectly homogeneous for pixels that are thematically mapped as one class. While the Gaussian model only accounts for speckle, several statistical models exist that also incorporate texture, either by assuming statistics that imply a non-Gaussian scattering vector, or explicitly, by modelling texture as a separate RV. The latter case leads to a doubly stochastic model with a so-called compounded distribution.

The well known product model, described e.g. in [26], [27], has been shown to be both mathematically tractable and successful for modelling and prediction purposes in coherent imaging. In the multilook polarimetric version, which is extensively reviewed in [25], it decomposes $\mathbf{Z}$ as a product of two independent stochastic variables,

$$\mathbf{Z} = T\mathbf{W}\,, \tag{11}$$

[1] A cone is defined as a subset of a vector space that is closed under multiplication by positive scalars.

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

$p_T(t)$ of texture variable T		$p_{\mathbf{C}}(\mathbf{C})$ of covariance matrix $\mathbf{C}$		Ref.
Constant	$\delta(t-1)$	$\mathcal{W}_d^{\mathbb{C}}(\mathbf{\Sigma}, L)$	$\frac{L^{Ld}}{\Gamma_d(L)} \frac{\lvert\mathbf{C}\rvert^{L-d}}{\lvert\mathbf{\Sigma}\rvert^{L}} \operatorname{etr}(-L\mathbf{\Sigma}^{-1}\mathbf{C})$	[23]
$\bar{\gamma}(\alpha)$	$\frac{\alpha^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\alpha t)$	$\mathcal{K}_d(\mathbf{\Sigma}, L, \alpha)$	$\frac{2\lvert\mathbf{C}\rvert^{L-d}(L\alpha)^{\frac{\alpha+Ld}{2}}}{\lvert\mathbf{\Sigma}\rvert^{L}\Gamma_d(L)\Gamma(\alpha)} \left(\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\right)^{\frac{\alpha-Ld}{2}} K_{\alpha-Ld}\left(2\sqrt{L\alpha \operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})}\right)$	[24]
$\bar{\gamma}^{-1}(\lambda)$	$\frac{(\lambda-1)^\lambda}{\Gamma(\lambda)} \frac{1}{t^{\lambda+1}} \exp\left(-\frac{\lambda-1}{t}\right)$	$G_d^0(\mathbf{\Sigma}, L, \lambda)$	$\frac{L^{Ld}\lvert\mathbf{C}\rvert^{L-d}}{\Gamma_d(L)\lvert\mathbf{\Sigma}\rvert^{L}} \frac{\Gamma(Ld+\lambda)(\lambda-1)^\lambda}{\Gamma(\lambda)} \left(L\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}) + \lambda - 1\right)^{-\lambda-Ld}$	[25]
$\bar{\mathcal{F}}(\xi, \zeta)$	$\frac{\Gamma(\xi+\zeta)}{\Gamma(\xi)\Gamma(\zeta)} \frac{\xi}{\zeta-1} \frac{\left(\frac{\xi}{\zeta-1}t\right)^{\xi-1}}{\left(\frac{\xi}{\zeta-1}t+1\right)^{\xi+\zeta}}$	$\mathcal{U}_d(\mathbf{\Sigma}, L, \xi, \zeta)$	$\frac{L^{Ld}\lvert\mathbf{C}\rvert^{L-d}}{\Gamma_d(L)\lvert\mathbf{\Sigma}\rvert^{L}} \frac{\Gamma(\xi+\zeta)}{\Gamma(\xi)\Gamma(\zeta)} \left(\frac{\xi}{\zeta-1}\right) \Gamma(Ld+\zeta)$ $\times U\left(Ld+\zeta, Ld-\xi+1, L\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\xi/(\zeta-1)\right)$	[9]

with individual distributions. The positive, scalar and unit mean RV T generates texture, assuming that its contribution is independent of polarisation and common for all channels. The matrix variable $\mathbf{W} \sim \mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma})$ models speckle. The PDF of $\mathbf{Z}$ depends on the PDF of the multilook texture RV T.

In [25], the family of generalised inverse Gaussian distributions is proposed as a model for T. Table I lists the gamma ($\bar{\gamma}$), inverse gamma ($\bar{\gamma}^{-1}$), and Fisher-Snedecor ($\bar{\mathcal{F}}$) distribution as possible choices of $p_T(t)$, giving both notation and expression of their PDF. We remark that the distributions have been normalised to unit mean, indicated by the overbar in the given symbol, which fixes and obsoletes one parameter of the unconstrained distribution. The table also presents the resulting distributions for $\mathbf{C}$, calculated from

$$\begin{aligned} p_{\mathbf{C}}(\mathbf{C}) &= \int_0^\infty p_{\mathbf{Z}|T}(L\mathbf{C}|t) p_T(t) |J_{\mathbf{Z}\to\mathbf{C}}|\, dt \\ &= |J_{\mathbf{Z}\to\mathbf{C}}| \int_0^\infty p_{\mathbf{W}}(tL\mathbf{C}) p_T(t)\, dt\,. \end{aligned} \tag{12}$$

These covariance matrix distributions are the matrix-variate $\mathcal{K}$ distribution, the matrix-variate $\mathcal{G}^0$ distribution, and the $\mathcal{U}$-distribution. The $\mathcal{K}$ distribution and the $\mathcal{U}$ distribution are named, respectively, after the second kind modified Bessel function of order ν, denoted $K_\nu(\cdot)$, and the second kind confluent hypergeometric Kummer function, denoted $U(\cdot,\cdot,\cdot)$. The complex Wishart distribution is also presented and interpreted as a special case of the product model, with a constant texture parameter, whose PDF is written in terms of a Dirac delta function, $\delta(t)$.

We remark that the PDF of a product of random variables is known as a Mellin convolution of the factor densities. Thus, (12) can be viewed as a Mellin convolution, as we shall return to later.

III. MELLIN KIND STATISTICS

A. Classical Statistics

A scalar statistical moment captures certain characteristics of a statistical distribution by projecting its PDF onto a scalar. In the univariate case, the νth-order raw moment of a real RV X is defined as

$$m_\nu\{X\} = \mathrm{E}\{X^\nu\} = \int_{-\infty}^{+\infty} x^\nu p_X(x)\, dx \tag{13}$$

and the νth-order central moment as

$$\begin{aligned} \bar{m}_\nu\{X\} &= \mathrm{E}\{(X - m_1)^\nu\} \\ &= \int_{-\infty}^{+\infty} (x - m_1)^\nu p_X(x)\, dx \end{aligned} \tag{14}$$

where $m_1 = \mathrm{E}\{X\}$. Let $\mathcal{F}\{\cdot\}(\omega)$ and $\mathcal{F}^{-1}\{\cdot\}(x)$ be the forward and inverse FT, respectively. The classical CF is defined as $\Phi_X(\omega) = \mathrm{E}\{e^{\jmath\omega X}\} = \mathcal{F}\{p_X(x)\}(\omega)$,[2] with ω a real number and $\jmath$ the imaginary unit. The CF always exists. When all moments exist, the CF can be written as

$$\begin{aligned} \Phi_X(\omega) &= \int_{-\infty}^{+\infty} e^{\jmath\omega x} p_X(x)\, dx \\ &= \sum_{\nu=0}^{\infty} \frac{(\jmath\omega)^\nu}{\nu!} m_\nu\{X\} \end{aligned} \tag{15}$$

[2]The exponential in the Fouriér transform may be defined with or without a negative sign in the exponent. We have chosen the latter version.

using the Maclaurin series expansion of the exponential function. The νth-order moment can in this case be retrieved from

$$m_\nu\{X\} = (-\jmath)^\nu \frac{d^\nu}{d\omega^\nu}\Phi_X(\omega)\bigg|_{\omega=0} . \tag{16}$$

A statistical distribution is uniquely specified by its CF if all of its moments are finite and the power series expansion for its CF converges absolutely near the origin [28]. Then

$$\begin{aligned} p_X(x) &= \mathcal{F}^{-1}\{\Phi_X(\omega)\}(x) \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\jmath\omega x}\Phi_X(\omega)\,d\omega\,. \end{aligned} \tag{17}$$

The CGF of X is defined as $\Psi_X(\omega) = \ln\Phi_X(\omega)$. When the moments $m_\nu\{X\}$ exist, so do the cumulants $c_\nu\{X\}$, that are coefficients of the power series expansion

$$\Psi_X(\omega) = \sum_{\nu=0}^{\infty}\frac{(\jmath\omega)^\nu}{\nu!}c_\nu\{X\}\,. \tag{18}$$

The cumulants can be retrieved from the CGF as

$$c_\nu\{X\} = (-\jmath)^\nu \frac{d^\nu}{d\omega^\nu}\Psi_X(\omega)\bigg|_{\omega=0} \tag{19}$$

by analogy with the moments. Moments and cumulants are related by a combinatorial version of Faá di Bruno's formula [29]:

$$\begin{aligned} c_\nu\{X\} &= m_\nu\{X\} \\ &\quad - \sum_{i=1}^{\nu-1}\binom{\nu-1}{i-1}c_i\{X\}m_{\nu-i}\{X\} \end{aligned} \tag{20}$$

and reversely through

$$m_\nu\{X\} = B_\nu\big(c_1\{X\},\ldots,c_\nu\{X\}\big) \tag{21}$$

where $B_\nu(\cdot)$ is the νth complete Bell polynomial [30]. All relations for classical univariate statistics are summarised in the diagram of Figure 1.

Moments and cumulants can be generalised to random vectors: $\mathbf{x}\sim p_{\mathbf{x}}(\mathbf{x})$, $\mathbf{x}\in\mathbb{R}^n$, random matrices: $\mathbf{X}\sim p_{\mathbf{X}}(\mathbf{X})$, $\mathbf{X}\in\mathbb{R}^{m\times n}$, and the corresponding complex cases: $\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})$, $\mathbf{z}\in\mathbb{C}^n$ and $\mathbf{Z}\sim p_{\mathbf{Z}}(\mathbf{Z})$, $\mathbf{Z}\in\mathbb{C}^{m\times n}$. The CF of a complex random vector $\mathbf{z}$ is defined as [31], [32]

$$\Phi_{\mathbf{z}}(\boldsymbol{\omega}) = \mathrm{E}\left\{e^{\jmath\,\Re\mathrm{e}\{\boldsymbol{\omega}^H\mathbf{z}\}}\right\}, \quad \boldsymbol{\omega}\in\mathbb{C}^n \tag{22}$$

and the CF of a complex random matrix $\mathbf{Z}$ as

$$\Phi_{\mathbf{Z}}(\boldsymbol{\Omega}) = \mathrm{E}\left\{e^{\jmath\,\Re\mathrm{e}\{\mathrm{tr}(\boldsymbol{\Omega}^H\mathbf{Z})\}}\right\}, \quad \boldsymbol{\Omega}\in\mathbb{C}^{m\times n} \tag{23}$$

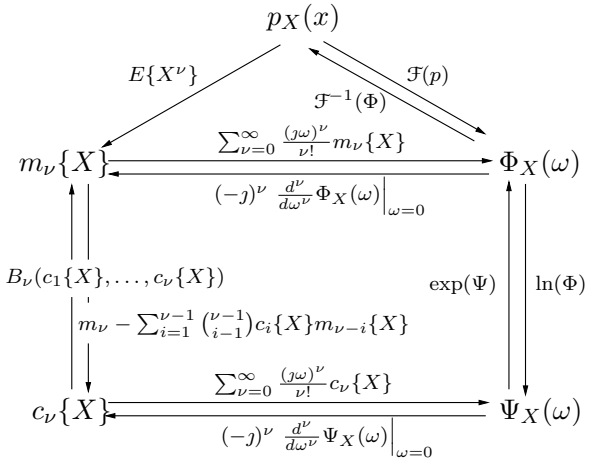

Fig. 1. Relations in univariate classical statistics.

where $\mathrm{tr}(\cdot)$ is the trace operator and $\Re\mathrm{e}\{\cdot\}$ extracts the real part of a complex expression, while the vector $\boldsymbol{\omega}$ and matrix $\boldsymbol{\Omega}$ are transform variables. We note that the CFs are defined in terms of the standard complex vector inner product $\boldsymbol{\omega}^H\mathbf{z}$ and complex matrix inner product $\mathrm{tr}(\boldsymbol{\Omega}^H\mathbf{Z})$ [31].

The νth-order moments of $\mathbf{z}$ and $\mathbf{Z}$ are retrieved from

$$m_\nu\{\mathbf{z}\} = (-\jmath)^\nu\frac{\partial^\nu}{\partial\boldsymbol{\omega}^\nu}\Phi_{\mathbf{z}}(\boldsymbol{\omega})\bigg|_{\boldsymbol{\omega}=\mathbf{0}} \tag{24}$$

$$m_\nu\{\mathbf{Z}\} = (-\jmath)^\nu\frac{\partial^\nu}{\partial\boldsymbol{\Omega}^\nu}\Phi_{\mathbf{Z}}(\boldsymbol{\Omega})\bigg|_{\boldsymbol{\Omega}=\mathbf{0}} \tag{25}$$

where $\partial^\nu/\partial\boldsymbol{\omega}^\nu$ and $\partial^\nu/\partial\boldsymbol{\Omega}^\nu$ is multi-index notation for sequential νth-order partial differentiation with respect to all elements in $\boldsymbol{\omega}$ and $\boldsymbol{\Omega}$, respectively. Non-scalar moments and cumulants of real random vectors and matrices are defined in [33], and the theory can be extended to the complex case, but this is outside the scope of our work.

B. Univariate Mellin Kind Statistics

The MT of the real valued function $f(x)$ defined on $\mathbb{R}^+$ is given as [1], [34]

$$F(s) = \mathcal{M}\{f(x)\}(s) = \int_0^\infty x^{s-1}f(x)\,dx \tag{26}$$

where the transform variable $s\in\mathbb{C}$. Under certain restrictions on $f(x)$, $F(s)$ is analytic in a strip parallel to the imaginary axis. The inverse MT is

$$\begin{aligned} f(x) &= \mathcal{M}^{-1}\{F(s)\}(x) \\ &= \frac{1}{2\pi\jmath}\int_{a-i\cdot\infty}^{a+i\cdot\infty} x^{-s}F(s)\,ds\,. \end{aligned} \tag{27}$$

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

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

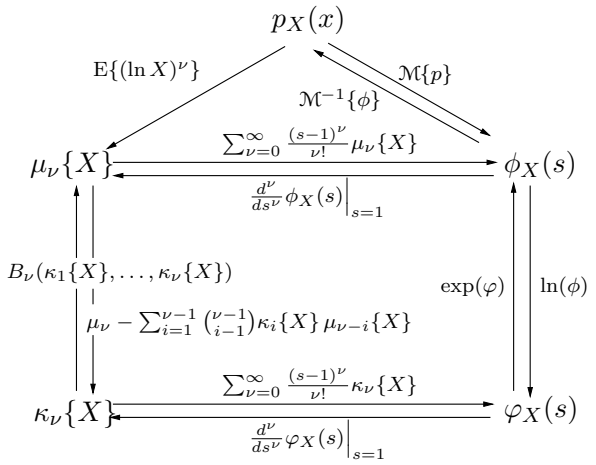

Fig. 2. Relations in univariate Mellin kind statistics.

The integration limits denote a line integral along any line $s = a \in \mathbb{R}$ parallel to the imaginary axis, which must lie within the analytic strip of $F(s)$.

Nicolas [2] proposed to replace the FT with the MT in the definition of the CF for the RV X, thus defining the Mellin kind CF as

$$\phi_X(s) = \mathrm{E}\{X^{s-1}\} = \mathcal{M}\{p_X(x)\}(s)\,. \tag{28}$$

The domain of the MT restricts this definition to positive RVs $X \in \mathbb{R}^+$. By expanding $\phi_X(s)$ as in the classical case, we obtain

$$\begin{aligned} \phi_X(s) &= \int_0^\infty e^{(\ln x)(s-1)} p_X(x)\, dx \\ &= \sum_{\nu=0}^{\infty} \frac{(s-1)^\nu}{\nu!} \mu_\nu\{X\} \end{aligned} \tag{29}$$

with the νth-order Mellin kind moment defined as

$$\begin{aligned} \mu_\nu\{X\} &= \mathrm{E}\{(\ln X)^\nu\} \\ &= \int_0^\infty (\ln x)^\nu p_X(x)\, dx\,. \end{aligned} \tag{30}$$

The derivation of (29) reveals that $\phi_X(s)$ is a power series expansion of the terms $\mu_\nu\{X\}$, appropriately termed log-moments, when they exist. When pursuing the analogy with the classical case, it is found that

$$\mu_\nu\{X\} = \frac{d^\nu}{ds^\nu} \phi_X(s)\Big|_{s=1}\,. \tag{31}$$

The Mellin kind CGF is further defined as $\varphi_X(s) = \ln \phi_X(s)$, from which the Mellin kind cumulants, also known as log-cumulants, can be retrieved as

$$\kappa_\nu\{X\} = \frac{d^\nu}{ds^\nu} \varphi_X(s)\Big|_{s=1} \tag{32}$$

given that the corresponding log-moment exists. When all log-cumulants exist, the Mellin kind CGF can be expanded as

$$\varphi_X(s) = \sum_{\nu=0}^{\infty} \frac{(s-1)^\nu}{\nu!} \kappa_\nu\{X\}\,. \tag{33}$$

The relation between the Mellin kind CF and CGF is the same as in the classical case, hence so is the relation between the log-moments and log-cumulants. Figure 2 summarises all relations for the univariate MKS.

Nicolas derived the MKS for the gamma, the inverse gamma, and the Fisher-Snedecor distribution [2], [3], among others. These results are listed in Table II for the unit mean version of these distributions.

C. Complex Matrix-Variate Mellin Kind Statistics

Mathai proposed a generalised transform (that he named the M-transform) for matrix-valued functions in [35], and referred to it in [36] as a generalised

MT. The complex version relevant to our study was presented in [21].

Definition 1 (Complex matrix-variate MT): Let $f(\mathbf{Z})$ be a real-valued scalar function defined on a cone of $d \times d$ Hermitian matrices that are either positive definite, negative definite, or null, and let f be symmetric in the sense $f(\mathbf{ZV}) = f(\mathbf{VZ})$, where $\mathbf{V}$ and $\mathbf{Z}$ belong to the same cone. The *complex matrix-variate MT* is then defined as

$$\mathcal{M}\{f(\mathbf{Z})\}(s) = \int_{\Omega_+} |\mathbf{Z}|^{s-d} f(\mathbf{Z})\, d\mathbf{Z} \tag{34}$$

with transform variable $s \in \mathbb{C}$, whenever the integral exists.

It is duly noted in [35] that since $\mathcal{M}\{f(\mathbf{Z})\}(s)$ is a function of the complex scalar transform variable s, whereas $f(\mathbf{Z})$ is defined on a matrix space, the transform is not unique. This problem is not associated with the multivariate MT defined in [37] as

$$\begin{aligned}&\mathcal{M}\{f(\mathbf{z})\}(s_1, \ldots, s_d)\\ &\quad = \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{d} z_i^{s_i - 1} f(\mathbf{z})\, dz_1 \cdots\, dz_d\end{aligned} \tag{35}$$

with $\mathbf{z} = [z_1, \ldots, z_d]^T \in \mathbb{C}^d$ and f defined on $\mathbb{C}^d$, which can in principle be extended to the matrix-variate case. Nevertheless, we shall refer to (34) as the matrix-variate MT.

The symmetry requirement in the definition of the matrix-variate MT restricts in theory the range of PDFs it can be applied to. In practice, however, it does not pose any problems for the compound Wishart type distributions used for multilook polarimetric radar data. In these functions (See Table I), the matrix variable $\mathbf{Z}$ occurs inside determinant and trace operators that are symmetric themselves, hence the overall PDFs are also symmetric in the required sense. We may therefore use the transform to define MKS for the complex matrix-variate case.

Definition 2 (Mellin kind matrix-variate CF): The Mellin kind CF of the complex random matrix $\mathbf{Z}$ is defined as

$$\phi_{\mathbf{Z}}(s) = \mathrm{E}\{|\mathbf{Z}|^{s-d}\} = \mathcal{M}\{p_{\mathbf{Z}}(\mathbf{Z})\}(s) \tag{36}$$

when $\mathbf{Z}$ and $p_{\mathbf{Z}}(\mathbf{Z})$ satisfy all requirements of the complex matrix-variate MT.

Definition 3 (Mellin kind matrix moments): The νth-order *Mellin kind matrix moment* of $\mathbf{Z}$ is defined as

$$\mu_\nu\{\mathbf{Z}\} = \left.\frac{d^\nu}{ds^\nu} \phi_{\mathbf{Z}}(s)\right|_{s=d}. \tag{37}$$

If all Mellin kind matrix moments exist, the Mellin kind CF can be written as the power series expansion

$$\begin{aligned}\phi_{\mathbf{Z}}(s) &= \int_{\Omega_+} e^{(s-d)\ln|\mathbf{Z}|} p_{\mathbf{Z}}(\mathbf{Z})\, d\mathbf{Z}\\ &= \sum_{\nu=0}^{\infty} \frac{(s-d)^\nu}{\nu!} \mu_\nu\{\mathbf{Z}\}\end{aligned} \tag{38}$$

in terms of the $\mu_\nu\{\mathbf{Z}\}$. The derivation of (38) reveals that

$$\begin{aligned}\mu_\nu\{\mathbf{Z}\} &= \mathrm{E}\{(\ln|\mathbf{Z}|)^\nu\}\\ &= \int_{\Omega_+} (\ln|\mathbf{Z}|)^\nu p_{\mathbf{Z}}(\mathbf{Z})\, d\mathbf{Z}\end{aligned} \tag{39}$$

which justifies the denotation of $\mu_\nu\{\mathbf{Z}\}$ as a *matrix log-moment* (MLM).

Definition 4 (Mellin kind matrix-variate CGF): The Mellin kind CGF of the complex random matrix $\mathbf{Z}$ is defined as

$$\varphi_{\mathbf{Z}}(s) = \ln \phi_{\mathbf{Z}}(s)\,. \tag{40}$$

Definition 5 (Mellin kind matrix cumulants): The νth-order *Mellin kind matrix cumulant* of $\mathbf{Z}$ is defined as

$$\kappa_\nu\{\mathbf{Z}\} = \left.\frac{d^\nu}{ds^\nu} \varphi_{\mathbf{Z}}(s)\right|_{s=d}. \tag{41}$$

When all Mellin kind matrix moments exist, so do the Mellin kind matrix cumulants, and the Mellin kind CGF can be expanded as

$$\varphi_{\mathbf{Z}}(s) = \ln \phi_{\mathbf{Z}}(s) = \sum_{\nu=0}^{\infty} \frac{(s-d)^\nu}{\nu!} \kappa_\nu\{\mathbf{Z}\} \tag{42}$$

in terms of the $\kappa_\nu\{\mathbf{Z}\}$, that are also called *matrix log-cumulants* (MLCs).

As we see, there is a complete analogy with the MKS derived in Section III-B for the univariate case, as summarised in Figure 3. The Mellin kind matrix-variate CF and CGF are related by the same logarithmic transformation as in the univariate case. Thus, the conversion between MLMs and MLCs is also given in terms of Faá di Bruno's formula and the complete Bell polynomial, by analogy with (20) and (21).

Example (Complex Wishart distribution): The Mellin kind CF of a complex Wishart matrix $\mathbf{W} \sim \mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma})$ is derived in Appendix A as

$$\begin{aligned}\phi_{\mathbf{W}}(s) &= \mathcal{M}\{p_{\mathbf{W}}(\mathbf{W})\}(s) \\ &= \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} |\mathbf{\Sigma}|^{s-d} .\end{aligned} \tag{43}$$

The MLCs are found to be

$$\kappa_1\{\mathbf{W}\} = \psi_d^{(0)}(L) + \ln|\mathbf{\Sigma}| \tag{44}$$

$$\kappa_\nu\{\mathbf{W}\} = \psi_d^{(\nu-1)}(L) , \quad \nu > 1 \tag{45}$$

where we introduce the νth-order *multivariate polygamma function* as

$$\psi_d^{(\nu)}(L) = \sum_{i=0}^{d-1} \psi^{(\nu)}(L-i) . \tag{46}$$

This is a convenient extension of the ordinary polygamma functions, defined as the logarithmic derivatives of the gamma function:

$$\psi^{(\nu)}(L) = \frac{d^{\nu+1} \ln\Gamma(L)}{dL^{\nu+1}} , \quad \nu \geq 0 . \tag{47}$$

Let $\tilde{\mathbf{W}} = \mathbf{W}/L$ be the scaled Wishart matrix whose PDF is given in (9). The MLCs of $\tilde{\mathbf{W}}$ are derived as

$$\kappa_1\{\tilde{\mathbf{W}}\} = \psi_d^{(0)}(L) + \ln|\mathbf{\Sigma}| - d\ln L \tag{48}$$

$$\kappa_\nu\{\tilde{\mathbf{W}}\} = \psi_d^{(\nu-1)}(L) , \quad \nu > 1 . \tag{49}$$

The MLMs of $\mathbf{W}$ and $\tilde{\mathbf{W}}$ can be found by inserting the MLCs into the equivalent formula of (21).

Log-statistics of $\mathbf{W}$ and $\tilde{\mathbf{W}}$ were first derived in [38] without utilising the Mellin transform, and not for a general order ν. They were also used in [39], but interpreted as log-moments and log-cumulants of the positive scalar RV $|\mathbf{W}|$ rather than of the matrix $\mathbf{W}$. A detailed derivation is given in Appendix A.

D. Some Properties of the Matrix-Variate Product Model

We shall now look at some fundamental properties of the MT which makes it a natural replacement of the FT when working with a multiplicative signal model, and extend this exposition to the complex matrix-variate case.

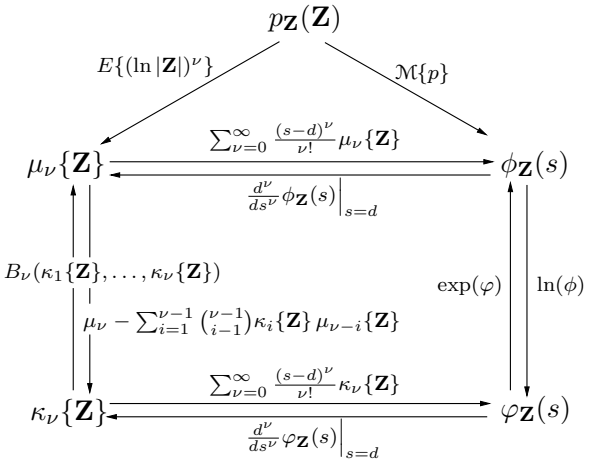

Fig. 3. Relations in matrix-variate Mellin kind statistics.

1) Univariate Additive Model: Let X, U and V be real scalar RVs whose moments and cumulants all exist, and assume that U and V are statistically independent. For the additive stochastic signal model,

$$X = U + V \tag{50}$$

we find that the CF, the CGF, and the cumulants of X, as defined in the classical case using the FT, can be written as

$$\Phi_X(\omega) = \Phi_U(\omega) \cdot \Phi_V(\omega) \tag{51}$$

$$\Psi_X(\omega) = \Psi_U(\omega) + \Psi_V(\omega) \tag{52}$$

$$c_\nu\{X\} = c_\nu\{U\} + c_\nu\{V\} . \tag{53}$$

These relations also hold when the signal model is generalised to the multivariate, matrix-variate, and complex case. The PDF of X is given by the convolution

$$\begin{aligned}p_X(x) &= (p_U * p_V)(x) \\ &= \int_{-\infty}^{+\infty} p_U(u) p_V(x-u)\, du\end{aligned} \tag{54}$$

where $*$ denotes the convolution operator, which corresponds to a multiplication in the FT domain, as seen in (51).

2) Univariate Product Model: Now consider the stochastic product model:

$$X = U \cdot V, \tag{55}$$

with the additional constraints that X, U, and $V \in \mathbb{R}^+$. Observe that the Mellin kind CF, CGF and

cumulants defined with the MT take the form

$$\phi_X(s) = \phi_U(s) \cdot \phi_V(s) \tag{56}$$
$$\varphi_X(s) = \varphi_U(s) + \varphi_V(s) \tag{57}$$
$$\kappa_\nu\{X\} = \kappa_\nu\{U\} + \kappa_\nu\{V\}\,. \tag{58}$$

The usual interpretation [2] is to view the MT as a Laplace transform computed on a logarithmic scale. The logarithmic transformation translates the product model into an additive one, which explains why the Mellin kind CF inherits the multiplicative property of the classical CF, whereas the CGF and the cumulants take over the additive property from their classical counterparts.

The PDF of X can be found from

$$\begin{aligned} p_X(x) &= \int_0^\infty p_{X|V}(x|v)\, p_V(v)\, dv \\ &= \int_0^\infty p_U\Big(\frac{x}{v}\Big)\, p_V(v)\, \frac{dv}{v} \end{aligned} \tag{59}$$

which is known as the Mellin convolution. The operation is denoted $p_X(x) = (p_U \mathbin{\hat{\star}} p_V)(x)$. The product in (56) is the MT domain equivalent.

3) Matrix-variate Product Model: Before we are ready to consider the *matrix-variate product model*, we shall establish a matrix-variate Mellin convolution theorem using the matrix-variate MT of Definition 1. We also find it natural to include some closely related correlation theorems. We start by defining the *matrix-variate Mellin convolution*.

Definition 6 (Matrix-variate Mellin convolution): Let $f(\mathbf{U})$ and $g(\mathbf{U})$ be two functions defined on the cone of positive definite (or negative definite) complex Hermitian matrices. Further assume that $\mathbf{U}$ and $\mathbf{V}$ both belong to the domain of f and g, and that the functions are symmetric in the sense that $f(\mathbf{UV}) = f(\mathbf{VU})$. We define the matrix-variate Mellin convolution of f and g as

$$\begin{aligned} &(f \mathbin{\hat{\star}} g)(\mathbf{U}) \\ &\quad= \int_{\Omega_+} |\mathbf{V}|^{-d} f(\mathbf{V}^{-\frac{1}{2}}\mathbf{U}\mathbf{V}^{-\frac{1}{2}}) g(\mathbf{V})\, d\mathbf{V} \\ &\quad= \int_{\Omega_+} |\mathbf{V}|^{-d} g(\mathbf{V}^{-\frac{1}{2}}\mathbf{U}\mathbf{V}^{-\frac{1}{2}}) f(\mathbf{V})\, d\mathbf{V} \end{aligned} \tag{60}$$

which is an associative and commutative operation.

Theorem 1 (Matrix-variate Mellin convolution): Under the assumptions presented in Definition 6, then

$$\begin{aligned} &\mathcal{M}\{(f \mathbin{\hat{\star}} g)(\mathbf{U})\}(s) \\ &\qquad= \mathcal{M}\{f(\mathbf{U})\}(s) \cdot \mathcal{M}\{g(\mathbf{U})\}(s)\,. \end{aligned} \tag{61}$$

Proof: Introduce the substitution $\mathbf{X} = \mathbf{UV}$ and note that $\mathbf{X}$ must belong to the same matrix space as $\mathbf{U}$ and $\mathbf{V}$. Furthermore, we have $\mathbf{U} = \mathbf{V}^{-\frac{1}{2}}\mathbf{X}\mathbf{V}^{-\frac{1}{2}}$ and $d\mathbf{U} = d\mathbf{X}/|\mathbf{V}|^d$ [21]. This yields

$$\begin{aligned} &\mathcal{M}\{f(\mathbf{U})\}(s) \cdot \mathcal{M}\{g(\mathbf{V})\}(s) \\ &= \int_{\Omega_+} (|\mathbf{X}|/|\mathbf{V}|)^{s-d} f(\mathbf{V}^{-\frac{1}{2}}\mathbf{X}\mathbf{V}^{-\frac{1}{2}}) |\mathbf{V}|^{-d}\, d\mathbf{X} \\ &\qquad \times \int_{\Omega_+} |\mathbf{V}|^{s-d} g(\mathbf{V})\, d\mathbf{V} \\ &= \int_{\Omega_+} |\mathbf{X}|^{s-d} \Big[\int_{\Omega_+} |\mathbf{V}|^{-d} f(\mathbf{V}^{-\frac{1}{2}}\mathbf{X}\mathbf{V}^{-\frac{1}{2}}) g(\mathbf{V})\, d\mathbf{V} \Big] d\mathbf{X} \\ &= \mathcal{M}\{(f \mathbin{\hat{\star}} g)(\mathbf{X})\}(s) \end{aligned} \tag{62}$$

where in the last transition, we identify the term in the square brackets as the Mellin convolution. ■

We have shown that the MT provides a convolution theorem for the product model, like the FT does for the additive model, and that this extends to matrix-variate theory. By further analogy with the Fourier transform, the Mellin transform also has a correlation theorem. The following operation reduces in the univariate case to the Mellin correlation, as Nicolas defines it in [2], [3].

Definition 7 (Type I matrix-variate Mellin correlation): Under the assumptions presented in Definition 6, we define the *type I matrix-variate Mellin correlation* of f and g as

$$\begin{aligned} &(f \mathbin{\hat{\otimes}} g)(\mathbf{U}) \\ &\qquad= \int_{\Omega_+} |\mathbf{V}|^{d} f(\mathbf{V}^{\frac{1}{2}}\mathbf{U}\mathbf{V}^{\frac{1}{2}}) g(\mathbf{V})\, d\mathbf{V}\,. \end{aligned} \tag{63}$$

This operation is neither associative nor commutative.

Theorem 2 (Type I matrix-variate Mellin correlation): Under the assumptions presented in Definition 6, then

$$\begin{aligned} &\mathcal{M}\{(f \mathbin{\hat{\otimes}} g)(\mathbf{U})\}(s) \\ &\qquad= \mathcal{M}\{f(\mathbf{U})\}(s) \cdot \mathcal{M}\{g(U)\}(2d - s)\,. \end{aligned} \tag{64}$$

The proof is given in [40].

We also present an alternative definition. It reduces in the univariate case to a relation often referred to as a Mellin correlation (See e.g. [41]).

Definition 8 (Type II matrix-variate Mellin correlation): Under the assumptions presented in Definition 6, we define the *type II matrix-variate Mellin correlation* of f and g as

$$(f \mathbin{\hat{\circledast}} g)(\mathbf{U}) = \int_{\Omega_+} |\mathbf{V}|^{-d} f(\mathbf{V}^{\frac{1}{2}}\mathbf{U}\mathbf{V}^{\frac{1}{2}}) g(\mathbf{V})\, d\mathbf{V}\,. \tag{65}$$

This is neither an associative nor commutative operation.

Theorem 3 (Type II matrix-variate Mellin correlation): Under the assumptions presented in Definition 6, then

$$\mathcal{M}\{(f \mathbin{\hat{\circledast}} g)(\mathbf{U})\}(s) = \mathcal{M}\{f(\mathbf{U})\}(s) \cdot \mathcal{M}\{g(-\mathbf{U})\}(s)\,. \tag{66}$$

The proof follows in the same manner as for the Theorem 1, and is therefore omitted. A more general theorem, which reduced to both Theorems 2 and 3 was presented in [21, Th. 6.2].

We will now explain the relevance to a matrix-variate product model expressed by the (ordinary) matrix product

$$\mathbf{X} = \mathbf{U}\,\mathbf{V} \tag{67}$$

where $\mathbf{X}$, $\mathbf{U}$, and $\mathbf{V}$ are complex and positive definite Hermitian matrices. With the same approach as in (59) we establish that the PDF of $\mathbf{X}$ is

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{X}) &= \int_{\Omega_+} p_{\mathbf{X}|\mathbf{V}}(\mathbf{X}|\mathbf{V}) p_{\mathbf{V}}(\mathbf{V})\, d\mathbf{V} \\ &= \int_{\Omega_+} |\mathbf{V}|^{-d} p_{\mathbf{U}}(\mathbf{V}^{-\frac{1}{2}}\mathbf{X}\mathbf{V}^{-\frac{1}{2}}) p_{\mathbf{V}}(\mathbf{V})\, d\mathbf{V}\,. \end{aligned} \tag{68}$$

This is exactly $(p_{\mathbf{U}} \mathbin{\hat{\star}} p_{\mathbf{V}})(\mathbf{X})$, as should be expected from (62), which justifies the definition of the matrix-variate Mellin convolution. Assume that all MLMs and MLCs of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{X}$ exist, as given in Definition 3 and 5. It follows from Theorem 1 that

$$\phi_{\mathbf{X}}(s) = \phi_{\mathbf{U}}(s) \cdot \phi_{\mathbf{V}}(s) \tag{69}$$

$$\varphi_{\mathbf{X}}(s) = \varphi_{\mathbf{U}}(s) + \varphi_{\mathbf{V}}(s) \tag{70}$$

$$\kappa_\nu\{\mathbf{X}\} = \kappa_\nu\{\mathbf{U}\} + \kappa_\nu\{\mathbf{V}\}\,. \tag{71}$$

in the matrix-variate case.

Example (Multilook polarimetric product model): We now return to the multilook polarimetric product model for the radar data covariance matrix $\mathbf{Z} = L\mathbf{C}$. The model can be written as: $\mathbf{Z} = \mathbf{T}\mathbf{W}$, with $\mathbf{T} = T\,\mathbf{I}_d$, where T is the texture RV and $\mathbf{I}_d$ is the $d \times d$ identity matrix. We note that the matrix $\mathbf{T}$ contains only one functionally independent entry, namely T. Without entering the stringent argument in terms of differential calculus, we state that an integral $\int_{\Omega_+} f(|\mathbf{T}|)\, d\mathbf{T}$ can be replaced with $\int_{\mathbb{R}^+} f(T^d)\, dT$. We thus have

$$\begin{aligned} \phi_{\mathbf{T}}(s) &= \int_{\Omega_+} |\mathbf{T}|^{s-d} p_{\mathbf{T}}(\mathbf{T})\, d\mathbf{T} \\ &= \int_0^\infty t^{d(s-d)} p_T(t)\, dt \\ &= \sum_{\nu=0}^{\infty} \frac{[d(s-d)]^\nu}{\nu!} \mu_\nu\{T\} \\ &= \phi_T(d(s-d)+1)\,. \end{aligned} \tag{72}$$

This implies that

$$\mu_\nu\{\mathbf{T}\} = \frac{d^\nu}{ds^\nu}\phi_{\mathbf{T}}(s)\bigg|_{s=d} = d^\nu \mu_\nu\{T\}\,. \tag{73}$$

Faá di Bruno's formula is used to prove

$$\kappa_\nu\{\mathbf{T}\} = d^\nu \kappa_\nu\{T\} \tag{74}$$

and the matrix-variate version of the formula yields the MLCs of $\mathbf{Z}$ as

$$\kappa_\nu\{\mathbf{Z}\} = \mu_\nu\{\mathbf{Z}\} - \sum_{i=1}^{\nu-1} \binom{\nu-1}{i-1} \kappa_i\{\mathbf{Z}\}\mu_{\nu-i}\{\mathbf{Z}\}\,. \tag{75}$$

The first MLCs are expressed as

$$\kappa_1\{\mathbf{Z}\} = \mu_1\{\mathbf{Z}\} \tag{76}$$

$$\kappa_2\{\mathbf{Z}\} = \mu_2\{\mathbf{Z}\} - \mu_1^2\{\mathbf{Z}\} \tag{77}$$

$$\kappa_3\{\mathbf{Z}\} = \mu_3\{\mathbf{Z}\} - 3\mu_1\{\mathbf{Z}\}\mu_2\{\mathbf{Z}\} + 2\mu_1^3\{\mathbf{Z}\}\,. \tag{78}$$

We use $\mathbf{Z} = \mathbf{T}\mathbf{W}$ and (71) to prove that

$$\kappa_\nu\{\mathbf{Z}\} = \kappa_\nu\{\mathbf{W}\} + d^\nu \kappa_\nu\{T\} \tag{79}$$

but, since the observable of multilook polarimetric radar is $\mathbf{C} = \mathbf{Z}/L = T\tilde{\mathbf{W}}$, we are more interested in the $\kappa_\nu\{\mathbf{C}\}$. With the $\kappa_\nu\{\mathbf{W}\}$ expanded, the MLCs of $\mathbf{C}$ evaluate under the product model to

$$\kappa_1\{\mathbf{C}\} = \psi_d^{(0)}(L) + \ln|\mathbf{\Sigma}| - d(\ln L - \kappa_1\{T\}) \tag{80}$$

$$\kappa_\nu\{\mathbf{C}\} = \psi_d^{(\nu-1)}(L) + d^\nu \kappa_\nu\{T\}\,, \quad \nu > 1 \tag{81}$$

for a general texture variable with unspecified distribution. To obtain the MLCs of a specific distribution, the texture variable log-cumulants $\kappa_\nu\{T\}$, such as those listed in Table II, must be inserted.

We finally note that sample MLMs, denoted $\langle\mu_\nu\{\mathbf{C}\}\rangle$, are calculated with the simple sample mean estimator

$$\langle\mu_\nu\{\mathbf{C}\}\rangle = \frac{1}{N}\sum_{i=1}^{N}(\ln|\mathbf{C}_i|)^\nu \tag{82}$$

given a sample of N covariance matrices: $\mathscr{C} = \{\mathbf{C}_i\}_{i=1}^N$. Sample MLCs $\langle\kappa_\nu\{\mathbf{C}\}\rangle$ are obtained from (75) by combining sample MLMs instead of theoretical MLMs.

IV. Applications

In this section we discuss application of matrix-variate MKS. Before presenting specific examples of MoMLC algorithms for parameter estimation and demonstrating their effectiveness, we introduce the *MLC diagram*. The MLC diagram is not an application in its own right, but serves as a visualisation tool, which efficiently explains some uses of MKS and provides intuition about the MoMLC. It is a straight-forward extension of the log-cumulant diagrams used by Nicolas [2], [3] for univariate MKS.

A. Matrix Log-Cumulant Diagrams

The MLC diagram generally displays a q-dimensional space where each dimension represents one particular MLC with unique order ν. Let $\nu_1, \cdots, \nu_q$ be the orders of the chosen MLCs. In this MLC space, we plot: (i) The manifolds spanned by the theoretical MLCs that can be attained under given models, and (ii) points that represent the empirical sample MLCs computed from data samples.

Define $\boldsymbol{\vartheta}$ as the vector that contains all texture parameters of a certain distribution model. Thus, $\boldsymbol{\vartheta}_\mathcal{K} = [\alpha]$, $\boldsymbol{\vartheta}_{\mathcal{G}^0} = [\lambda]$ and $\boldsymbol{\vartheta}_\mathcal{U} = [\xi, \zeta]^T$ are the respective texture parameter vectors of the $\mathcal{K}$, $\mathcal{G}^0$ and $\mathcal{U}$ distribution. Assume that the parameters L and $\boldsymbol{\Sigma}$ are fixed, such that the theoretical MLCs only vary through $\boldsymbol{\vartheta}$. The MLC space manifold spanned by a general model is denoted

$$\mathcal{M}(\boldsymbol{\vartheta}) = \left\{\left(\kappa_{\nu_1}(\boldsymbol{\vartheta}), \kappa_{\nu_2}(\boldsymbol{\vartheta}), \cdots, \kappa_{\nu_q}(\boldsymbol{\vartheta})\right)\right\} \tag{83}$$

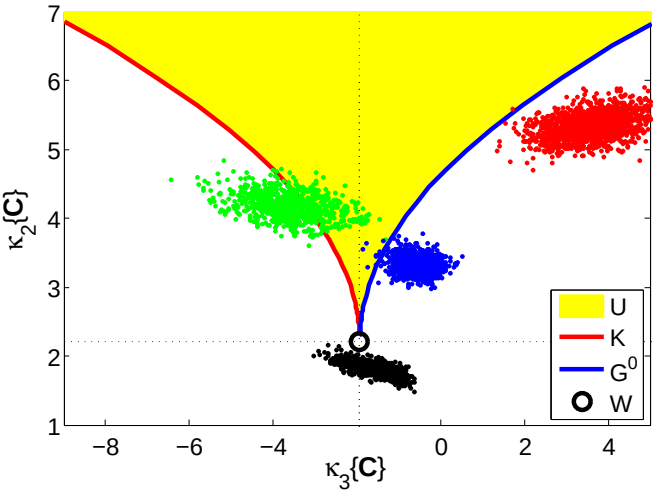

Fig. 4. Matrix log-cumulant diagram showing the manifolds of theoretical MLCs for the complex Wishart, $\mathcal{K}$, $\mathcal{G}^0$ and $\mathcal{U}$ distribution, as well as a collection of sample MLCs representing forest (green), ocean (blue), urban area (red) and a wheat crop (black).

where we have changed the notation of the theoretical νth-order MLC from $\kappa_\nu\{\mathbf{C}\}$ to $\kappa_\nu(\boldsymbol{\vartheta})$ to emphasise that the points that constitute the manifold are functions of $\boldsymbol{\vartheta}$. The dimension of the manifold $\mathcal{M}(\boldsymbol{\vartheta})$ is the same as the dimension of the vector $\boldsymbol{\vartheta}$, that is, the number of texture parameters. We next define the vector of sample MLCs as

$$\langle\boldsymbol{\kappa}(\mathscr{C})\rangle = \left[\langle\kappa_{\nu_1}(\mathscr{C})\rangle, \cdots, \langle\kappa_{\nu_q}(\mathscr{C})\rangle\right] \tag{84}$$

and note that the sample MLCs have also been given a new notation, $\langle\kappa_\nu(\mathscr{C})\rangle$, to stress that they are computed from the data sample $\mathscr{C}$.

Like Nicolas [2], [3], we concentrate on diagrams that plot the third-order log-cumulant against the second-order log-cumulant. We have shown that under the polarimetric product model, MLCs of order two and higher are independent of the scale matrix $\boldsymbol{\Sigma}$. Assuming that the equivalent number of looks, L, is a global constant for the data set, this diagram is of particular interest since it displays the solitary impact of the texture parameters upon the models. Therefore, it also provides valuable insight about how the texture parameters can be estimated.

As seen in Figure 4, the manifolds of our selected distribution models have different dimensions. The Wishart distribution has no texture parameters and is therefore represented by a point (black circle), which can be viewed as a zero-dimensional manifold. The texture of the $\mathcal{K}$ and $\mathcal{G}^0$ distribution is parametrised by one parameter, thus they are represented by a curve (red and blue, respectively), which

is a one-dimensional manifold. The $\mathcal{U}$ distribution has two texture parameters and is represented by a surface (yellow area), which is a two-dimensional manifold. This is valid also when plotting higher-dimensional tuples of MLCs, that is, for MLC spaces with dimension higher than two. The dashed coordinates are centred at the point which represents the Wishart distribution. The impact of nonzero texture on the MLCs is measured relative to these axes.

Given a sample $\mathscr{C} = \{\mathbf{C}_i\}_{i=1}^N$ of covariance matrices, we can compute the sample MLCs of order $\nu_1, \cdots, \nu_q$ and plot them as a point in MLC space. This has been done in Figure 4 for one forest sample (shown as green points) and one wheat crop sample (black) taken from a polarimetric NASA/JPL AIRSAR C-band image of Flevoland, The Netherlands. We have also plotted sample MLCs computed from an ocean sample (blue) and an urban area sample (red) extracted from an image of San Francisco, United States, captured by the same sensor. Both images are from 1989. Multiple points have been obtained for each class by bootstrap sampling [42] of $\mathscr{C}$.

MoMLC parameter estimation can now be visualised as a projection of the sample MLCs onto the manifolds representing the models. The manifolds are functions of the texture parameters, and the parameter values at the projection point is assigned as an estimate. An estimator based on a single νth-order MLC relies on a projection in the direction normal to the νth-order coordinate. This is illustrated in Figure 5, displaying estimators for the $\mathcal{K}$ distribution texture parameter α based on the point $(\langle\kappa_3(\mathscr{C})\rangle, \langle\kappa_2(\mathscr{C})\rangle)$ in MLC space, which is shown as the black symbol '$\times$'. The estimators denoted $\hat{\alpha}(\langle\kappa_2\rangle)$ and $\hat{\alpha}(\langle\kappa_3\rangle)$ are based on the second-order and third-order MLC equation, respectively. The dashed arrows visualise their projection of the sample MLC point onto the red curve representing the $\mathcal{K}$ distribution. Note that an estimator requires at least as many sample MLCs as the number of texture parameters to be estimated. For instance, the $\mathcal{K}$ and $\mathcal{G}_0$ distribution require one, while the $\mathcal{U}$ distribution requires two sample MLCs. The estimators $\hat{\alpha}(\langle\kappa_2\rangle)$ and $\hat{\alpha}(\langle\kappa_3\rangle)$ use exactly the required number.

As indicated, it is possible to design an estimator which is based on more sample MLCs than there are texture parameters, which implies that more information about $\boldsymbol{\vartheta}$ is extracted and $\mathscr{C}$ is utilised more

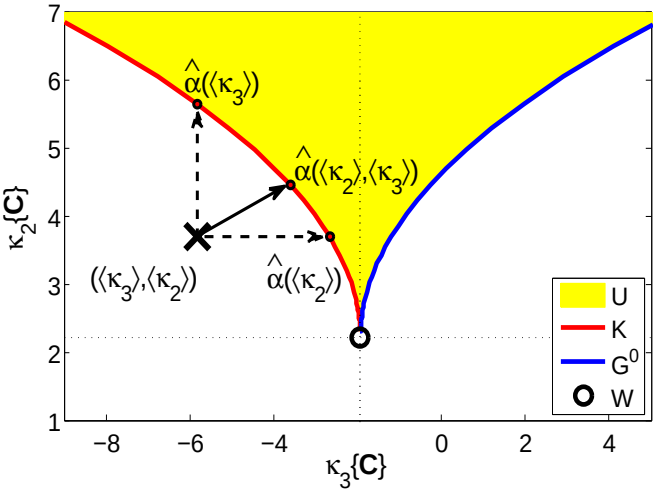

Fig. 5. MLC space interpretation of three estimators of the $\mathcal{K}$ distribution texture parameter α. The first estimator is based on $\langle\kappa_2\{\mathbf{C}\}\rangle$, the other is based on $\langle\kappa_3\{\mathbf{C}\}\rangle$, and the third is based on both.

efficiently. One way to combine the information contained in multiple sample MLCs is to derive a squared Mahalanobis distance (d_M^2) between the sample MLC point and the points on the model manifold. The estimation problem then reduces to a minimisation of the distance measure with respect to $\boldsymbol{\vartheta}$. To find an expression for d_M^2, we must derive the (approximate) mean values and covariance matrix of the sample MLCs.

The minimum of d_M^2 defines a new projection of the sample MLC point onto the model manifold, as illustrated by the solid arrow in Figure 5. The sample MLC point is projected onto the point on the $\mathcal{K}$ distribution curve that minimises d_M^2, and the associated value of α defines the estimate $\hat{\alpha}(\langle\kappa_2\rangle, \langle\kappa_3\rangle)$. We can see this estimate as a weighted mean of $\hat{\alpha}(\langle\kappa_2\rangle)$ and $\hat{\alpha}(\langle\kappa_3\rangle)$. The information content of an individual sample MLC is proportional to its precision (i.e., inverse variance), and determines its contribution to the overall estimate. The shape of the sample MLC clusters in Figure 4 shows that the sample variance increases with MLC order, as expected.

A detailed derivation of the Mahalanobis distance is given in [43], where we also discuss the coupling of the estimation problem and the problem of measuring goodness-of-fit (GoF) for distribution models. The geometrical interpretation of distances in MLC space in terms of model fit is intuitive. We also find it much easier to observe deviations between data and model in MLC space than by

comparing data histograms with model densities, which is the alternative normally resorted to in the literature. This point is highlighted in [39]. In [43] we derive the sample distribution of d_M^2, such that formal GoF testing can be performed, which conforms with visual inspection of model fit in the MLC diagram.

B. Parameter Estimation

In this section, we discuss MKS-based estimation algorithms for parameters of the distributions presented in Table I.

1) Equivalent number of looks: The equivalent number of looks, L, can be estimated from the first-order MLC equation of the complex Wishart distribution. This yields the maximum likelihood solution proposed in [38], [44]:

$$\psi_d^{(0)}(\hat{L}) - d\ln\hat{L} = \langle\kappa_1\{\mathbf{C}\}\rangle - \ln|\mathbf{\Sigma}| \tag{85}$$

where we must insert the first-order sample MLC and an estimate of $\mathbf{\Sigma}$ before solving for L by numerical methods. We use the maximum likelihood estimate of $\mathbf{\Sigma}$, defined as the sample mean of $\mathscr{C}$.

In principle, we can also solve for L from the MLC equations of the product model in (11) and avoid the Wishart constraint. However, these MLC equations contain texture parameters already from the first order, and all unknown parameters must therefore be estimated jointly from a system of equations. Higher-order MLCs can also be used to improve the estimator in (85). None of these approaches have been attempted in practice.

In the following, we shall assume that an estimate of L has been provided and treat it as a known constant.

2) Matrix-variate $\mathcal{K}$ Distribution: Under this distribution, the texture parameter α is related to the second-order MLC through

$$\kappa_2\{\mathbf{C}\} = d^2\psi^{(1)}(\alpha) + \psi_d^{(1)}(L) \tag{86}$$

and the estimate $\hat{\alpha}_{A_1}$ is obtained by solving

$$\psi^{(1)}(\hat{\alpha}_{A_1}) = \frac{\langle\kappa_2\{\mathbf{C}\}\rangle - \psi_d^{(1)}(L)}{d^2}\,. \tag{87}$$

Alternative estimators are proposed in Frery et al. [45] and Doulgeris et al. [46], where the former is just a mono-pol version of the latter. Doulgeris' estimator is

$$\hat{\alpha}_D = \frac{d(Ld+1)}{L\widehat{\mathrm{Var}}\{\tau\} - d} \tag{88}$$

where $\tau = \mathrm{tr}\left(\hat{\mathbf{\Sigma}}^{-1}\mathbf{C}\right)$. The derivation is shown in Appendix B. Another approach taken by Freitas and Frery et al. [25], [45] is to derive estimators from fractional moments of the mono-pol intensity C. By combining the half- and quarter-order moments they found that

$$\frac{\Gamma^2\left(\hat{\alpha}_F+\frac{1}{4}\right)}{\Gamma(\hat{\alpha}_F)\,\Gamma\left(\hat{\alpha}_F+\frac{1}{2}\right)}\frac{\Gamma^2\left(L+\frac{1}{4}\right)}{\Gamma(L)\,\Gamma\left(L+\frac{1}{2}\right)} - \frac{\langle C^{\frac{1}{4}}\rangle^2}{\langle C^{\frac{1}{2}}\rangle} = 0 \tag{89}$$

which can be solved for $\hat{\alpha}_F$. This method provides one estimate per polarimetric channel. The final estimate is an average of the mono-pol estimates.

Averaging over mono-pol estimates can also be carried out for the mono-pol version of $\hat{\alpha}_{A_1}$ (i.e., with $d = 1$), which is the estimator derived by Nicolas from univariate MKS [2], [3]. We denote this estimator as $\hat{\alpha}_N$ and include it in the comparison in order to quantify the gain of using the full polarimetric information contained in $\mathbf{C}$ with respect to the information contained in intensity channels only. On a historic note, we remark that the mono-pol MKS-based estimator of Nicolas was proposed earlier by Kreithen and Hogan [47] and Blacknell [48], although without relating it to Mellin transform theory.

The final estimator we present is the one we have proposed in [43] based on multiple MLCs, as discussed in Section IV-A. It is defined as

$$\hat{\alpha}_{A_2} = \arg\left\{\min_{\alpha}\left\{d_M^2\right\}\right\} \tag{90}$$

where the squared Mahalanobis distance

$$d_M^2 = \big(\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa}\big)^T\mathbf{K}^{-1}\big(\langle\boldsymbol{\kappa}\rangle - \boldsymbol{\kappa}\big) \tag{91}$$

contains the sample MLC vector $\langle\boldsymbol{\kappa}\rangle = [\langle\kappa_2\rangle, \langle\kappa_3\rangle]^T$, its mean vector $\boldsymbol{\kappa} = \mathrm{E}\{\langle\boldsymbol{\kappa}\rangle\} = [\kappa_2, \kappa_3]^T$, and the covariance matrix

$$\begin{aligned}\mathbf{K} &= \mathrm{Cov}\{\langle\boldsymbol{\kappa}\rangle\} \\ &= \begin{bmatrix} \kappa_4+2\kappa_2^2 & \kappa_5+6\kappa_2\kappa_3 \\ \kappa_5+6\kappa_2\kappa_3 & \kappa_6+9\kappa_2\kappa_4+9\kappa_3^2+6\kappa_2^3 \end{bmatrix}.\end{aligned} \tag{92}$$

The sample MLCs are fixed after a data sample is collected, and the minimisation is performed by varying $\boldsymbol{\kappa}$ and $\mathbf{K}$, that both depend on α through the theoretical MLCs.

Figure 6 and 7 show bias and variance of all estimators, obtained from Monte Carlo simulations with $L = 10$ and $\alpha = 10$. They clearly show that the estimators based on the full polarimetric covariance matrix ($\hat{\alpha}_{A_1}$, $\hat{\alpha}_D$ and $\hat{\alpha}_{A2}$) outperform those based

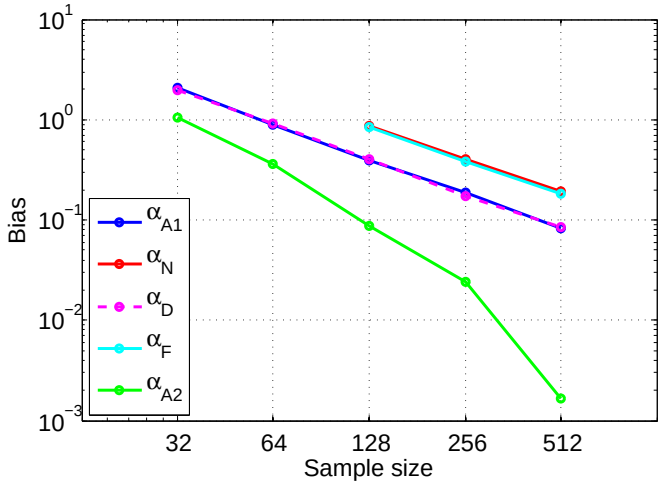

Fig. 6. Bias of estimators for the $\mathcal{K}$ distribution texture parameter α as function of sample size N.

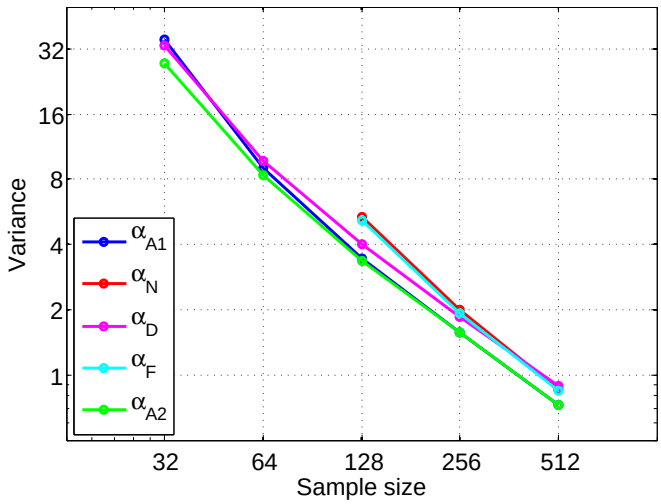

Fig. 7. Variance of estimators for the $\mathcal{K}$ distribution texture parameter α as function of sample size N.

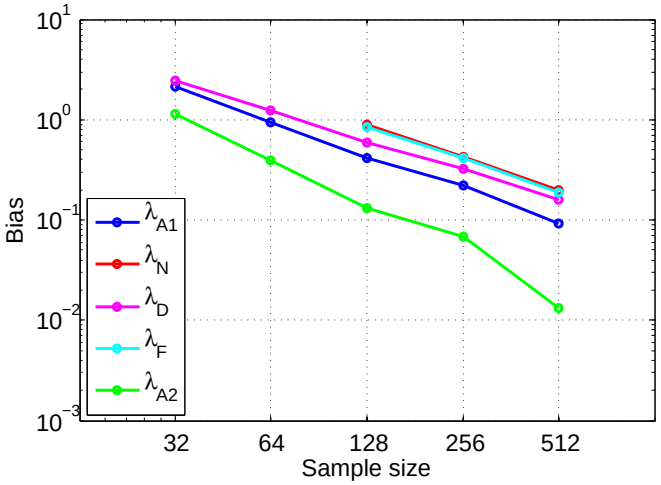

Fig. 8. Bias of estimators for the $\mathcal{G}^0$ distribution texture parameter λ.

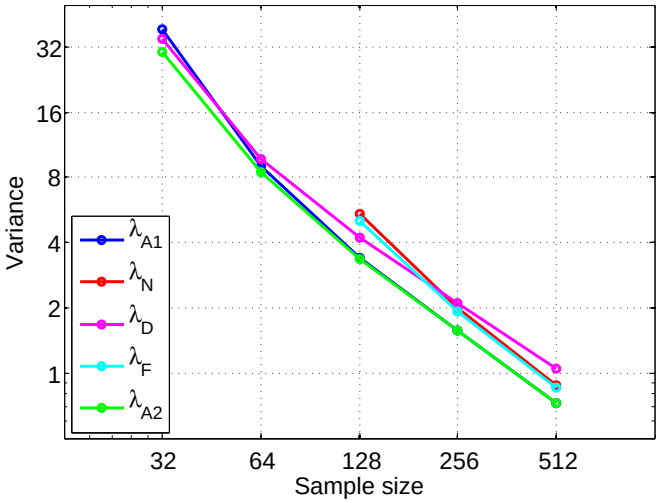

Fig. 9. Variance of estimators for the $\mathcal{G}^0$ distribution texture parameter λ.

on intensities only ($\hat{\alpha}_N$ and $\hat{\alpha}_F$), both in terms of bias and variance. From the latter group, $\hat{\alpha}_F$ ranks slightly better than $\hat{\alpha}_N$. For the truly polarimetric estimators, we see that $\hat{\alpha}_{A_2}$ has the superior bias properties, while the bias of $\hat{\alpha}_{A_1}$ and $\hat{\alpha}_D$ is very similar. Estimator $\hat{\alpha}_{A_2}$ is best also when it comes to variance, but is approached by $\hat{\alpha}_{A_1}$ as the sample size increases.

3) Matrix-variate $\mathcal{G}^0$ Distribution: For this distribution with texture parameter λ, the second-order MLC is

$$\kappa_2\{\mathbf{C}\} = d^2\psi^{(1)}(\lambda) + \psi_d^{(1)}(L) \tag{93}$$

which leads to an estimator $\hat{\lambda}_{A_1}$ by solving

$$\psi^{(1)}(\hat{\lambda}_{A_1}) = \frac{\langle\kappa_2\{\mathbf{C}\}\rangle - \psi_d^{(1)}(L)}{d^2} \tag{94}$$

that is identical to $\hat{\alpha}_{A_1}$. The method of Doulgeris, derived in Appendix B, yields

$$\hat{\lambda}_D = \frac{2L\widehat{\mathrm{Var}}\{\tau\} + d(Ld-1)}{L\widehat{\mathrm{Var}}\{\tau\} - d} \tag{95}$$

while the fractional moment estimator is an average of the mono-pol estimates defined as the solution of

$$\frac{\Gamma^2\left(\hat{\lambda}_F - \frac{1}{4}\right)}{\Gamma(\hat{\lambda}_F)\Gamma\left(\hat{\lambda}_F - \frac{1}{2}\right)} \frac{\Gamma^2\left(L+\frac{1}{4}\right)}{\Gamma(L)\Gamma\left(L+\frac{1}{2}\right)} - \frac{\langle C^{\frac{1}{4}}\rangle^2}{\langle C^{\frac{1}{2}}\rangle} = 0. \tag{96}$$

In addition, an estimator $\hat{\lambda}_N$ is obtained by averaging the mono-pol estimates produced by $\hat{\lambda}_{A_1}$ for $d = 1$, while $\hat{\lambda}_{A_2}$ is defined in the same way as $\hat{\alpha}_{A_2}$, as given by (90).

We have performed Monte Carlo simulations for the estimators of λ with $\mathcal{G}^0$ distributed data parametrised by $L = 10$ and $\lambda = 10$. The bias and variance results in Figure 8 and 9 are very similar to those reported for the estimators of α. The main difference is that $\hat{\lambda_D}$, which does not use MKS, is superseded in terms of variance by the MKS estimators based on intensities only, $\hat{\lambda}_N$ and $\hat{\lambda}_F$, for $N > 200$. The preferred estimator is $\hat{\lambda}_{A_2}$, due to its superior bias and variance. Estimator $\hat{\lambda}_{A_1}$ has a comparably low variance for $N > 100$, which makes it a good alternative, due to a slightly lower complexity.

4) $\mathcal{U}$ Distribution: This distribution has two texture parameters, ξ and ζ. The estimation procedure therefore requires two MLC equations:

$$\kappa_2\{\mathbf{C}\} = d^2 \left(\psi^{(1)}(\xi) + \psi^{(1)}(\zeta)\right) + \psi_d^{(1)}(L) \quad (97)$$

$$\kappa_3\{\mathbf{C}\} = d^3 \left(\psi^{(2)}(\xi) - \psi^{(2)}(\zeta)\right) + \psi_d^{(2)}(L) \quad (98)$$

from which we can jointly determine the estimates $\hat{\xi}_{A_1}$ and $\hat{\zeta}_{A_1}$ by solving the equation system:

$$\psi^{(1)}(\hat{\xi}_{A_1}) + \psi^{(1)}(\hat{\zeta}_{A_1}) = \frac{\langle \kappa_2\{\mathbf{C}\}\rangle - \psi_d^{(1)}(L)}{d^2} \quad (99)$$

$$\psi^{(2)}(\hat{\xi}_{A_1}) - \psi^{(2)}(\hat{\zeta}_{A_1}) = \frac{\langle \kappa_3\{\mathbf{C}\}\rangle - \psi_d^{(2)}(L)}{d^3}. \quad (100)$$

The alternative estimators are $\hat{\xi}_N$ and $\hat{\zeta}_N$, that is., the averaged mono-pol estimates obtained from $\hat{\xi}_{A_1}$ and $\hat{\zeta}_{A_1}$ with $d = 1$. These have been implemented by the authors of [49].

The results for the $\mathcal{U}$ distribution estimators are similar to those reported for the $\mathcal{K}$ and $\mathcal{G}^0$ distributions, and are therefore omitted.

V. Conclusions

We have used a matrix-variate Mellin transform previously introduced by Mathai to extend the framework that we call *Mellin kind statistics* from the univariate to the matrix-variate case describing multilook polarimetric radar data. We have further defined the Mellin kind characteristic function and cumulant generating function for the matrix-variate case, and used them to define *matrix log-moments* and *matrix log-cumulants*. We have then proven the matrix-variate Mellin convolution theorem, and used it to develop expressions for Mellin kind statistics of the multilook polarimetric product model. Specific expressions for important distributions, such as the matrix-variate $\mathcal{K}$ distribution, $\mathcal{G}^0$ distribution and $\mathcal{U}$ distribution, have been given.

Mellin kind moments and cumulants are computed on a logarithmic scale, and the impact of speckle and texture therefore can be separated in the matrix-variate log-cumulant domain, which provides a valuable analysis tool for the doubly stochastic product model. Simulations have demonstrated the superior bias and variance properties possessed by estimators derived with the method of matrix log-cumulants. We have also used matrix log-cumulant space as a visualisation tool to provide intuition about estimation algorithms and model assessment that uses Mellin kind statistics. The mathematical tractability and the simplicity of the obtained expressions show, together with the excellent estimator properties documented, that the matrix-variate Mellin transform is a natural tool for analysis of multilook polarimetric radar data.

Appendix A
Mellin Kind Statistics for the Complex Wishart Distribution

Let $\mathbf{W} \sim \mathcal{W}_d^{\mathbb{C}}(L, \mathbf{\Sigma})$ have the complex Wishart distribution given in (7). The matrix-valued Mellin transform of $p_{\mathbf{W}}(\mathbf{W}; L, \mathbf{\Sigma})$, and hence the Mellin kind CF of the random matrix $\mathbf{W}$, is then

$$\begin{aligned}\phi_{\mathbf{W}}(s) &= \mathcal{M}_{\mathbf{W}}\{p_{\mathbf{W}}(\mathbf{W})\}(s) \\ &= \int_{\Omega_+} |\mathbf{W}|^{s-d} p_{\mathbf{W}}(\mathbf{W})\, d\mathbf{W} \\ &= \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} \frac{|\mathbf{\Sigma}|^{L+s-d}}{|\mathbf{\Sigma}|^L} \\ &\quad \times \int_{\Omega_+} p_{\mathbf{W}}(\mathbf{W}; L+s-d, \mathbf{\Sigma})\, d\mathbf{W} \\ &= \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)} |\mathbf{\Sigma}|^{s-d}.\end{aligned} \quad (101)$$

Accordingly, the Mellin kind CGF is

$$\begin{aligned}\varphi_{\mathbf{W}}(s) &= \ln \phi_{\mathbf{W}}(s) \\ &= \ln \Gamma_d(L+s-d) \\ &\quad - \ln \Gamma_d(L) + (s-d) \ln |\mathbf{\Sigma}|.\end{aligned} \quad (102)$$

We shall use the result

$$\begin{aligned}\frac{d}{dL}\Gamma_d(L) &= \frac{d}{dL}\left(\pi^{d(d-1)/2}\prod_{i=0}^{d-1}\Gamma(L-i)\right)\\ &= \pi^{d(d-1)/2}\sum_{i=0}^{d-1}\left(\frac{d}{dL}\Gamma(L-i)\prod_{\substack{j=0\\ j\neq i}}^{d-1}\Gamma(L-j)\right)\\ &= \pi^{d(d-1)/2}\prod_{j=0}^{d-1}\Gamma(L-j)\sum_{i=0}^{d-1}\psi^{(0)}(L-i)\\ &= \Gamma_d(L)\,\psi_d^{(0)}(L)\end{aligned} \tag{103}$$

which is obtained by straightforward application of the product rule of differentiation. We have also utilised the well-known relation

$$\frac{d}{dL}\Gamma(L) = \Gamma(L)\,\psi^{(0)}(L) \tag{104}$$

and the multivariate polygamma function introduced in (46). Remark that (103) is a multivariate version of (104). We also need the result

$$\frac{d}{dL}\psi_d^{(\nu)}(L) = \psi_d^{(\nu+1)}(L) \tag{105}$$

whose proof is trivial.

Equations (103) and (105) are used to deduce the derivatives of $\varphi_{\mathbf{W}}(s)$, denoted as $\varphi_{\mathbf{W}}^{(\nu)}(s) = \frac{d^\nu}{ds^\nu}\varphi_{\mathbf{W}}(s)$. The MLCs can then be written as $\kappa_\nu\{\mathbf{W}\} = \varphi_{\mathbf{W}}^{(\nu)}(d)$. By repeated differentiation of (102) and induction we find that

$$\kappa_1\{\mathbf{W}\} = \psi_d^{(0)}(L) + \ln|\boldsymbol{\Sigma}| \tag{106}$$

$$\kappa_\nu\{\mathbf{W}\} = \psi_d^{(\nu-1)}(L)\,, \quad \nu > 1\,. \tag{107}$$

Let $\mathbf{X}$ be a $d\times d$ complex positive definite matrix and $\mathbf{A}$ an equal size real constant matrix. The scaling property of the matrix-variate MT,

$$\mathcal{M}\{f(\mathbf{AX})\}(s) = |\mathbf{A}|^{-s}\mathcal{M}\{f(\mathbf{X})\}(s)\,, \tag{108}$$

is easily verified by evaluating the integral with a simple substitution of variables. For $\mathbf{A} = a\,\mathbf{I}_d$ with a real and positive scalar constant a, we get $\mathcal{M}\{f(a\mathbf{X})\}(s) = a^{-ds}\mathcal{M}\{f(\mathbf{X})\}(s)$. This is used to show that

$$\begin{aligned}&\mathcal{M}\{p_{\tilde{\mathbf{W}}}(\tilde{\mathbf{W}})\}(s)\\ &\quad= L^{d^2}\mathcal{M}\{p_{\mathbf{W}}(L\tilde{\mathbf{W}})\}(s)\\ &\quad= L^{-d(s-d)}\mathcal{M}\{p_{\mathbf{W}}(\mathbf{W})\}(s)\,.\end{aligned} \tag{109}$$

Recall the definition of $\tilde{\mathbf{W}} = \mathbf{W}/L$, which gives

$$\begin{aligned}\phi_{\tilde{\mathbf{W}}}(s) &= L^{-d(s-d)}\phi_{\mathbf{W}}(s)\\ &= \frac{\Gamma_d(L+s-d)}{\Gamma_d(L)}\left(\frac{|\boldsymbol{\Sigma}|}{L^d}\right)^{s-d}\,.\end{aligned} \tag{110}$$

This is used to show that

$$\varphi_{\tilde{\mathbf{W}}}(s) = \varphi_{\mathbf{W}}(s) - (s-d)d\ln L \tag{111}$$

and the MLCs of $\tilde{\mathbf{W}}$ follow immediately as given in (48) and (49).

APPENDIX B
DOULGERIS' PARAMETER ESTIMATORS

Doulgeris et al. [46] derived their estimator for the texture parameter of the matrix-variate $\mathcal{K}$ distribution from moments of the Hotelling-Lawley trace. This is an important test statistic in multivariate statistics, defined as

$$\tau = \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C})\,. \tag{112}$$

It is easily shown that $\mathrm{E}\{\tau\} = d$, and the variance of τ is

$$\mathrm{Var}\{\tau\} = \mathrm{E}\{T^2\}\left(d^2 + \frac{d}{L}\right) - d^2\,. \tag{113}$$

Given a choice of the texture RV T, we may solve for the texture parameter to obtain an estimator. For instance, with $T \sim \bar{\gamma}(\alpha)$ we have $\mathrm{E}\{T^2\} = (\alpha+1)/\alpha$, which yields the following estimator for the $\mathcal{K}$ distribution parameter α:

$$\hat{\alpha}_D = \frac{d(Ld+1)}{L\widehat{\mathrm{Var}}\{\tau\} - d}\,. \tag{114}$$

This estimator was given in [46]. The variance of τ is estimated with a standard variance estimator from a population of Hotelling-Lawley traces, $\{\tau_i = \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C}_i)\}_{i=1}^N$.

When $T \sim \bar{\gamma}^{-1}(\lambda)$, we have $\mathrm{E}\{T^2\} = (\lambda-1)/(\lambda-2)$, which is used to derive

$$\hat{\lambda}_D = \frac{2L\widehat{\mathrm{Var}}\{\tau\} + d(Ld-1)}{L\widehat{\mathrm{Var}}\{\tau\} - d} \tag{115}$$

for the $\mathcal{G}^0$ distribution parameter λ. The method is not pursued for the $\mathcal{U}$ distribution, since it would require derivation of higher moments of τ to solve for both texture parameters.

ACKNOWLEDGEMENT

The authors wish to thank Anthony Paul Doulgeris at the University of Tromsø and Gabriele Moser at the University of Genoa for their valuable contribution through discussions of the topic and comments on the manuscript.

REFERENCES

[1] B. Epstein, "Some applications of the Mellin transform in statistics," *Ann. Math. Statist.*, vol. 19, no. 3, pp. 370–379, Sep. 1948.

[2] 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.

[3] ——, "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.

[4] R. D. Pierce, "Application of the positive alpha-stable distribution," in *Proc. IEEE Signal Proc. Workshop on Higher-Order Stat., SPW-HOS '97*, Banff, Canada, Jul. 1997, pp. 420–424.

[5] E. W. Stacy, "A generalization of the gamma distribution," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1187–1192, Sep. 1962.

[6] E. W. Stacy and G. A. Mihram, "Parameter estimation for a generalized gamma distribution," *Technometrics*, vol. 7, no. 3, pp. 349–358, Aug. 1965.

[7] G. Moser, J. Zerubia, and S. B. Serpico, "Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 188–200, Jan. 2006.

[8] ——, "SAR amplitude probability density function estimation based on a generalized Gaussian model," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1429–1442, Jun. 2006.

[9] 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.

[10] J.-M. Nicolas, "A Fisher-MAP filter for SAR image processing," in *Proc. IEEE Int. Geosc. Remote Sens. Symp., IGARSS'03*, vol. 3, Toulouse, France, Jul. 2003, pp. 1996–1998.

[11] A. Achim, E. E. Kuruoğlu, and J. Zerubia, "SAR image filtering based on the heavy-tailed Rayleigh model," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2686–2693, Sep. 2006.

[12] G. Chen and X. Liu, "Wavelet-based SAR image despeckling using Cauchy pdf modeling," in *IEEE Radar Conf. 2008*, Rome, Italy, 26-30 May 2008, pp. 1–5.

[13] 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.

[14] D. Benboudjema, F. Tupin, W. Pieczynski, M. Sigelle, and J.-M. Nicolas, "Unsupervised segmentation of SAR images using triplet Markov fields and Fisher noise distributions," in *Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS'07*, vol. 1, Barcelona, Spain, Jul. 2007, pp. 3891–3894.

[15] F. Galland, J.-M. Nicolas, H. Sportouche, M. Roche, F. Tupin, and P. Réfrégier, "Unsupervised synthetic aperture radar image segmentation using Fisher distributions," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2966–2972, Aug. 2009.

[16] F. Bujor, E. Trouvé, L. Valet, J.-M. Nicolas, and J.-P. Rudant, "Application of log-cumulants to the detection of spatiotemporal discontinuities in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2073–2084, Oct. 2004.

[17] G. Moser and S. B. Serpico, "Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2972–2982, Oct. 2006.

[18] ——, "Unsupervised change detection from multichannel SAR data by Markovian data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2114–2128, Jul. 2009.

[19] R. Abdelfattah and J.-M. Nicolas, "Interferometric SAR coherence magnitude estimation using second kind statistics," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, part 2, pp. 1942–1953, Jul. 2006.

[20] C. Valade and J.-M. Nicolas, "Homomorphic wavelet transform and new subband statistics models for SAR image compression," in *Proc. IEEE Int. Geosc. Remote Sens. Symp., IGARSS'04*, vol. 1, Anchorage, USA, Sep. 2004, pp. 285–288.

[21] A. M. Mathai, *Jacobians of Matrix Transformations and Functions of Matrix Arguments*. New York, USA: World Scientific, 1997, ch. 6, Def. 6.3.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] R. Touzi, W. M. Boerner, J.-S. Lee, and E. Lüneburg, "A review of polarimetry in the context of synthetic aperture radar: Concepts and information extraction," *Can. J. Remote Sens.*, vol. 30, no. 3, pp. 380–407, 2004.

[28] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, USA: McGraw-Hill, 2002.

[29] W. P. Johnson, "The curious history of Faà di Bruno's formula," *Am. Math. Monthly*, vol. 109, no. 3, pp. 217–234, Mar. 2002.

[30] L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions*. Dordrecht, The Netherlands: Reidel Publishing Company, 1974, ch. 3.3, pp. 133–136.

[31] H. H. Andersen, M. Højbjerre, D. Sørensen, and P. S. Eriksen, *Linear and Graphical Models for the Multivariate Complex Normal Distribution*, ser. Lecture Notes in Statistics. New York, USA: Springer, 1995.

[32] C.-Y. Chi, C.-C. Feng, C.-H. Chen, and C.-Y. Chen, *Blind Equalization and System Identification*. London: Springer, 2006, ch. 3, pp. 103–104.

[33] T. Kollo and D. von Rosen, *Advanced Multivariate Statistics with Matrices*. Dordrecht, The Netherlands: Springer, 2005.

[34] J. Bertrand, P. Bertrand, and J.-P. Ovarlez, "The Mellin transform," in *The Transform and Applications Handbook*, 2nd ed., A. D. Poularikas, Ed. Boca Raton, US: CRC Press, 2000, ch. 11.

[35] A. M. Mathai, "Some results on functions of matrix argument," *Math. Nachr.*, vol. 84, no. 1, pp. 171–177, 1978.
[36] ——, "Distribution of the canonical correlation matrix," *Ann. Inst. Statist. Math.*, vol. 33, part A, pp. 35–43, 1981.
[37] B. Mathur and S. Krishna, "On multivariate fractional integration operators," *Indian J. Pure Appl. Math.*, vol. 8, pp. 1078–1082, 1977.
[38] 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.
[39] S. N. Anfinsen, T. Eltoft, and A. P. Doulgeris, "A relaxed Wishart model for polarimetric SAR data," in *Proc. 4th Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '09)*, vol. ESA SP-668, Frascati, Italy, 8 pp., Jan. 2009.
[40] 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.
[41] D. Casasent and D. Psaltis, "Deformation invariant, space-invariant optical pattern recognition," in *Progress in Optics*, E. Wolf, Ed. New York, USA: North-Holland, 1978, vol. XVI, ch. V.
[42] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, USA: Chapman & Hall, 1993.
[43] S. N. Anfinsen and T. Eltoft, "Goodness-of-fit tests for multilook polarimetric radar data based on the Mellin transform," *IEEE Trans. Geosci. Remote Sens.*, 2010, submitted, Available: http://www.phys.uit.no/~stiann/sna-gof-submitted.pdf.
[44] S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Estimation of the equivalent number of looks in polarimetric SAR imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS'08*, vol. 4, Boston, USA, Jul. 2008, pp. 487–490.
[45] A. C. Frery, A. H. Correia, and C. C. Freitas, "Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3098–3109, Oct. 2007.
[46] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.
[47] D. E. Kreithen and G. G. Hogan, "Statistical analysis of Ka-band sea clutter," in *Proc. IEEE OCEANS '91*, vol. 2. Honolulu, USA: IEEE, Oct. 1991, pp. 1217–1222.
[48] D. Blacknell, "Comparison of parameter estimator for K-distribution," *IEE Proc. Radar, Sonar, Navig.*, vol. 141, no. 1, pp. 45–52, Feb. 1994.
[49] O. Harant, L. Bombrun, M. Gay, R. Fallourd, E. Trouvé, and F. Tupin, "Segmentation and classification of polarimetric SAR data based on the KummerU distribution," in *Proc. 4th Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '09)*, vol. ESA SP-668, Frascati, Italy, 6 pp., Apr. 2009.

Chapter 7

Paper 3:
Goodness-of-Fit Tests for Multilook Polarimetric Radar Data Based on the Mellin Transform

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.

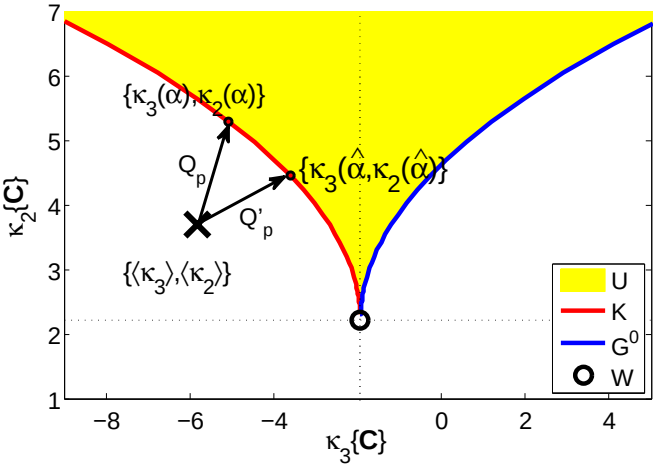

Fig. 1. MLC space geometrical interpretation of the goodness-of-fit test proposed for simple and composite hypotheses. The test statistics Q_p and Q'_p are interpreted as distance between data (sample MLC vector) and model (point on the manifold of population MLC vectors).

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

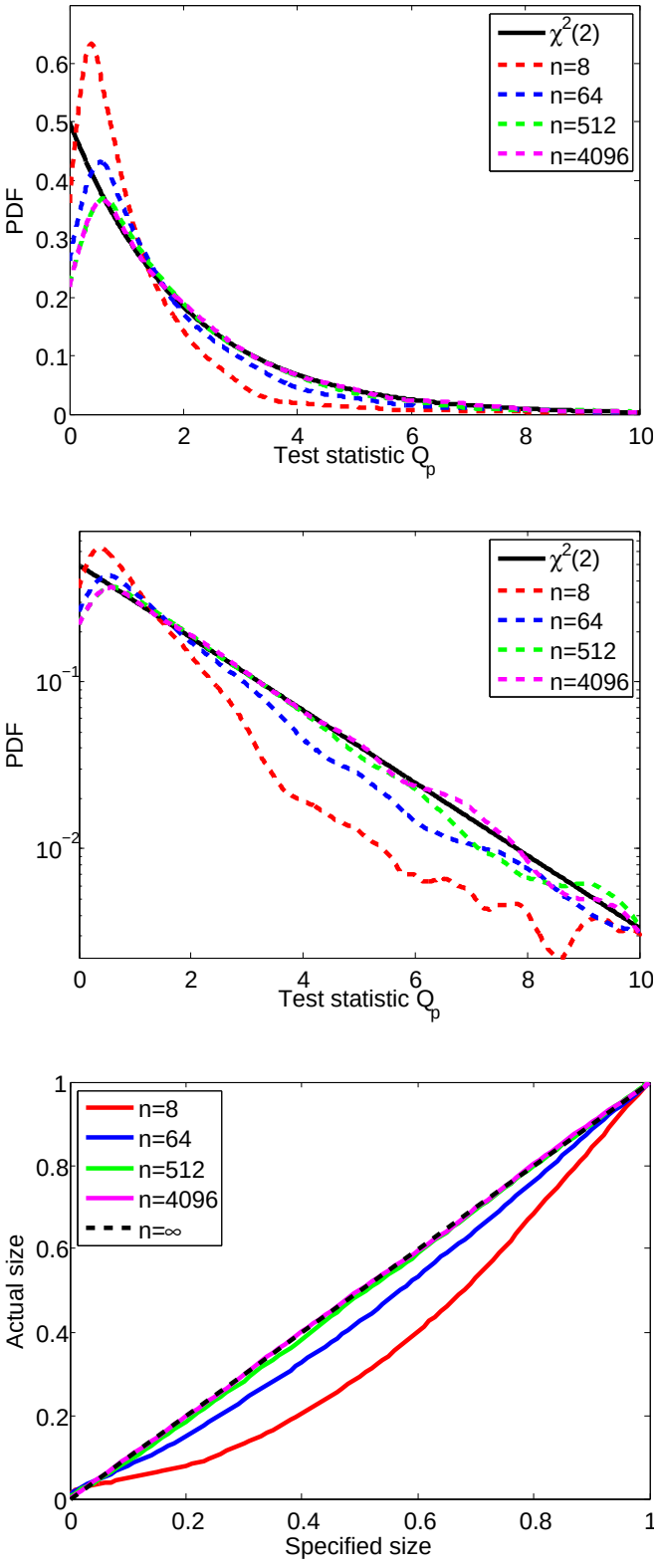

Fig. 2. Assessing validity of $\chi^2(2)$ distribution model for Q_2 with finite sample sizes. The simulated data follow the distribution specified under $H_0 : \mathbf{C} \sim s\mathcal{W}_d^{\mathbb{C}}(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0)$. The plots show estimated sampling distributions 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

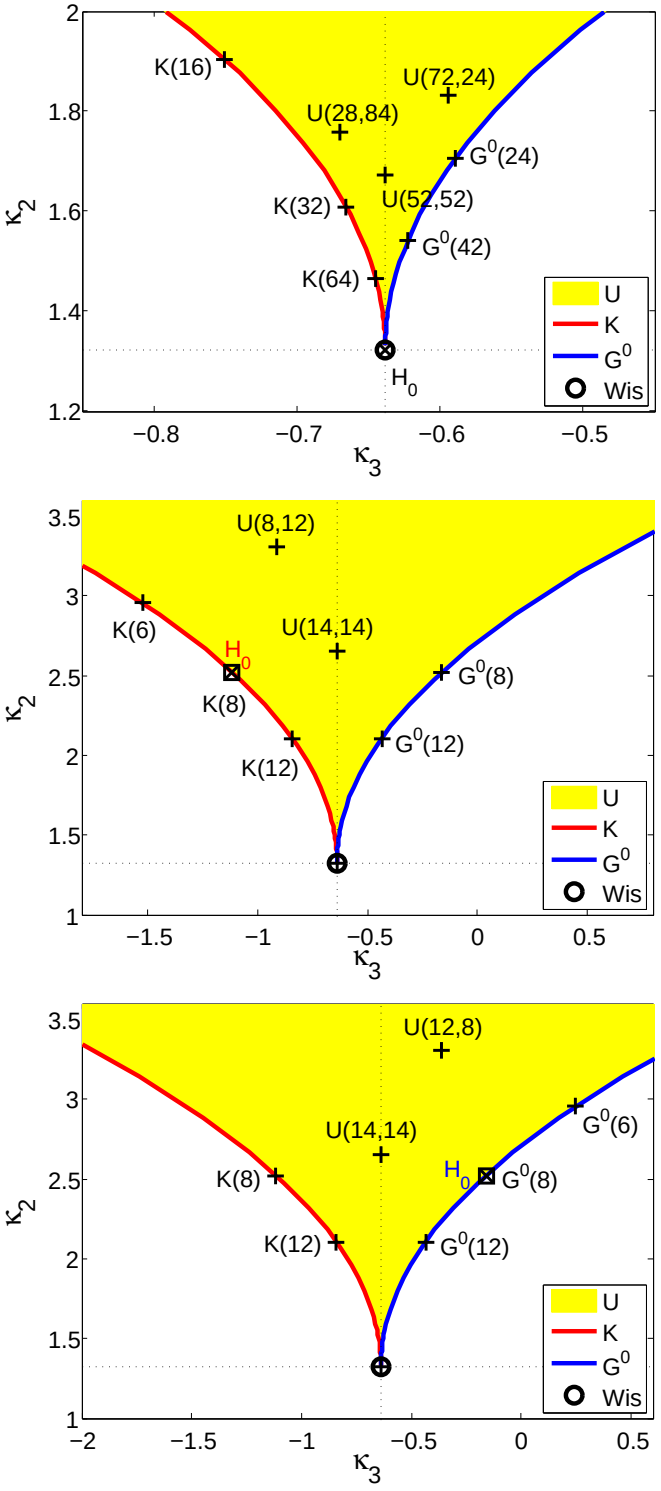

Fig. 3. Data applied to the $\chi^2(2)$ tests of $H_0 : \mathbf{C} \sim \mathcal{W}_d^{\mathbf{C}}(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0)$ (top), $H_0 : \mathbf{C} \sim \mathcal{K}(L=4, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0, \alpha=8)$ (middle), and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L=4,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

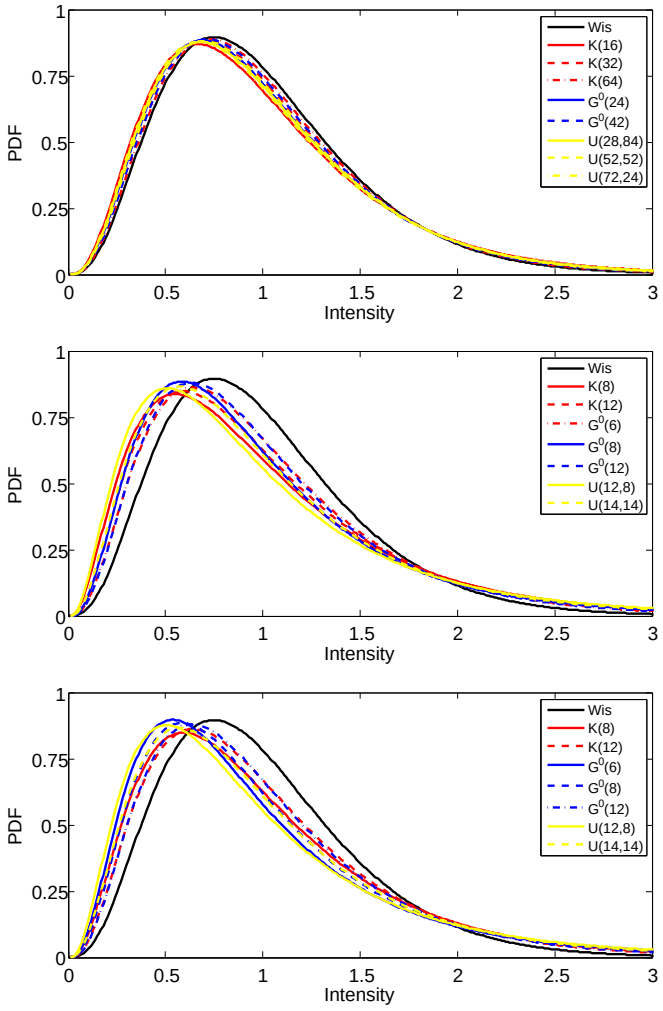

Fig. 4. Data applied to the $\chi^2(2)$ tests of the simple hypotheses H_0 : $\mathbf{C} \sim \mathcal{W}_d^{\mathbf{C}}(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0)$ (top), $H_0 : \mathbf{C} \sim \mathcal{K}(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0, \alpha=8)$ (middle), and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L=4,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

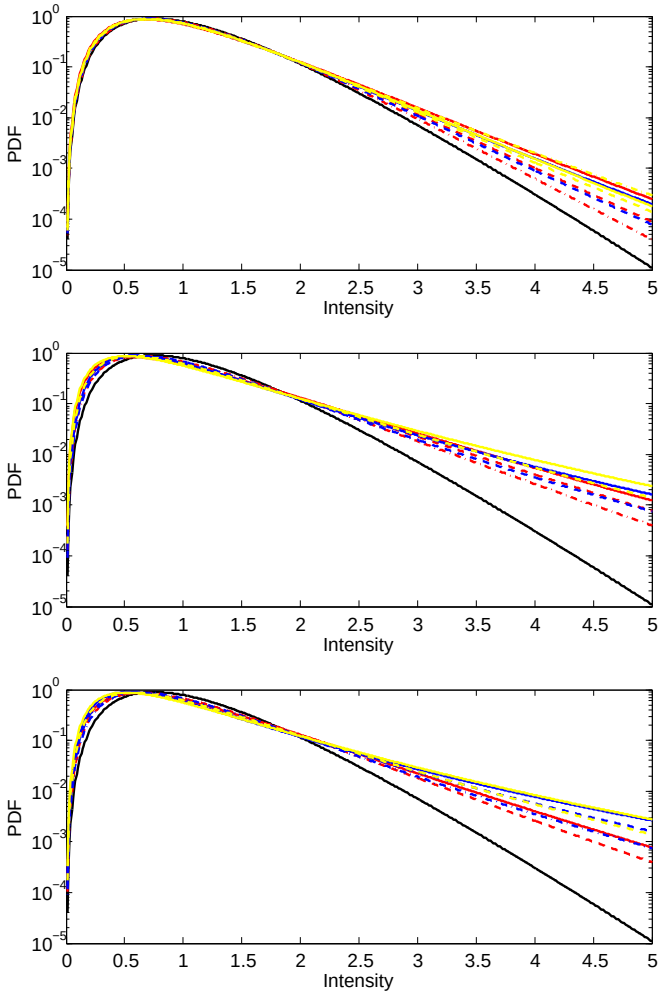

Fig. 5. Data applied to the $\chi^2(2)$ tests of the simple hypotheses H_0 : $\mathbf{C} \sim \mathcal{W}_d^{\mathbf{C}}(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0)$ (top), $H_0 : \mathbf{C} \sim \mathcal{K}(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0, \alpha=8)$ (middle), and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L=4, \mathbf{\Sigma}=\mathbf{\Sigma}_0, \lambda=8)$ (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:

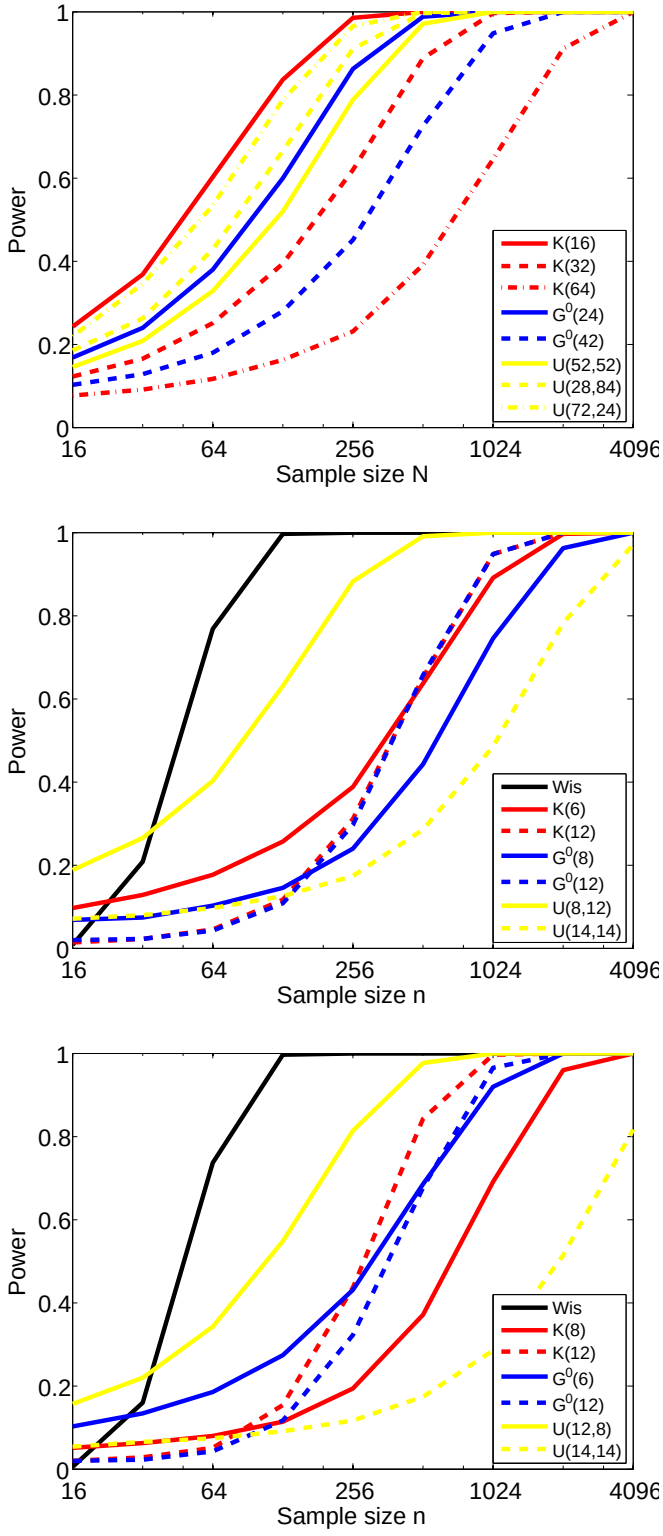

Fig. 6. Power of the Q_p test of the simple hypotheses $H_0 : \mathbf{C} \sim \mathcal{W}_d^{\mathbb{C}}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0)$ (top), $H_0 : \mathbf{C} \sim \mathcal{K}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0, \alpha = 8)$ (middle), and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L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

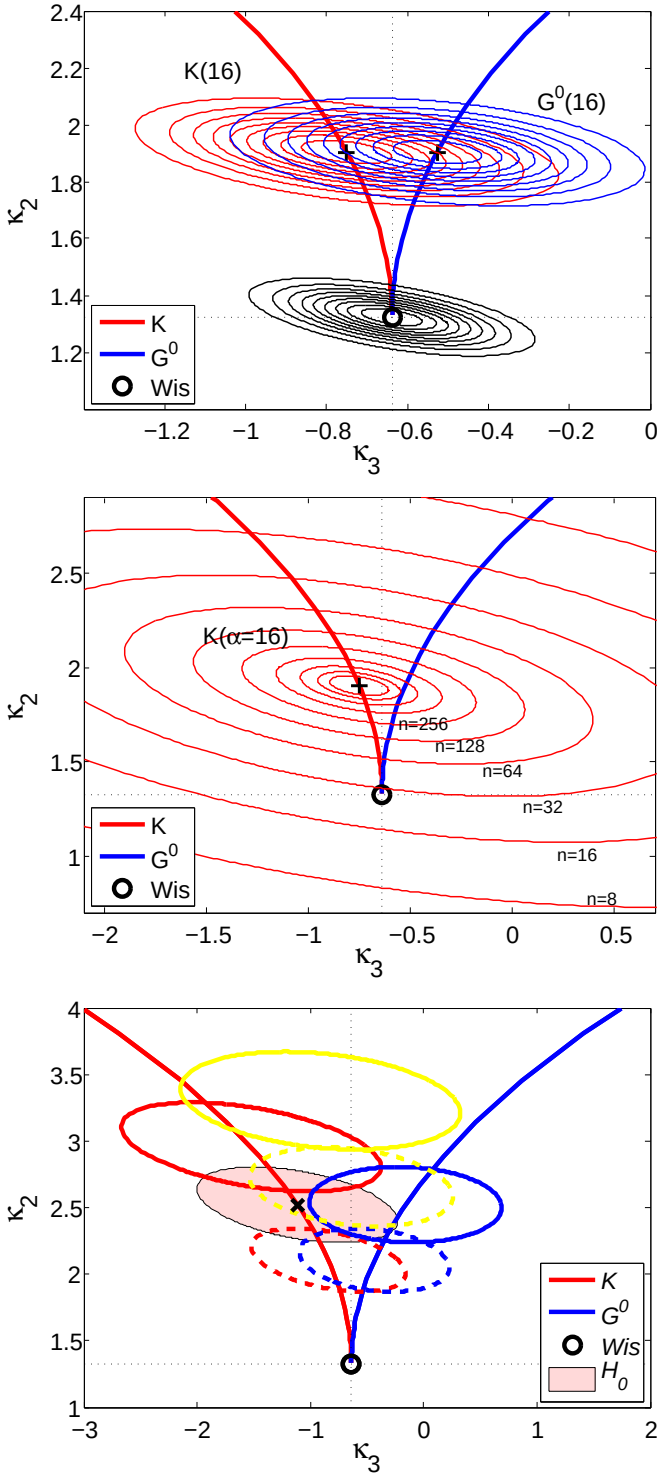

Fig. 7. Top panel: Equiprobability curves defining the $\{10, 20, 30, \cdots, 90\}$th percentiles of the models: $\mathbf{C} \sim \mathcal{W}_d^{\mathbb{C}}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0)$ (black ellipses), $\mathbf{C} \sim \mathcal{W}_d^{\mathbb{C}}(L = 4, \mathbf{\Sigma} = \mathbf{\Sigma}_0, \alpha = 16)$ (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 $n = 1024$.

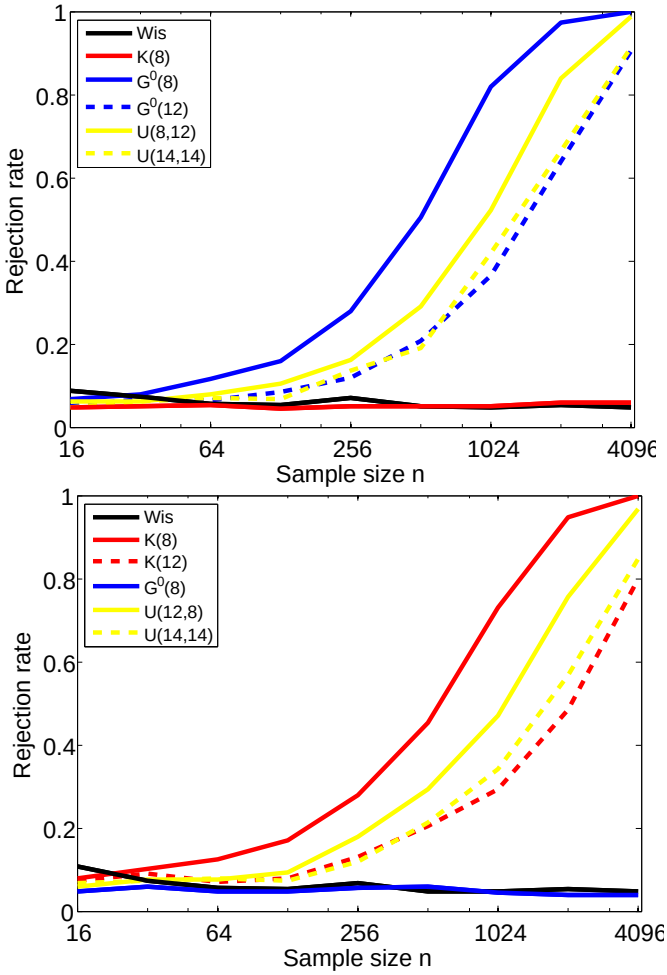

Fig. 8. Rejection rate of the Q'_p test of the composite hypotheses: $H_0 : \mathbf{C} \sim \mathcal{K}(L, \mathbf{\Sigma}, \alpha)$ (top) and $H_0 : \mathbf{C} \sim \mathcal{G}^0(L, \mathbf{\Sigma}, \lambda)$ (bottom) at the $\alpha_c = 5\%$ significance level with $\nu = \{2, 3\}$ for various data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

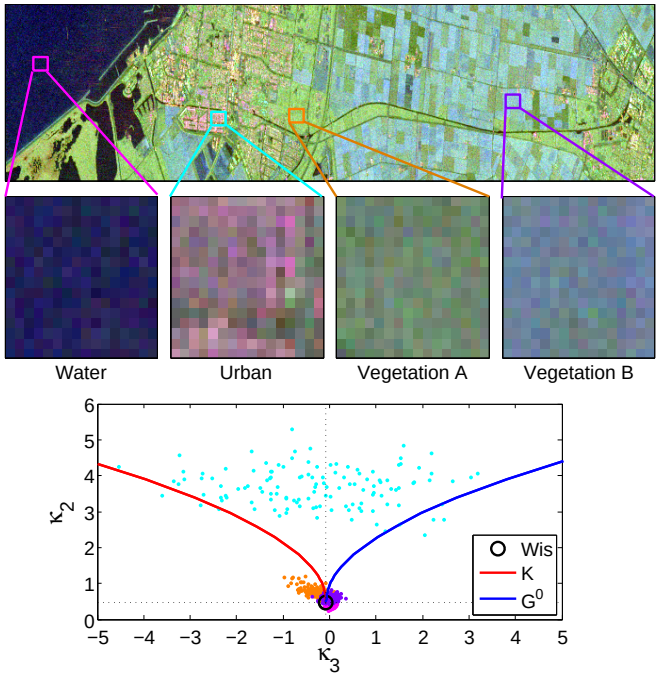

Fig. 9. **Top:** Subset of Radarsat-2 fine quad polarisation mode image of Flevoland, The Netherlands, acquired on 2 March 2008. **Middle:** Homogeneous samples of a water body, urban area, and two vegetated areas (labelled A and B). **Bottom:** MLC diagram with sample MLCs computed from the image samples.

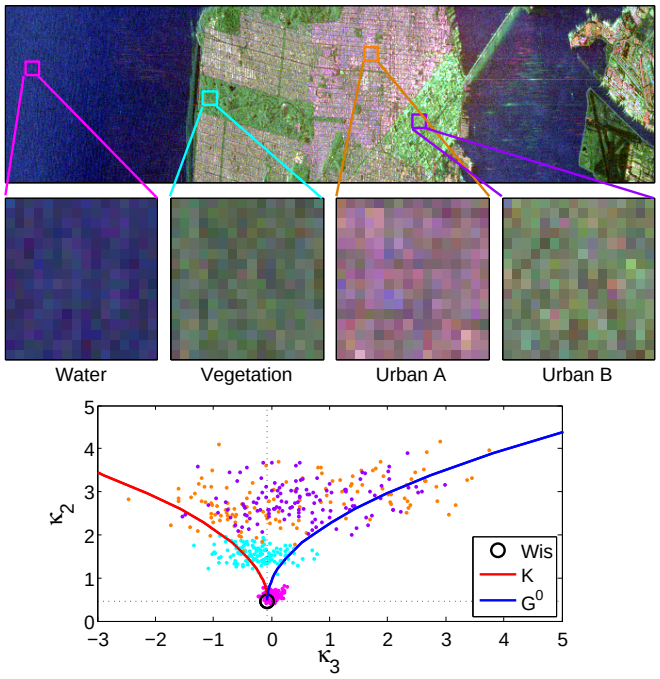

Fig. 10. **Top:** Subset of Radarsat-2 fine quad polarisation mode image of San Francisco, United States, acquired on 9 March 2008. **Middle:** Homogeneous samples of a water body, a vegetated area, and two urban areas (labelled A and B). **Bottom:** MLC diagram with sample MLCs computed from the image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

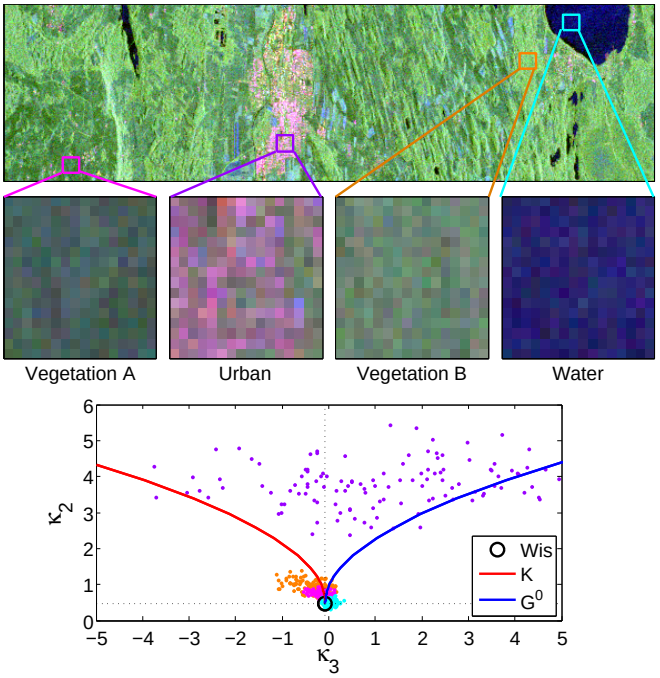

Fig. 11. **Top:** Subset of Radarsat-2 fine quad polarisation mode image of Oberpfaffenhofen, Germany, acquired on 6 March 2008. **Middle:** Homogeneous samples of a vegetated area (labelled A), an urban area, a vegetated area (labelled B), and a water body. **Bottom:** MLC diagram with sample MLCs computed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

Chapter 8

Conclusions

This chapter gives concluding remarks and outlines some directions of future research.

8.1 Concluding Remarks

In this thesis, it has been shown that the Mellin kind statistics framework is a natural tool for analysis of the matrix distributions derived under the multilook polarimetric product model. The simple and elegant mathematical expression we obtain, the performance of the parameter estimators, and the interpretability of the results are taken as proof of this statement.

Some excellent parameter estimators have been derived by the method of matrix log-cumulants, and a pioneering goodness-of-fit test has been constructed. However, by looking at the long list of applications that have grown out of Nicolas' univariate Mellin kind statistics (See the second paragraph of Section 1.1), it can be expected that more algorithms for analysis of multilook polarimetric images will follow. We have already identified some target areas and applications.

8.2 Future Research

The first paper where the contours of the matrix-variate Mellin kind statistics framework could be seen was [Anfinsen et al., 2009]. This paper has been included in Appendix A. It discusses statistical modelling of speckle filtered multilook polarimetric images and demonstrates that the filtering alters the data such that new models are required. This is a topic which is worth pursuing. By looking at the characteristics of the speckle filtered data in matrix log-cumulant space, the matrix-variate $\mathcal{W}$ and $\mathcal{M}$ distributions are launched as potential models for filtered speckle in heterogeneous and extremely heterogeneous areas. Another task would be to find statistical models for partially developed speckle, and the Mellin kind statistics could possible be helpful in the characterisation of such a model.

An interesting observation is that the log-determinant transformation of the polarimetric covariance matrix compresses the matrix data to a scalar band. By storing this band only, we can still compute matrix log-moments and matrix log-cumulants and make inferences based on them. This compression could be utilised to make fast algorithms for different image analysis tasks that still maintain high performance, since the matrix log-cumulants have a high content of statistical information.

It was mentioned in Section 2.3.5 that the texture parameters of the radar image distributions can be used as textural features in various image analysis problems. The matrix log-cumulants would contain the same information, but are cheaper to compute and not confined to a specific distribution model. This is an area that we want to explore further. Compound matrix distributions have already been applied to classification [Doulgeris et al., 2008] and segmentation [Bombrun and Beaulieu, 2008, Harant et al., 2009, Vasile et al., 2009]. In polarimetric change detection on the other hand, the test statistics proposed in [Conradsen et al., 2003, Kersten et al., 2005] are developed under the assumption of Wishart statistics. Extensions could possibly be achieved by means of Mellin kind statistics.

As a final remark, we note again that the log-determinant compression induced by the matrix-variate Mellin transform reduces the data dimension to one. The statistical information contained in the polarimetric covariance matrix could possibly be better preserved by using the multivariate Mellin transform from [Mathur and Krishna, 1977], even though the off-diagonal complex correlations will be discarded in such an approach. Thus, there is a chance we could use the Mellin transform to squeeze even more information out of polarimetric radar images.

Appendix A

A Relaxed Wishart Model for Polarimetric SAR Data

A RELAXED WISHART MODEL FOR POLARIMETRIC SAR DATA

Stian Normann Anfinsen, Torbjørn Eltoft, and Anthony Paul Doulgeris

University of Tromsø, Department of Physics and Technology, NO-9037 Tromsø, Norway,
E-mail: {stian.normann.anfinsen, torbjorn.eltoft, anthony.p.doulgeris}@uit.no

ABSTRACT

In this paper we demonstrate that simple yet flexible modelling of multilook polarimetric synthetic aperture radar (PolSAR) data can be obtained through a relaxation of the Wishart model. The degrees of freedom of the complex Wishart distribution is treated as a spatially nonstationary parameter, which is allowed to vary between thematic classes and segments of the PolSAR scene.

Key words: synthetic aperture radar; polarimetry; statistical modelling; Wishart distribution.

1. INTRODUCTION

The Wishart distribution is the de facto statistical model for multilook PolSAR data. It is based on the assumption that the complex scattering coefficients are jointly circular Gaussian. However, this is only satisfied for homogeneous areas with fully developed speckle and no texture, which renders the model inadequate in many cases. Improved modelling is achieved by using more complex models that account for texture, such as the polarimetric $\mathcal{G}$ distribution family [1], with the polarimetric $\mathcal{K}$ distribution [2] as a special case. These models allow for better adaption to data whose distribution is heavy-tailed and non-Gaussian, but this comes at the cost of higher mathematical complexity.

The comparatively higher mathematical tractability of the Wishart distribution motivates us to pursue a relaxed Wishart model as an alternative. In the context of multilook PolSAR data, the degrees of freedom of the Wishart distribution is interpreted as the equivalent number of looks, a constant, global value that quantifies the effective number of data samples averaged in the multilooking process. In contrast, we treat it as a free parameter, which varies between, and possibly also within, classes and segments of the PolSAR scene. This reflects the highly variable degree of smoothing imposed on the data by nonlinear speckle filters. The choice can also be justified by looking at the degrees of freedom as a shape parameter of the distribution, which is determined not only by the degree of averaging, but also by texture. Thus, the influence of multilooking, speckle filtering, and texture is assimilated into one parameter, which can be estimated efficiently with a recently proposed estimator [3, 4]

Sec. 2 reviews some existing density models for multilook PolSAR data and proposes the relaxed Wishart distribution as an alternative. In Sec. 3 we derive certain matrix moments that are used to illustrate the adaptivity of the different density models, and as a new domain for visual goodness-of-fit assessment. Sec. 4 presents experiments with airborne NASA/JPL AIRSAR data, and in Sec. 5 we give our conclusions.

2. STATISTICAL MODELLING

The full-polarimetric SAR instrument separately transmits orthogonally polarised microwaves pulses, and measures orthogonal components of the received signal. For each pixel, the measurements result in a matrix of scattering coefficients. These are complex-valued, dimensionless numbers that describe the transformation of the transmitted (incoming) electromagnetic (EM) field to the received (backscattered) EM field for all combinations of transmit and receive polarisation.

The transformation can be expressed as

$$\begin{bmatrix} E_h^r \\ E_v^r \end{bmatrix} = \frac{e^{jkr}}{r} \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} \begin{bmatrix} E_h^t \\ E_v^t \end{bmatrix}, \quad (1)$$

where k denotes wavenumber and r is the distance between radar and target. The subscript of the EM field components E_i^j denotes horizontal (h) or vertical (v) polarisation, which is the most common set of orthogonal polarisations, while the superscript indicates transmitted (t) or received (r) wave. The scattering coefficients S_{ij} are subscripted with the associated receive and transmit polarisation, in that order. Together, they form the scattering matrix, denoted $\mathbf{S} = [S_{ij}]$.

The scattering matrix can be reduced to one of the vectors

$$\mathbf{s} = \begin{bmatrix} S_{hh} \\ (S_{hv}+S_{vh})/\sqrt{2} \\ S_{vv} \end{bmatrix} \text{ or } \mathbf{k} = \frac{1}{\sqrt{2}} \begin{bmatrix} S_{hh}+S_{vv} \\ S_{hh}-S_{vv} \\ S_{hv}+S_{vh} \end{bmatrix}. \quad (2)$$

The lexicographic scattering vector, denoted $\mathbf{s}$, is the vectorised version of $\mathbf{S}$ after the cross-polarisation terms S_{hv}

and S_{vh} have been averaged, assuming reciprocity of the target. The scaling with a factor $\sqrt{2}$ is done to preserve total power of the signal. The Pauli basis scattering vector, denoted $\mathbf{k}$, is a linear transformation of $\mathbf{s}$, which provides physical interpretations of its elements in terms of basic scattering mechanisms [5].

2.1. Gaussian Model

It is commonly assumed that the scattering vector elements are jointly circular complex Gaussian, even though this model only encompasses variability due to speckle, and not texture, which is discussed in the Sec. 2.2. The matrix $\mathbf{S}$ and the vectors $\mathbf{s}$ and $\mathbf{k}$ are single-look complex format representations of PolSAR data. The following derivations shall use $\mathbf{s}$ as the scattering vector, but would be equivalent for $\mathbf{k}$.

Multilook PolSAR data is commonly represented by

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

known as the sample covariance matrix. It is formed as the mean Hermitian outer product of the single-look scattering vectors $\{\mathbf{s}_i\}_{i=1}^{L}$, where L is the nominal number of looks. The superscript H means complex conjugate transpose. Assume that $\mathbf{s}$ is zero mean and circular complex multivariate Gaussian, denoted $\mathbf{s} \sim \mathcal{N}_d^{\mathbb{C}}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{0}$ is a column vector of zeros, d is the dimension of $\mathbf{s}$, and $\mathbf{\Sigma} = \mathrm{E}\{\mathbf{s}\mathbf{s}^H\}$ is the covariance matrix of $\mathbf{s}$. The probability density function (pdf) of $\mathbf{s}$ is thus

$$p_{\mathbf{s}}(\mathbf{s}; \mathbf{\Sigma}) = \frac{1}{\pi^d |\mathbf{\Sigma}|} \exp(-\mathbf{s}^H \mathbf{\Sigma}^{-1} \mathbf{s}) \,, \tag{4}$$

where $|\cdot|$ is the determinant operator. It follows that if $L \geqslant d$ and the $\mathbf{s}_i$ in (3) are independent, then $\mathbf{C}$ follows the nonsingular complex Wishart distribution [6]:

$$\begin{aligned} &p_{\mathbf{W}}(\mathbf{C}; L, \mathbf{\Sigma}) \\ &\quad = \frac{L^{Ld} |\mathbf{C}|^{L-d}}{|\mathbf{\Sigma}|^L \Gamma_d(L)} \exp\left(-L \operatorname{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{C}\right)\right) \,, \end{aligned} \tag{5}$$

where $\operatorname{tr}(\cdot)$ is the trace operator. The normalisation constant $\Gamma_d(L)$ is the multivariate Gamma function, defined as

$$\Gamma_d(L) = \pi^{d(d-1)/2} \prod_{i=0}^{d-1} \Gamma(L-i) \,, \tag{6}$$

where $\Gamma(L)$ is the standard Euler gamma function. In reality, the $\mathbf{s}_i$ are correlated, and this is compensated for by replacing L with an equivalent number of looks, $L_e \leqslant L$, in order to obtain consistency between moments of the theoretical model and sample moments of the data. This approximation provides a good model for the distribution of $\mathbf{C}$, denoted $\mathbf{C} \sim \mathcal{W}_d^{\mathbb{C}}(L_e, \mathbf{\Sigma})$.

2.2. Product Model

In addition to speckle, the randomness of a SAR measurement can also be attributed to texture. The notion of texture represents the natural spatial variation of the radar cross section, which is generally not perfectly homogeneous for pixels that are thematically mapped as one class. Several statistical models exist that incorporate texture, either by assuming statistics that imply a non-Gaussian scattering vector, or explicitly modelling texture as a separate random variable (rv). The latter case leads to a doubly stochastic model with a compounded distribution.

The well known product model, reviewed in [7, 8, 9], is shown to be both mathematically tractable and successful for modelling purposes. In the multilook polarimetric version [1], the polarimetric covariance matrix $\mathbf{C}$ is decomposed as a product of two independent stochastic processes with individual distributions:

$$\mathbf{C} = z\mathbf{W} \,. \tag{7}$$

One process, $\mathbf{W} \sim \mathcal{W}_d^{\mathbb{C}}(L_e, \mathbf{\Sigma})$, models speckle. The other process generates texture, represented by the scalar rv $z \in \mathbb{R}^+$, under the assumption that texture is independent of polarisation. The pdf of $\mathbf{C}$ depends on the distribution of z, which is normalised to unit mean.

Gamma Distributed Texture

The first covariance matrix distribution derived from the product model used the gamma distribution to model z [2]. A gamma distributed rv $z > 0$ has density

$$p_z(z; \alpha, \mu) = \left(\frac{\alpha}{\mu}\right)^{\alpha} \frac{z^{\alpha-1}}{\Gamma(\alpha)} \exp\left(-\frac{\alpha}{\mu} z\right) \,, \tag{8}$$

with shape parameter $\alpha > 0$ and mean value $\mu = \mathrm{E}\{z\} > 0$. This is denoted $z \sim \mathcal{G}(\alpha, \mu)$. The unitary mean texture rv is thus $z \sim \mathcal{G}(\alpha, 1)$. This leads to the matrix-variate $\mathcal{K}$ distribution for $\mathbf{C}$ [1, 2]:

$$\begin{aligned} &p_{\mathbf{C}}(\mathbf{C}; L_e, \mathbf{\Sigma}, \alpha) \\ &\quad = \frac{2|\mathbf{C}|^{L_e-d} (L_e\alpha)^{\frac{\alpha+L_e d}{2}}}{|\mathbf{\Sigma}|^{L_e} \Gamma_d(L_e) \Gamma(\alpha)} \left(\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})\right)^{\frac{\alpha-L_e d}{2}} \\ &\quad \times K_{\alpha-L_e d}\left(2\sqrt{L_e \alpha \operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C})}\right) . \end{aligned} \tag{9}$$

$K_n(\cdot)$ is the modified Bessel function of the second kind with order n. See [1, 10] for a detailed derivation.

Inverse Gamma Distributed Texture

The family of generalised inverse Gaussian distributions was proposed in [1] as a model for z. The gamma distribution is one special case. The inverse gamma distribution is another, which has been promoted in particular as

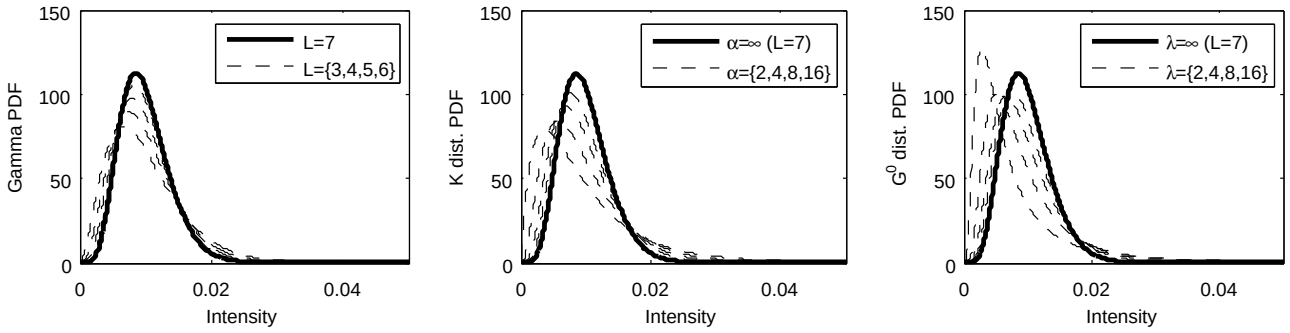

Figure 1. Examples of single intensity marginal densities for a polarimetric covariance matrix modelled by the relaxed Wishart distribution (left), matrix-variate $\mathcal{K}$ distribution (middle), and matrix-variate $\mathcal{G}^0$ distribution (right).

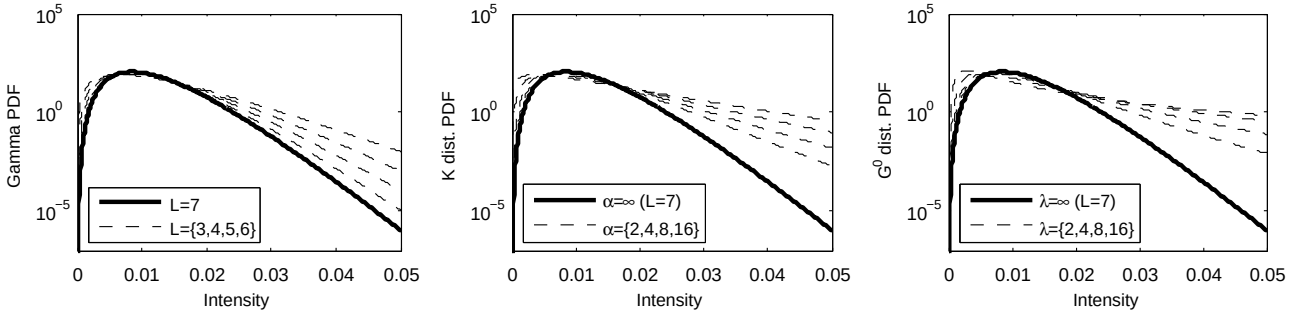

Figure 2. Same as Fig. 1 with logarithmic second axis to emphasize differences at the tails.

a good model for strongly heterogeneous clutter [1, 11]. Its pdf is given by

$$p_z(z;\lambda,\nu) = (\lambda\nu)^\lambda \frac{z^{-\lambda-1}}{\Gamma(\lambda)} \exp\left(-\frac{\lambda\nu}{z}\right) . \qquad (10)$$

This is denoted $z \sim \mathcal{G}^{-1}(\lambda,\nu)$, with shape parameter $\lambda > 0$ and $\nu > 0$. The normalised texture rv becomes $z \sim \mathcal{G}^{-1}(\lambda,(\lambda-1)/\lambda)$, which leads to the matrix-variate $\mathcal{G}^0$ distribution for $\mathbf{C}$ [1]:

$$\begin{aligned} &p_{\mathbf{C}}(\mathbf{C};L_e,\mathbf{\Sigma},\lambda) \\ &\quad = \frac{L_e^{L_e d}|\mathbf{C}|^{L_e-d}}{|\mathbf{\Sigma}|^{L_e}} \frac{\Gamma(L_e d+\lambda)(\lambda-1)^\lambda}{\Gamma_d(L_e)\Gamma(\lambda)} \\ &\quad \times \left(L_e \operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{C}) + \lambda - 1\right)^{-\lambda-L_e d} . \end{aligned} \qquad (11)$$

For interpretation purposes, we note that $z \to 1$ and the distributions in Eqs. (9) and (11) converge in distribution to the complex Wishart distribution in Eq. (5) as $\alpha \to \infty$ and $\lambda \to \infty$, respectively. Thus, high values of α and λ imply little texture, whereas low values refer to significant texture and non-Gaussianity.

2.3. Relaxed Wishart Model

The standard Wishart model in Eq. (5) is parametrised by a constant L_e, which is estimated for the data set as a whole [3, 4]. We introduce a relaxed Wishart ($\mathcal{RW}$) model, whose functional form is identical. The difference is that L_e is replaced with a variable shape parameter, $\mathscr{L} \leqslant L_e$. Depending on the application, $\mathscr{L}$ is allowed to vary between classes (classification), segments (segmentation), or pixels (e.g., change detection). The new distribution is denoted by $\mathbf{C} \sim \mathcal{RW}_d^{\mathrm{C}}(\mathscr{L},\mathbf{\Sigma})$.

The motivation for this approach is explained by Fig. 1. It is not possible to visualise the effect of the distribution parameters directly on the pdfs in Eqs. (5), (9), and (11). We therefore plot their marginial densities for a single polarisation intensity. The respective marginals are gamma distributed, $\mathcal{K}_I$ distributed, and $\mathcal{G}_I^0$ distributed. For the latter two, the superscripted I denotes the multilook intensity version of the given distribution family.

In all the plots, the continuous curve represents the limiting case defined by the standard Wishart model, with a gamma distributed marginal pdf. In the left panel, the dashed curves show the evolution of the pdf under the $\mathcal{RW}$ model as $\mathscr{L}$ is lowered from the limit of $\mathscr{L} = L_e$. The same evolution is illustrated for the $\mathcal{K}$ distribution (middle panel) and the $\mathcal{G}^0$ distribution (right panel) for decreasing values of the respective texture parameters, α and λ. We observe that the effect of varying $\mathscr{L}$ resembles that induced by α and λ, even though a greater variation in shape is possible for the distributions based on the product model. Fig. 2 uses a logarithmic scale to highlight the heavy tails of the $\mathcal{K}$ and $\mathcal{G}^0$ distribution., which is less prominent for the marginal pdf of the $\mathcal{RW}$ distribution. We still conclude that $\mathscr{L}$ can be interpreted as a texture parameter alongside α and λ. Thus, the $\mathcal{RW}$ distribution implicitly models texture up to a moderate level.

3. GOODNESS-OF-FIT EVALUATION

This section discusses evaluation of the goodness-of-fit (GoF) for the matrix-variate density models of multilook PolSAR data. GoF testing in the literature has been limited to visual inspection of how well marginal densities of intensity fit histograms of the data. Classical statistical distribution tests, such as the Kolmogorov-Smirnov test or the Anderson-Darling test, are impractical in this case. As noted in [1], this is because they require binning of the domain of $\mathbf{C}$, which is the cone of positive definite matrices.

We here propose an alternative space where GoF evaluation can be performed. The idea is that GoF can be assessed by comparing theoretical moments of the models with sample moments computed from the data. We first define a new kind of matrix moments, that we call log-determinant cumulants. Closed form expressions for the candidate models are then derived. These are seen to have favourable properties that can be utilised to visualise the texture modelling capabilities of the models.

3.1. Log-determinant Cumulants

The following derivation is based on the application of second kind statistics, following the terminology introduced in [12]. Whereas the commonly known characteristic function is defined as the Fourier transform of a pdf, the second kind characteristic function is the Mellin transform of the pdf. This function can be used to generate moments and cumulants of the second kind, also termed log-moments and log-cumulants.

Let ξ be a real, positive rv with pdf $p_\xi(\xi)$. Start by defining the rth-order log-moment of ξ as

$$m_r(\xi) = \mathrm{E}\{(\ln \xi)^r\} = \left. \frac{d^r \phi_\xi(s)}{ds^r} \right|_{s=1} , \qquad (12)$$

where $\phi_\xi(s)$ is the Mellin transform of $p_\xi(\xi)$ and $s \in \mathbb{C}$ [12]. Then define the rth-order log-cumulant of ξ as

$$\kappa_r(\xi) = \left. \frac{d^r \psi_\xi(s)}{ds^r} \right|_{s=1} , \qquad (13)$$

where $\psi_\xi(s) = \ln \phi_\xi(s)$. Relations between some low-order log-moments and log-cumulants are given by

$$\kappa_1 = m_1 , \qquad (14)$$
$$\kappa_2 = m_2 - m_1^2 , \qquad (15)$$
$$\kappa_3 = m_3 - 3m_1 m_2 + 2m_1^3 . \qquad (16)$$

It follows from a fundamental property of the Mellin transform [12] that for a product of independent random variables, $\xi = \rho \cdot \zeta$, with $\rho, \zeta > 0$:

$$\kappa_r(\xi) = \kappa_r(\rho) + \kappa_r(\zeta) , \quad \forall r \in \mathbb{N} . \qquad (17)$$

This equips us to derive the log-cumulant of $|\mathbf{C}|$, which will be referred to as the log-determinant cumulant (LDC) of $\mathbf{C}$.

Note that $|\mathbf{C}| = |z\mathbf{W}| = z^d |\mathbf{W}|$. Thus,

$$\kappa_r(|\mathbf{C}|) = d^r \kappa_r(z) + \kappa_r(|\mathbf{W}|) . \qquad (18)$$

The log-cumulants of z have been derived in [12]. For $z \sim \mathcal{G}(\alpha, \mu)$ it was shown that

$$\kappa_1(z) = \ln\left(\frac{\mu}{\alpha}\right) + \Psi^{(0)}(\alpha) , \qquad (19)$$
$$\kappa_r(z) = \Psi^{(r-1)}(\alpha) , \quad r > 1 , \qquad (20)$$

where $\Psi^{(r)}(z)$ is Euler's polygamma function of order r. The log-cumulants of $z \sim \mathcal{G}^{-1}(\lambda, \nu)$ were found as

$$\kappa_1(z) = \ln \lambda\nu - \Psi^{(0)}(\lambda) , \qquad (21)$$
$$\kappa_r(z) = (-1)^r \, \Psi^{(r-1)}(\lambda) . \qquad (22)$$

The LDCs of the Wishart distributed $\mathbf{W}$ can be deduced from results found in [3, 4] as

$$\kappa_1(|\mathbf{W}|) = \ln|\mathbf{\Sigma}| + \sum_{i=0}^{d-1} \Psi^{(0)}(L_e - i) - d \ln L_e \qquad (23)$$

$$\kappa_r(|\mathbf{W}|) = \sum_{i=0}^{d-1} \Psi^{(r-1)}(L_e - i) , \quad r > 1 . \qquad (24)$$

This completes the expression in Eq. (18) for our candidate models.

3.2. Log-determinant Cumulant Diagram

Note that the LDCs are matrix-variate generalisation of the log-cumulants derived in [12] for the single polarisation product model. As in the one-dimensional case, we can utilise the fact the LDCs do not depend on the scale parameter $\mathbf{\Sigma}$ for $r > 1$. More specifically, the $\kappa_{r>1}(|\mathbf{C}|)$ depend only on the texture parameters:

$$\kappa^{\mathcal{W}}_{r>1}(|\mathbf{C}|) = \sum_{i=0}^{d-1} \Psi^{(r-1)}(L_e - i) \qquad (25)$$

$$\kappa^{\mathcal{K}}_{r>1}(|\mathbf{C}|) = d^r \Psi^{(r-1)}(\alpha) + \sum_{i=0}^{d-1} \Psi^{(r-1)}(L_e - i) \qquad (26)$$

$$\kappa^{\mathcal{G}^0}_{r>1}(|\mathbf{C}|) = (-d)^r \Psi^{(r-1)}(\lambda) + \sum_{i=0}^{d-1} \Psi^{(r-1)}(L_e - i) \qquad (27)$$

$$\kappa^{\mathcal{RW}}_{r>1}(|\mathbf{C}|) = \sum_{i=0}^{d-1} \Psi^{(r-1)}(\mathscr{L} - i) , \qquad (28)$$

where the superscript of κ indicates which model the expression describes.

By plotting two LDCs of different orders against each other, we obtain a curve in LDC space which depicts the

paired LDC values that can be attained under a given model. We refer to this as an LDC diagram. Sample LDCs calculated from data can be overlaid the model curves, and the diagram used to assess how well the data are described by different models, and which model provides the best fit. Diagrams of second and third order log-cumulants were plotted in [12, 13], and we will use the same orders for our LDC diagrams. Remark that the bias and variance of the sample LDCs are expected to increase rapidly with order.

4. EXPERIMENTS

4.1. Marginal Densities of Intensity

We have extracted three test samples from an L-band quadrature polarisation image aquired by the airborne NASA/JPL AIRSAR sensor over Flevoland, the Netherlands, in August 1989. The samples are taken from some of the more textured areas in the image. One is from a forest area and the other two from different crops. Marginal densities of the intensity in the HH, HV, and VV channels for the forest sample are shown in Fig. 3. These densities describe unfiltered data. Fig. 4 describes the same data after they have been filtered with a refined Lee filter [14] of window size 7×7.

The first observation is that the data are well described by all the models in Fig. 3. By zooming in on the densities, it may be concluded by visual inspection that the $\mathcal{K}$ model provides the best fit, followed by the $\mathcal{RW}$ model. The ENL estimated for the data set, and used to parametrise the standard Wishart, $\mathcal{K}$, and $\mathcal{G}^0$ model, is $L_e = 3.3$. This constant is replaced with $\mathscr{L} = 2.53$ for the $\mathcal{RW}$ model. The texture parameters α and λ are estimated with the method described for the $\mathcal{K}$ distribution in [10].

Fig. 4 shows that the models have very different GoF for speckle filtered data. The Wishart model is the worst fit, and none of the distributions based on the product model produce an adequate result either. Only the $\mathcal{KW}$ model seems to do a good job. The ENL was estimated to $L_e = 48$, and is reduced to $\mathscr{L} = 27$ for the $\mathcal{RW}$ model. The marginal densities of the other two test samples yield very similar results, both before and after speckle filtering, and are therefore not shown.

4.2. Log-determinant Cumulant Diagrams

Fig. 5 shows a LDC diagram where $\kappa_3(|\mathbf{C}|)$ is plotted against $\kappa_2(|\mathbf{C}|)$, with analogy to the log-cumulant diagrams in [12, 13]. The Wishart model has no texture parameter, and its LDCs are therefore constant, equal to the contribution $\kappa_r(|\mathbf{W}|)$, $r = 1, 2$ of the Wishart distributed speckle matrix. These constants are indicated in the figure by the dotted lines, intersecting at the point which describes the Wishart model. The possible LDC pairs of the $\mathcal{K}$, $\mathcal{G}^0$, and $\mathcal{RW}$ models lie on a curve parametrised by α, λ, and $\mathscr{L}$, respectively. The asymptotic behaviour of these curves, as the texture parameters decrease towards their lower limits, is indicated on the figure. At the upper limit, the curves all converge to the Wishart case. Sample LDCs of the three test samples are plotted as points in green (forest), cyan, and magenta (two different crops). We plot a collection of sample LDC estimates, obtained by bootstrap sampling of the full test samples, in order to illustrate the dispersion of the sample estimates. ENL estimates for each test sample are shown in the figure.

Diagrams of data before and after speckle filtering are presented in the left and right panel, respectively. For the unfiltered data, the LDC diagram clearly indicates that the $\mathcal{RW}$ distribution is the best model for the forest sample. The other test samples are less textured, and all models are adequate. For the speckle filtered data, the LDC diagram suggests that the $\mathcal{RW}$ model fits best for the forest sample and the first crop sample (cyan), while the $\mathcal{K}$ model performs best for the second crop sample (magenta). For the crop samples, both the $\mathcal{RW}$ model and the $\mathcal{K}$ model fit reasonably well. The Wishart and the $\mathcal{G}^0$ model are inadequate in all cases. The LDC diagram indicates good fit of the $\mathcal{K}$ model to the crop samples, which is not compatible with observations of the marginal densities (not shown). This prompts us to reconsider the estimator for α (and λ) in future work.

The success of the $\mathcal{RW}$ distribution in modelling of speckle filtered data, and the relative failure of the others, can be explained by a discussion of the nature of adaptive speckle filters. Adaptive speckle filters apply variable smoothing by consideration of local homogeneity. Hence, the ENL is mapped from a constant value to a dispersed range of values. This is not modelled appropriately, neither by the Wishart distribution nor the other distributions based on the product model. The $\mathcal{RW}$ distribution, on the other hand, apparently represents a better approach.

5. CONCLUSIONS

We have proposed a relaxed Wishart distribution where the equivalent number of looks of the standard Wishart model has been replaced by a varable shape parameter. We have further derived the log-determinant cumulants of the polarimetric covariance (or coherency) matrix under the product model, and demonstrated how they can be utilised in visual inspection of goodness-of-fit for matrix-variate distributions. Experimental results show that for a moderate level of texture, the newly proposed density can compete with densities derived from the product model with regards to modelling of unfiltered PolSAR data. For data that are processed with an adaptive speckle filter, the relaxed Wishart model is shown to perform better. Based on the very promising results, we suggest that the relaxed Wishart distribution should be tested more extensively on other data sets and with different speckle filters. It should also be applied to model-based classification, change detection, and other image analysis tasks.

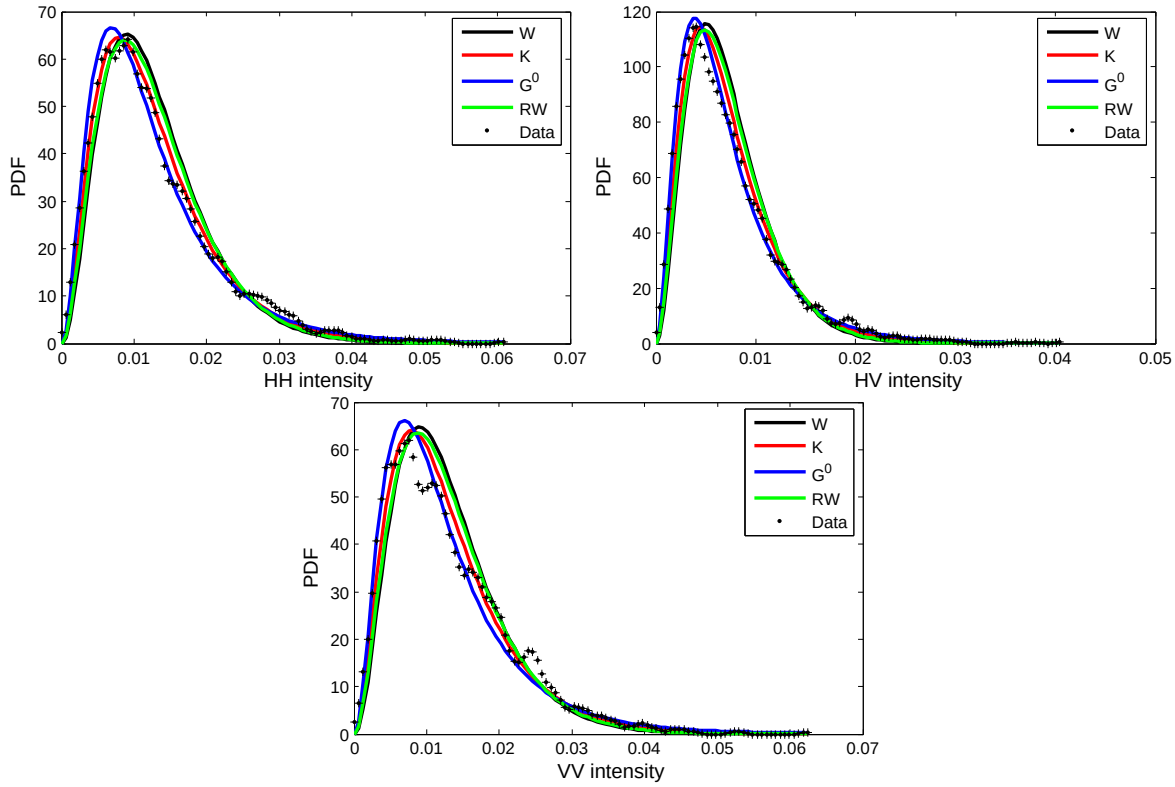

Figure 3. Comparison of marginal densities of the Wishart, $\mathcal{RW}$, $\mathcal{K}$, and $\mathcal{G}^0$ distribution with data histograms for a textured forest area in the AIRSAR L-band image of Flevoland. No speckle filter applied.

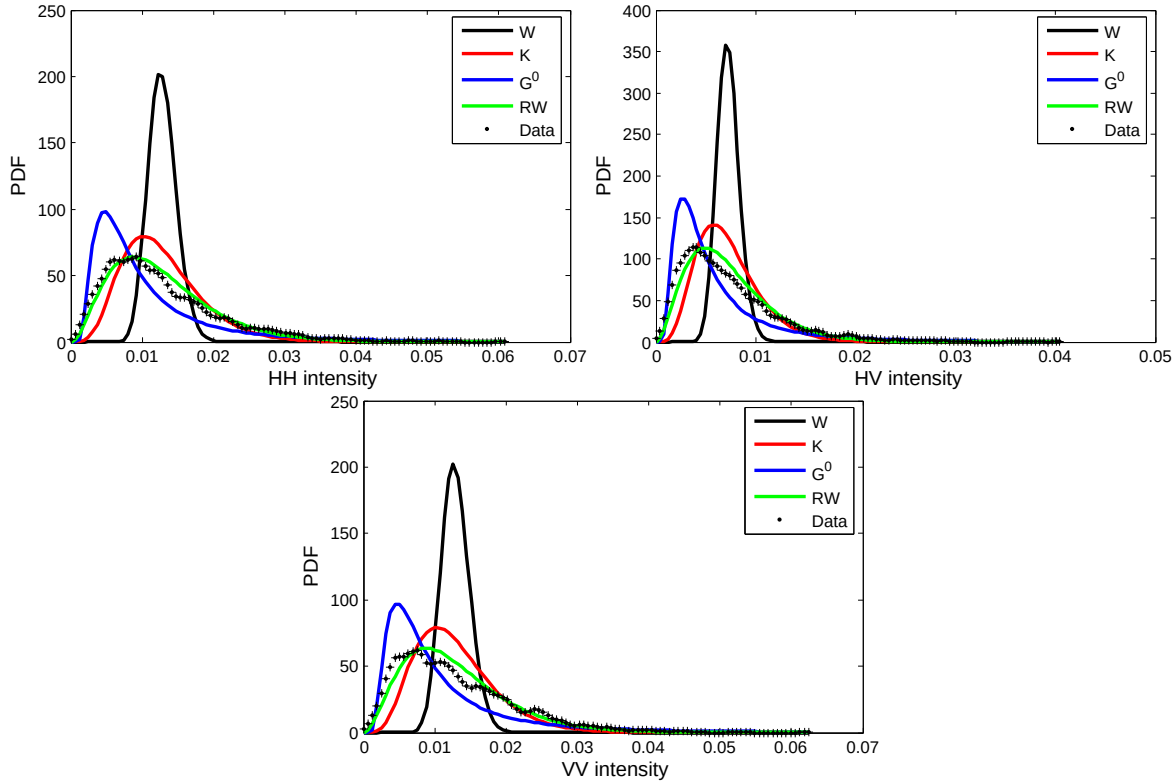

Figure 4. Comparison of marginal densities of the Wishart, $\mathcal{RW}$, $\mathcal{K}$, and $\mathcal{G}^0$ distribution with data histograms for a textured forest area in the AIRSAR L-band image of Flevoland. Modified Lee filter with window size 7×7 applied.

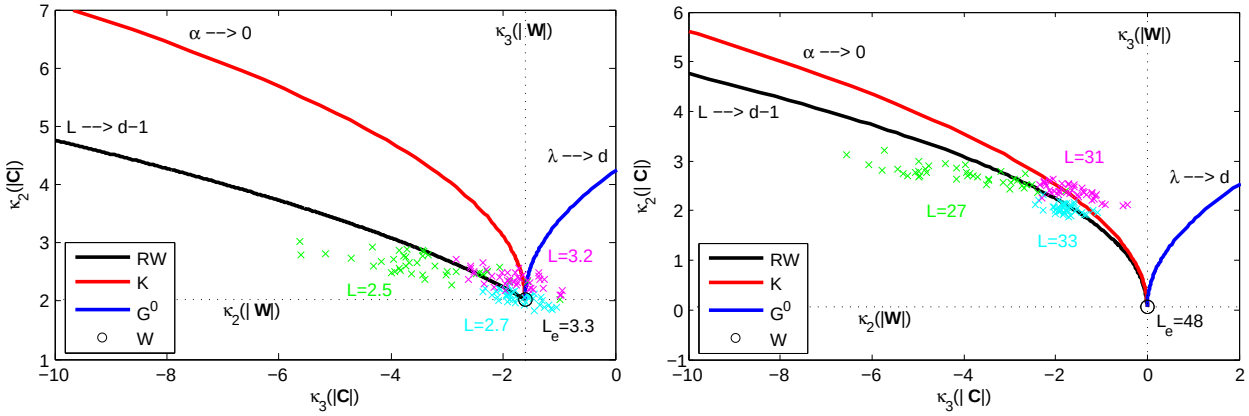

Figure 5. Diagram of second and third-order log-determinant cumulants (LDCs) for the Wishart, $\mathcal{RW}$, $\mathcal{K}$, and $\mathcal{G}^0$ distribution. Sample LDCs for three test samples from the AIRSAR L-band image of Flevoland are plotted in green (forest), cyan and magenta (two different crops). Results are shown before (left panel) and after (right panel) speckle filtering with a 7×7 refined Lee filter.

ACKNOWLEDGEMENTS

The authors would like to thank NASA/JPL-Caltech for making available the AIRSAR data set used in the paper. We further thank the European Space Agency and the POLSARPRO software development team for their continued efforts in publishing open source software and PolSAR data samples.

REFERENCES

1. C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric $\mathcal{G}$ distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, Feb. 2005.

2. 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 Sensing Symp., IGARSS'94*, vol. 4, Pasadena, USA, Aug. 1994, pp. 2179–2181.

3. S. N. Anfinsen, A. P. Doulgeris, and T. Eltoft, "Estimation of the equivalent number of looks in PolSAR imagery," in *Proc. IEEE Int. Geosc. Remote Sensing Symp., IGARSS'08*, vol. 4, Boston, USA, July 2008, pp. 487–490.

4. ——, "Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery," *IEEE Trans. Geosci. Remote Sensing*, in review.

5. J.-S. Lee, M. R. Grunes, T. L. Ainsworth, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart distribution," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 5, pp. 2249–2259, Sept. 1999.

6. N. 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.

7. C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, 2nd ed. SciTech Publishing, 2004.

8. R. Touzi, W. M. Boerner, J.-S. Lee, and E. Lüneburg, "A review of polarimetry in the context of synthetic aperture radar: Concepts and information extraction," *Can. J. Remote Sensing*, vol. 30, no. 3, pp. 380–407, 2004.

9. S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain radar clutter," *J. Electrom. Waves Applic.*, vol. 3, no. 8, pp. 747–768, 1989.

10. A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sensing*, vol. 46, no. 10, pp. 2999–3009, Oct. 2008.

11. 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 Sensing*, vol. 35, no. 3, pp. 648–659, May 1997.

12. 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.

13. ——, "A Fisher-MAP filter for SAR image processing," in *Proc. IEEE Int. Geosc. Remote Sensing Symp., IGARSS'03*, vol. 3, Toulouse, France, July 2003, pp. 1996–1998.

14. J.-S. Lee, M. R. Grunes, and G. de Grandi, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 5, pp. 2363–2373, Sept. 1999.

Appendix B

Introduction to Second Kind Statistics: Application of Log-Moments and Log-Cumulants to Analysis of Radar Images

Introduction to Second Kind Statistics: Application of Log-Moments and Log-Cumulants to Analysis of Radar Images

Jean-Marie Nicolas and Stian Normann Anfinsen (translator)

Abstract—Statistical methods classically used to analyse a probability density function (pdf) are founded on the Fourier transform, on which useful tools such the first and second characteristic function are based, yielding the definitions of moments and cumulants. Yet this transformation is badly adapted to the analysis of probability density functions defined on $\mathbb{R}^+$, for which the analytic expressions of the characteristic functions may become hard, or even impossible to formulate. In this article we propose to substitute the Fourier transform with the Mellin transform. It is then possible, inspired by the precedent definitions, to introduce *second kind statistics*: second kind characteristic functions, second kind moments (or log-moments), and second kind cumulants (or log-cumulants). Applied to traditional distributions like the gamma distribution or the Nakagami distribution, this approach gives results that are easier to apply than the classical approach. Moreover, for more complicated distributions, like the $\mathcal{K}$ distributions or the positive α-stable distributions, the second kind statistics give expressions that are truly simple and easy to exploit. The new approach leads to innovative methods for estimating the parameters of distributions defined on $\mathbb{R}^+$. It is possible to compare the estimators obtained with estimators based on maximum likelihood theory and the method of moments. One can thus show that the new methods have variances that are considerably lower than those mentioned, and slightly higher than the Cramér-Rao bound.

Index Terms—Probability density functions defined on $\mathbb{R}^+$, gamma distribution, Nakagami distribution, characteristic functions, parameter estimation, Mellin transform

1 Introduction

Estimation of the parameters of a probability density functions (pdf) is a topic of major significance in pattern recognition. Starting from these estimates, segmentation and classification algorithms can be implemented, both in the field of signal processing and image processing. In signal processing, the intrinsic knowledge of the nature of the data (provided by an acoustic sensor, electromagnetic sensor, etc.) allows us to make realistic assumptions about the suitable distribution models. In particular, many techniques are based on the additive noise model with a noise term that is assumed to be Gaussian. Traditionally, if one describes a random phenomenon by a pdf, one will also introduce the concept of the characteristic function, defined as the Fourier transform $\mathcal{F}$ of the pdf. For example, if $p_x(u)$ is the pdf modelling a random variable X, the characteristic function $\Phi_x(v)$ is obtained by the relation [1]:

$$\Phi_x(v) = \mathcal{F}[p(u)](v) = \int_{-\infty}^{+\infty} e^{jvu} p_x(u)\, du\,. \tag{1}$$

The second characteristic function is defined as the logarithm of the characteristic function:

$$\Phi_x(v) = \log(\Psi_x(v))\,. \tag{2}$$

By taking account of properties of the Fourier transform, it is easy to show that moments of order n are obtained by derivation of the characteristic function:

$$\begin{aligned} m_n &= \int_{-\infty}^{+\infty} u^n p_x(u) du \\ &= (-j)^n \frac{d^n \Psi_X(v)}{dv^n}\bigg|_{v=0} \end{aligned} \tag{3}$$

and cumulants of order n by derivation of the second characteristic function:

$$\kappa_{x(r)} = (-j)^n \frac{d^r \Psi_x(v)}{dv^r}\bigg|_{v=0}\,.$$

Moreover, if a phenomenon is analysed, described by a pdf q_y, which is perturbated by an additive noise, described by its pdf r_z, one knows that the output signal is described by the pdf p_x given as

$$p_x = q_y * r_z\,, \tag{4}$$

with the operator $*$ denoting convolution. It is known that the characteristic functions and the cumulants can be written:

$$\Phi_x(s) = \Phi_y(s)\Phi_z(s) \tag{5}$$

$$\Psi_x(s) = \Psi_y(s) + \Psi_z(s) \tag{6}$$

$$\kappa_{x(r)} = \kappa_{y(r)} + \kappa_{z(r)}\ \forall r \tag{7}$$

The author is with the École Nationale Supérieure des Télécommunications, Département TSI, 46 rue Barrault, 75634 Paris cedex 13 (e-mail: jean-marie.nicolas@telecom-paristech.fr).

The original paper was published as: J.-M. Nicolas, "Introduction aux statistiques de deuxième espèce: applications des logs-moments et des logs-cumulants à l'analyse des lois d'images radar", Traitement du Signal, vol. 19, no. 3, pp. 139–167, 2002.
Translated from French by Stian Normann Anfinsen, University of Tromsø, Department of Physics and Technology, NO-9037 Tromsø, Norway (e-mail: stian.normann.anfinsen@uit.no).

However, in image processing the problems are different. It should be noted first of all that the pixel values are positive or zero (we will not discuss here the analysis of images defined by complex values), and that the noise is often multiplicative. Also, the preceding model must undergo some adaptations to be applicable as it is. One approach often proposed is to perform a logarithmic transformation, which is possible since the pdf is defined on $\mathbb{R}^+$. Several remarks can then be made:

- The analytical calculation of the characteristic functions defined on $\mathbb{R}^+$ is sometimes hard, even impossible for certain distributions, as we will see in section 3.2.
- No complete methodology is proposed for logarithmically transformed data. Calculation of moments on logarithmic scale (that one may call log-moments) is carried out analytically starting from Eq. (3). It requires a change of variable (thus a rewriting of the pdf for this new variable) and is carried out in a specific way for each pdf. This approach requires a good knowledge of integral transforms and of the properties of special functions.
- In traditional statistics, the Gaussian distribution is the reference, which corresponds to the log-normal distribution on a logarithmic scale. However, in many examples, this law does not describe the studied phenomenon well. In particular, the speckle (clutter) observed in images obtained by coherent illumination (e.g., laser, radar, ultrasound) follows, for intensity images, the gamma distribution, which we will study in more detail in this article and which tends asymptotically towards a degenerated Gaussian distribution.

As we will show, a new methodology based on another integral transformation, the Mellin transform [2], makes it possible to perform a more effective analysis of practically important distributions defined on $\mathbb{R}^+$. This methodology, that we propose to call *second kind statistics*, uses the same framework as traditional statistics for the definition of the characteristic functions (simply by replacing the Fourier transform with the Mellin transform in Eq. (1)) and the same construction of the moments and cumulants (by derivation of the characteristic functions). This leads naturally to the definitions of *second kind moments* and *second kind cumulants*. We shall see why we propose to call these new entities *log-moments* and *log-cumulants*. Thanks to this approach, it is possible to analyse in a simpler way distributions with two or three parameters that have traditionally been used for imagery: the gamma distribution, Nakagami distribution, and $\mathcal{K}$ distribution. Then, we will see how to tackle more complex problems like the distributions of the Pearson system, additive mixtures and distributions with heavy tails (i.e., distributions for which the moments cannot be defined starting from a certain order[1]). Finally, we will analyse why the parameter estimators of these distributions based on the log-moments and log-cumulants have a lower variance than those obtained from the traditional moments and cumulants.

The remark can be made that a formalism with such similarity to the existing definitions cannot lead to intrinsically new results. It should be stressed that the essential contribution of this framework is to offer a signal and image processing methodology which proves, for certain applications, considerably easier to use than the traditional approaches. The major goal of this article is to illustrate its simplicity of implementation as well as its flexibility in use.

2 Definition of the Second Kind Characteristic Functions

The objective of this section is to propose a formalism of *second kind statistics* based on the Mellin transform and redefine some elements of traditional statistics, namely the characteristic function yielding moments and cumulants, as outlined in the introduction.

2.1 First Characteristic Function of the Second Kind

Let X be a positive-valued random variable whose pdf, $p_x(u)$, is defined for $u \in \mathbb{R}^+$. The *first characteristic function of the second kind* is defined as the Mellin transform $\mathcal{M}$ of $p_x(u)$:

$$\phi_x(s) = \mathcal{M}[p_x(u)](s) = \int_0^{+\infty} u^{s-1} p_x(u)\, du \qquad (8)$$

provided that this integral converges, which is verified in general only for values of s located inside a strip delimited by two lines parallel to the secondary axis, i.e.

$$s = a + jb, \quad a \in]a1; a2[, \quad b \in \mathbb{R}$$

with a_2 commonly approaching $+\infty$, just as a_1 approaches $-\infty$. As the Mellin transform has an inverse [2], knowing $\phi_x(s)$, one can deduce $p_x(u)$ thanks to the relation:

$$p_x(u) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} u^{-s} \phi_x(s)\, ds$$

given that c is confined within the strip where the first characteristic function is defined (i.e, $c \in]a_1; a_2[$). Note that if $p_x(u)$ is a pdf, the second kind characteristic function satisfies the fundamental property:

$$\phi_x(s)|_{s=1} = 1\,.$$

1. Translator's remark: Note that author uses a strict definition of heavy-tailed distributions. An alternative and more common definition is that heavy-tailed distributions are not exponentially bounded. That is, they have heavier tails than the exponential distribution. Since the context of the discussion is modelling of multilook intensity radar data, it would be natural to replace the exponential distribution with the (generalised) gamma distribution in this criterion.

By analogy, the *second kind moments*, $\tilde{m}_\nu$ ($\nu \in \mathbb{N}$) are defined by the relation:

$$\tilde{m}_\nu = \left. \frac{d^\nu \phi_x(s)}{ds^\nu} \right|_{s=1} . \tag{9}$$

By virtue of a fundamental property of the Mellin transform [Col59]:

$$\mathcal{M}[f(u)(\log u)^\nu](s) = \frac{d^\nu \mathcal{M}[f(u)](s)}{ds^\nu}$$

which is evaluated at $s = 1$, the second kind moments can be written in two different ways:

$$\tilde{m}_\nu = \left. \frac{d^\nu \phi_x(s)}{ds^\nu} \right|_{s=1} \tag{10}$$

$$= \int_0^{+\infty} (\log u)^\nu \, p_x(u) \, du \, . \tag{11}$$

Eq. (11) suggests that we refer to the second kind moments as *log-moments*, which is adopted for the remainder of the article.

We now introduce the second kind mean or *log-mean* $\tilde{\bar{m}}$. This auxiliary variable is defined by the following relation

$$\log \tilde{\bar{m}} = \tilde{m}_1 \quad \Leftrightarrow \quad \tilde{\bar{m}} = e^{\tilde{m}_1} .$$

Note that this entity, which is in fact the geometric mean, takes its values in $\mathbb{R}^+$ (a suitable scale for the variable u), whereas the log-moments take their values in $\mathbb{R}$ (on logarithmic scale). It is thus possible to compare the mean $\bar{m}$ and the log-mean $\tilde{\bar{m}}$, and the practical importance will be demonstrated for the gamma distribution.

Just as one traditionally defines the central moments, we introduce the definition of the central log-moments of order n, $\tilde{M}_n$:

$$\begin{aligned} \tilde{M}_n &= \int_0^{+\infty} (\log u - \tilde{m}_1)^n \, p_x(u) \, du \\ &= \int_0^{+\infty} \left(\log \frac{u}{\tilde{\bar{m}}} \right)^n \, p_x(u) \, du \, . \end{aligned} \tag{12}$$

In particular, one readily finds the expression

$$\tilde{M}_2 = \tilde{m}_2 - \tilde{m}_1^2 .$$

Thanks to this formalism, it is possible to obtain an analytical expression for the log-moments by simple derivation of the second kind characteristic function. We will look at the classical interpretation of the Mellin transform.

2.2 A First Interpretation of the Mellin Transform

By comparison of the moment definition in Eq. (3) and the definition of the first characteristic of the second kind in Eq. (8), one can write the generalised moments, m_ν:

$$m_\nu = \phi_x(s)|_{s=\nu+1} = \int_0^{+\infty} u^\nu p_x(u) \, du \, . \tag{13}$$

For $\nu \in \mathbb{N}$, these are the traditional moments. For $\nu \in \mathbb{R}^+$, we have the fractional moments, that have been used by some authors (like the use of FLOM: Fractional Low Order Moments, in [3]). Provided that the Mellin transform is defined for values of $\nu \in \mathbb{R}^-$, it is justified to use *lower order moments* [4]. Lastly, in addition to moments defined for a value $\nu = a$ (i.e., traditional moments, fractional moments, or lower order moments), one can define moments of complex order with $\nu = a + jb$ for all b, this because the pdf $p_x(u)$ is positive by definition, a property which is trivial to verify.

2.3 Second Kind Cumulants or Log-Cumulants

Still by analogy with classical statistic for scalar real random variables defined on $\mathbb{R}$, the *second characteristic function of the second kind* is defined as the natural logarithm of the first characteristic function of the second kind:

$$\psi_x(s) = \log(\phi_x(s)) \, . \tag{14}$$

The derivative of the second characteristic function of the second kind, evaluated at $s = 1$, defines *second kind cumulants* of order n:

$$\tilde{\kappa}_{x(n)} = \left. \frac{d^n \psi_x(s)}{ds^n} \right|_{s=1} . \tag{15}$$

Since formally, second kind cumulants are constructed according to the same rules as traditional cumulants, the relations between log-moments and log-cumulants are identical to the relations existing between moments and cumulants. For instance, the three first log-cumulants can be written as:

$$\begin{aligned} \tilde{\kappa}_1 &= \tilde{m}_1 \\ \tilde{\kappa}_2 &= \tilde{m}_2 - \tilde{m}_1^2 \\ \tilde{\kappa}_3 &= \tilde{m}_3 - 3\tilde{m}_1\tilde{m}_2 + 2\tilde{m}_1^3 \end{aligned}$$

As in the case of log-moments, we adopt the name *log-cumulants* for the second kind cumulants.

2.4 Some Properties of Log-Moments and Log-Cumulants

The utilisation of the Mellin transform requires knowledge about some of its specific properties. In particular, let us point out the definition of the Mellin convolution (which is an associative and commutative operation):

$$\begin{aligned} h = f \,\hat{\star}\, g \;\Leftrightarrow\; h &= \int_0^{\infty} f(y) g\left(\frac{u}{y}\right) \frac{dy}{y} \\ \Leftrightarrow\; h &= \int_0^{\infty} g(y) f\left(\frac{u}{y}\right) \frac{dy}{y} \, , \end{aligned} \tag{16}$$

Its fundamental property is similar to the convolution property of the Fourier transform:

$$\mathcal{M}[f \,\hat{\star}\, g](s) = \mathcal{M}[f](s) \, \mathcal{M}[g](s) \, .$$

Note that if f and g are pdfs, then h is also a pdf (i.e., $h(u) \geq 0 \; \forall u \in \mathbb{R}^+$ and $\mathcal{M}[h]|_{s=1} = 1$).

The use of this operator finds an immediate application in the study of multiplicative noise. Let Y and Z

be two independent random variables whose respective pdfs, q_y and r_z, are defined on $\mathbb{R}^+$. Consider a random variable X constructed by a multiplication of of these two variables. It is thus a model of multiplicative noise. It is then shown that the pdf of X, p_x, is obtained as the Mellin convolution of q_y and r_z [5], [6]:

$$p_x = q_y \,\hat{\star}\, r_z \,.$$

The properties deduced in the following are formally identical to those obtained in the case of a traditional convolution (Eqs. (5)-(7)). If ϕ_x is the second kind characteristic function of X, ϕ_y is the second kind characteristic function of Y and ϕ_z is the second kind characteristic function of Z, the following relations are obtained:

$$\begin{aligned} \phi_x(s) &= \phi_y(s)\,\phi_z(s) \\ \psi_x(s) &= \psi_y(s) + \psi_z(s) \\ \tilde{\kappa}_{x(n)} &= \tilde{\kappa}_{y(n)} + \tilde{\kappa}_{z(n)} \quad \forall n \in \mathbb{N} \end{aligned} \tag{17}$$

It is noted in particular that in the case of multiplicative noise, the log-cumulants are additive. This property is not surprising since the common method used to handle multiplicative noise, transformation into logarithmic scale, allows us to treat noise of multiplicative nature like additive noise.

Finally note the following property:

$$u\,(f \,\hat{\star}\, g) = (u\,f) \,\hat{\star} (u\,g)\,.$$

One can also, in a step similar to that of traditional convolution, define the inverse convolution (a non-commutative and non-associative operator). If the ratio

$$\frac{\mathcal{M}[f](s)}{\mathcal{M}[g](s)}$$

is defined in the vicinity of $s = 1$ such that the inverse Mellin transform can be evaluated, the following relation is posed:

$$h = f \,\hat{\star}^{-1}\, g \quad \Leftrightarrow \quad \mathcal{M}[h](s) = \frac{\mathcal{M}[f](s)}{\mathcal{M}[g](s)}\,.$$

With the above notation we have, given that the pdfs p_x, q_y and r_z exist:

$$p_x = q_y \,\hat{\star}^{-1}\, r_z \,,$$

from which we deduce:

$$\begin{aligned} \phi_x(s) &= \frac{\phi_y(s)}{\phi_z(s)} \\ \psi_x(s) &= \psi_y(s) - \psi_z(s) \\ \tilde{\kappa}_{x(n)} &= \tilde{\kappa}_{y(n)} - \tilde{\kappa}_{z(n)} \quad \forall n \in \mathbb{N} \end{aligned} \tag{18}$$

Finally, it can be useful to utilise the *Mellin correlation* (also a non-associative and non-commutative operation), which is defined by the relation:

$$h = f \,\hat{\otimes}\, g \;\Leftrightarrow\; \mathcal{M}[h](s) = \mathcal{M}[f](s)\,\mathcal{M}[g](2-s)\,. \tag{19}$$

A pdf must satisfy $\mathcal{M}[h]|_{s=1} = 1$, to which h complies. Starting from this relation and using the same notation, we can, provided that p_x satisfies

$$p_x = q_y \,\hat{\otimes}\, r_z \,,$$

deduce the following:

$$\begin{aligned} \phi_x(s) &= \frac{\phi_y(s)}{\phi_z(2-s)} \\ \psi_x(s) &= \psi_y(s) - \psi_z(2-s) \\ \tilde{\kappa}_{x(n)} &= \tilde{\kappa}_{y(n)} + (-1)^n \tilde{\kappa}_{z(n)} \quad \forall n \in \mathbb{N} \end{aligned} \tag{20}$$

The following expression can then be shown:

$$h = f \,\hat{\otimes}\, g \;\Leftrightarrow\; h = \int_0^\infty f(uy)\,g(y)\,y\,dy\,. \tag{21}$$

We also note the property:

$$u\,(f \,\hat{\otimes}\, g) = (u\,f) \,\hat{\otimes} \left(\frac{g}{u}\right) \tag{22}$$

In fact, the interpretation of the Mellin correlation is founded on the analysis of the inverse distribution, i.e., the distribution $p_I(u)$ of the random variable $Y = 1/X$, where the random variable X follows the distribution $p(u)$. The relation between these distributions are known to be:

$$p_I(u) = \frac{1}{u^2} p\left(\frac{1}{u}\right)\,.$$

By taking account of a fundamental property of the Mellin transform:

$$\mathcal{M}\left[\frac{1}{u} f\left(\frac{1}{u}\right)\right](s) = \mathcal{M}[f(u)](1-s)\,,$$

it is easily deduced that

$$\mathcal{M}[p_I](s) = \mathcal{M}[p](2-s)\,. \tag{23}$$

It is then seen that the Mellin correlation of a pdf q_y of the random variable Y and a pdf r_z of the random variable Z,

$$q_y \,\hat{\otimes}\, r_z \,,$$

is simply a way to establish the pdf of the random variable Y/Z.

Lastly, as for the traditional characteristic function, it is interesting to note that the second kind characteristic function can be expanded in terms of log-cumulants:

$$\begin{aligned} \psi_x(s) = \tilde{\kappa}_{x(1)}(s-1) + \frac{1}{2!}\tilde{\kappa}_{x(2)}(s-1)^2 \\ + \frac{1}{3!}\tilde{\kappa}_{x(3)}(s-1)^3 + \ldots \end{aligned}$$

2.5 Theorem of Existence of Log-Moments and Log-Cumulants

We have just seen that the theoretical introduction of the log-moments and log-cumulants does not pose any formal problem. However, the existence of these entities has not been proven, and an interrogation into the requirements for their existence is needed. In this section we will present a theorem of strong conditions, that generally verify the existence of the log-moments and log-cumulants for the distribution usually applied in signal and image processing.

TABLE 1
Properties of the Mellin convolution, the inverse Mellin convolution, and the Mellin correlation of two distributions defined on $\mathbb{R}^+$: p_A and p_B, with second kind characteristic functions ϕ_A and ϕ_B, and log-cumulants $\tilde{\kappa}_{A,n}$ and $\tilde{\kappa}_{B,n}$.

	Characteristic function	Cumulant
$p_A \hat{\star} p_B$	$\phi_A(s)\,\phi_B(s)$	$\tilde{\kappa}_{A,n} + \tilde{\kappa}_{B,n}$
$p_A \hat{\star}^{-1} p_B$	$\frac{\phi_A(s)}{\phi_B(s)}$	$\tilde{\kappa}_{A,n} - \tilde{\kappa}_{B,n}$
$p_A \hat{\otimes} p_B$	$\phi_A(s)\,\phi_B(2-s)$	$\tilde{\kappa}_{A,n} + (-1)^n \tilde{\kappa}_{B,n}$

Let $p(u)$ be a probability distribution defined on $\mathbb{R}^+$, whose second kind characteristic function is $\phi(s)$. This pdf satisfies the relations:

- $p(u) \geq 0 \quad \forall u \geq 0$
- $\int_0^{+\infty} p(u)\,du = \phi(s)|_{s=1}$

Theorem 1: If a pdf has a second kind characteristic function defined on the set $\Omega =]s_A, s_B[$, where $s = 1 \in \Omega$, then all of its log-moments and log-cumulants exist.

Proof: The existence of the log-moments and log-cumulants depends by the convergence of the integral

$$\int_0^{+\infty} (\log u)^n p(u)\,du\,.$$

In order to study this improper integral, we will study its behaviour at 0 and at the limit to infinity.

- **close to infinity:** Let $\alpha \in \Omega$ such that $\alpha > 1$. Thus, $\exists\, \alpha > 1$ such that:

$$\phi(\alpha) = \int_0^{+\infty} u^{\alpha-1} p(u)\,du < \infty$$

which amounts to saying that the moments of $p(u)$ (integer order or fractional) can be calculated for all orders between 1 and α. Assume an integer $n \geq 1$. Two cases then arise:

  - For $\forall\ x > 1$ we have $(\log x)^n < x^{\alpha-1}$. In this case, knowing that $p(u)$ is a pdf and satisfies $p(u) \geq 0$, one can write

$$\lim_{b\to\infty} \int_1^b (\log u)^n p(u)\,du \leq \lim_{b\to\infty} \int_1^b u^{\alpha-1} p(u)\,du \leq \phi(\alpha)$$

which demonstrates the convergence of the integral as $x \to \infty$.
  - There exists a constant $c > 1$ such that $(\log c)^n = c^{\alpha-1}$. Then, for $\forall\ x > c$ we have $(\log x)^n \leq x^{\alpha-1}$. By an identical argument as for the previous case, we deduce that

$$\lim_{b\to\infty} \int_c^b (\log u)^n p(u)\,du \leq \phi(\alpha)$$

which demonstrates the convergence of the integral as $x \to \infty$.

- **close to 0:**
  - First of all, consider the particular case where the pdf is bounded. Assume that

$$\exists\, A \quad : \quad \forall u \in [0,1]\,, \quad p(u) \leq A\,,$$

and calculate the limit

$$\lim_{a\to 0} \int_a^1 (\log u)^n p(u)\,du\,.$$

Since the pdf is bounded, we have for $\forall a \in]0,1[$ that

$$\left|\int_a^1 (\log u)^n p(u)\,du\right| \leq \left|\int_a^1 (\log u)^n A\,du\right| \leq A\left|\int_a^1 (\log u)^n\,du\right|\,.$$

The following property

$$\left|\lim_{a\to 0} \int_a^1 (\log u)^n\,du\right| = \Gamma(n+1)$$

proves the convergence at 0.
  - In the general case, the study of the convergence starts from the variables change $x \to \frac{1}{x}$, thereafter utilising the convergence property that we have just shown for the case $x \to \infty$.

We deduce that if a probability distribution with bounded values possesses moments (fractional or integer ordered) of order strictly larger than 0 and strictly smaller than 0, then all its log-moments and log-cumulants exist. □

Note that a far more elegant and concise proof, founded on the properties of analytical functions, can be worked out without major problems based on the assumption that $\phi(s)$ is holomorphic [7], and thus differentiable up to all orders at $s = 1$.

2.6 Comparison with Logarithmic Transformation

At this stage, one can wonder what the advantages of this new approach are, and whether a simple transformation into logarithmic scale would lead to the same result. We will show that in order to calculate a characteristic function after logarithmic transformation, one effectively has to calculate the Mellin transform of the original pdf.

We shall consider a random variable x with density defined over real positive numbers. Its pdf, $p_x(u)$, is thus defined for $u \in \mathbb{R}$, and the characteristic function is written:

$$\Phi_x(v) = \int_0^{+\infty} e^{jvu} p_x(u)\,du\,.$$

Then perform a logarithmic transformation. The new random variable y is described the pdf $q_y(w)$, defined for $w \in \mathbb{R}$, with $w = \log u$. This pdf results from p_x with the relation given by

$$q_y(w) = e^w p_x(e^w)\,.$$

Now calculate the characteristic function of the random variable y:

$$\begin{aligned}\Phi_y(v) &= \int_{-\infty}^{+\infty} e^{jvw} q_y(w)\, dw \\ &= \int_{-\infty}^{+\infty} e^{jvw} e^{w} p_x(e^w)\, dw \\ &= \int_{0}^{+\infty} e^{jv\log u} p_x(u)\, du \text{ with } u = e^w \\ &= \int_{0}^{+\infty} u^{jv} p_x(u)\, du\,.\end{aligned} \tag{24}$$

The relation in (24) is recognised as the Mellin transform of $p_x(u)$ at $s = 1 + jv$:

$$\Phi_y(v) = \phi_x(s)|_{s=1+jv}\,. \tag{25}$$

This relation shows that if one knows the Mellin transform of a pdf (i.e., its second kind characteristic function), then one knows the ordinary characteristic function in logarithmic scale.

On logarithmic scale, moments and cumulants are deduced by differentiation (simple or logarithmic) of expression (25), which is equivalent to what was defined in Eq. (10). This is another way to justify the terms *log-moments* and *log-cumulants*. We note, however, that the second kind statistics represent a generic method to find log-moments and log-cumulants directly without requiring a variable change (logarithmic transformation) and also without having to calculate the new distribution for the transformed variable.

Moreover, we will see that in the cases generally encountered in signal and image processing, and where the entities are defined on $\mathbb{R}^+$, it is easier to calculate the Mellin transform than the Fourier transform. Thus, our approach simplifies the working of the problem. In addition, when the Mellin transform is known, one automatically obtains:

- the moments, by inserting positive integers for the Mellin transform variable s, and
- the log-moments, by differentiating the Mellin transform with respect to s and evaluating at $s = 1$.

This should be appreciated by any practitioner, since, by applying a single transformation to the distribution, both moments and log-moments are obtained.

2.7 Comparison between Integral Transforms

The use in this context of an ignored transform: the Mellin transform, may surprise, since there exist more common invertible transforms, such as the Laplace transform, that could potentially play an important role in the study of distributions defined on $\mathbb{R}^+$. At this stage, it is important to observe what the relations between the Fourier transform ($\mathcal{F}$), the Laplace transform ($\mathcal{L}$), and the Mellin transform ($\mathcal{M}$) are. When it exists, the Laplace transform of a pdf $p(u)$ is written:

$$\mathcal{L}[p(u)](\sigma) = \int_0^{\infty} e^{-\sigma u} p(u)\, du$$

while the first characteristic function of this pdf is written

$$\mathcal{F}[p(u)](v) = \int_{-\infty}^{+\infty} e^{jvu} p(u)\, du\,.$$

The following relation is immediately deduced:

$$\mathcal{F}[p(u)](v)|_{v=-\frac{b}{2\pi}} = \mathcal{L}[p(u)](\sigma)|_{\sigma=jb}$$

Because the Laplace transform variable is a complex entity, one may consider that the Laplace transform could allow for an analytical continuation of the characteristic function [8]. However, the intrinsic properties of the Laplace transform are the same as those of the Fourier transform. A logarithmic transformation (in which case, one will have to use the bilateral Laplace transform) will in reality turn the Laplace transform into a Mellin transform:

$$\mathcal{L}[\tilde{p}(u)]|_{\sigma=a+jb} = \mathcal{M}[p(u)]|_{s=-a-jb} = \phi_x(s)|_{s=-a-jb}$$

where $\tilde{p}$ is the pdf on logarithmic scale.

There are such strong relations between these transforms that, most likely, nothing fundamentally new will be found by the use of the Mellin transform. Therefore, it seems that the choice should be dictated by practical considerations. We have seen that the Mellin transform makes it possible to obtain traditional moments and log-moments at the same time, without having to derive the pdf on logarithmic scale. Moreover, the Mellin transform of the experimental distributions commonly used in signal and image processing can be found in tables. This justifies a further look into the use of this rather ignored transform. This is the pragmatic view which motivates the derivations of this article.

3 Fundamental Examples

We will illustrate the new approach by applying it to distributions used to model synthetic aperture radar (SAR) images. These are the gamma and the generalised gamma distribution (intensity images with fully developed speckle), the Rayleigh and the Nakagami distribution (amplitude images with fully developed speckle), and finally the $\mathcal{K}$ distribution (an intensity distribution modelling fully developed speckle modulated multiplicatively by gamma distributed texture). Even if some of the results obtained are trivial, it seems important to be able to carry out comparisons with these simple and well-known cases, in particular in order to handle the problem of estimating the distribution parameters, an aspect which will be looked at in Section 5.

3.1 Gamma and Generalised Gamma Distribution

The two parameter gamma distribution, denoted $\mathcal{G}[\mu, L]$, is a type III solution of the Pearson system [9]. It is defined on $\mathbb{R}^+$ as

$$\mathcal{G}[\mu, L](u) = \frac{1}{\Gamma(L)} \frac{L}{\mu} \left(\frac{Lu}{\mu}\right)^{L-1} e^{-\frac{Lu}{\mu}} \quad \text{with } L > 0 \tag{26}$$

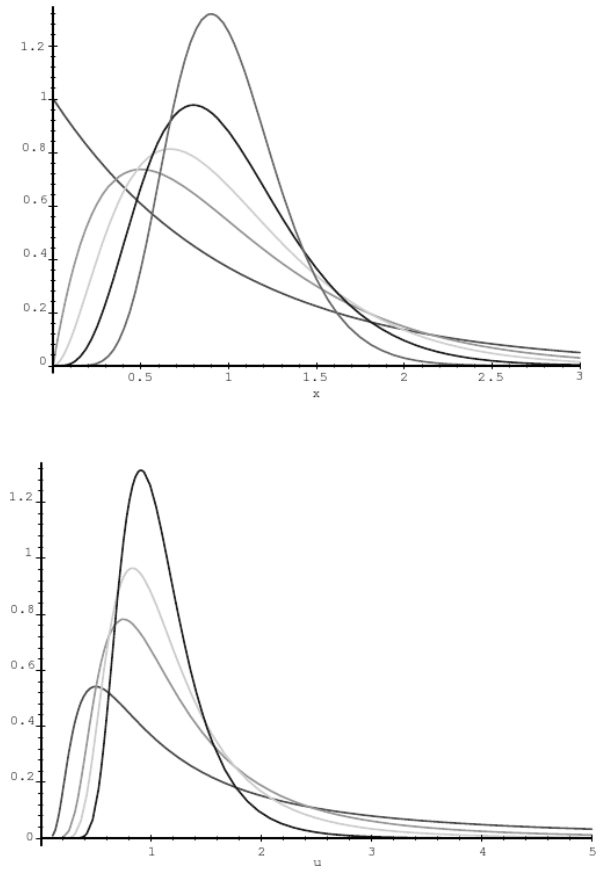

Fig. 1. Top: Gamma distribution $\mathcal{G}[\mu, L]$ with $\mu = 1$ and $L = 1, 2, 3, 5$ and 10. Bottom: Inverse gamma distribution $\mathcal{IG}[\mu, L]$ with $\mu = 1$ and $L = 1, 2, 3, 5$ and 10.

We see that μ is a scale parameter and that L is a shape parameter (Figure 1).

The particular case of $L = 1$ corresponds to the true gamma distribution[2], which is well-known from the radar literature as a model of fully developed speckle in single-look images. The case of $L = \frac{1}{2}$ gives the χ^2 distribution.

The Fourier transform tables show that the characteristic function is written as:

$$\Phi(\nu) = \left(\frac{L}{\mu}\right)^L \frac{e^{jL\arctan\left(\frac{\nu\mu}{L}\right)}}{\left(\nu^2 + \frac{L^2}{\mu^2}\right)^{\frac{L}{2}}} \tag{27}$$

whose complicated expression makes it difficult to use in practice.

On the other hand, by use of known Mellin transforms that can be found in tables [2], [10], the second kind characteristic function can be expressed in terms of the gamma function as:

$$\phi_x(s) = \mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)} \tag{28}$$

2. Translator's remark: Note that the author uses the terms 'generalised gamma distribution' and '(true) gamma distribution' for the distributions more commonly referred to as the 'gamma distribution' and the 'exponential distribution', respectively.

The classical moments $m_n, \forall n \in \mathbb{N}$ are much easier to derive from this function than from (27):

$$m_n = \mu_n \frac{\Gamma(L+n)}{L^n\Gamma(L)} \tag{29}$$

from which we have the well-known moments:

$$m_1 = \mu \qquad m_2 = \frac{L+1}{L}\mu^2\,.$$

This equation system is analytically invertible, and from the first two moments we derive the following relations for the parameters μ and L:

$$\mu = m_1 \tag{30}$$

$$L = \frac{m_1^2}{m_2 - m_1^2} = \frac{1}{\frac{m_2}{m_1^2} - 1} \tag{31}$$

Note that this distribution is asymmetric, and its mode value is given by:

$$m_{\text{mode}} = \frac{L-1}{L}\mu \leq \mu\,.$$

We also remark that the second kind characteristic function can be separated into a first term, μ^{s-1}, and a second term that only depends on L, the shape parameter. As L goes to infinity, $\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}$ goes towards 1, and $\mathcal{G}[\mu, L]$ converges in distribution to the homothetic distribution $\mathcal{H}[\mu]$:

$$\phi_x(s) \to \mu^{s-1}$$
$$\Leftrightarrow \mathcal{G}[\mu, L](u) \to \mathcal{H}[\mu](u) = \frac{1}{\mu}\delta(\mu u - 1)$$

We note that the homothetic distribution can be seen as a degenerate Gaussian distribution (i.e. with zero variance). It seems to confirm what many experts of radar imaging has pointed out, that the gamma distribution tends towards a Gaussian distribution as L goes to infinity, but by the alternative denotation we avoid abuse of language that can lead to confusion.

Another major point specific to the distributions defined on $\mathbb{R}^+$ rests on the fact that the Mellin transform of $\mathcal{G}[\mu, L]$ is defined for $s > 1 - L$. It is seen that, for $L > 1$, it is possible to have negative values of $s - 1$ and thus *lower order moments*. Qualitatively, the lower order moments – i.e. positive powers of $\frac{1}{u}$ – mainly reflect the weight of the distribution between 0 and μ, while the traditional moments - i.e. positive powers of u - rather analyse the distribution between μ and ∞. Thanks to the lower order moments, it is possible to analyse selectively the left or the right tail of a probability distribution. The importance of this observation for asymmetrical distributions such as the gamma distribution is evident.

The first two log-cumulants of $\mathcal{G}[\mu, L]$ are expressed by the following relations, where $\Phi(\cdot)$ is the digamma function and $\Psi(r, \cdot)$ is the r-th order polygamma function, i.e. the r-th order derivative of the digamma function:

$$\tilde{\kappa}_{x(1)} = \log(\mu) + \Psi(L) - \log(L) \tag{32}$$

$$\tilde{\kappa}_{x(2)} = \Psi(1, L) \geq 0 \tag{33}$$

$$\tilde{\kappa}_{x(3)} = \Psi(2, L) \leq 0 \tag{34}$$

and it is trivial to show that

$$\tilde{\kappa}_{x(r)} = \Psi(r-1, L) \quad \forall\, r > 1\,,$$

which expresses the property that the log-cumulants depend only on L from second order and upwards.

We note that the property

$$\lim_{L\to\infty} (\Psi(L) - \log(L)) = 0$$

associated with the fact that the polygamma functions go towards 0 at infinity, can be used to show that the gamma distribution converges towards the homothetic distribution as L goes to infinity.

Remark that the third order log-cumulant is negative. This illustrates that, for the gamma distribution, the left tail is heavier than the right tail of the distribution, which decreases very quickly as the argument approaches infinity.

The log-mean is written:

$$\tilde{\bar{m}} = \mu \frac{e^{\Psi(L)}}{L} \tag{35}$$

It is interesting to note the two following points:

- $\frac{e^{\Psi(L)}}{L} \leq 1$: The log-mean is less than the mean value. Note that this property is valid for all L.
- $\frac{e^{\Psi(L)}}{L} \geq \frac{L-1}{L}$: The log-mean is larger than the mode value.

A more complete analysis would show that the log-mean is also lower than the median value, defined by

$$\int_0^{m_{med}} \mathcal{G}(u)\, du = 0.5\,.$$

It can also be justified to use the log-mean instead of the traditional mean. This gives interesting results in certain applications of image processing [11].

Finally, by a logarithmic transformation, the gamma distribution $\mathcal{G}[\mu, L](u)$ becomes the Fisher-Tippett distribution $\mathcal{G}_{\mathcal{FT}}[\tilde{\mu}, L](w)$ with $\tilde{\mu} = \log \mu$ and $w = \log u$:

$$\mathcal{G}_{\mathcal{FT}}[\bar{\mu}, L](w) = \frac{L^L}{\Gamma(L)} e^{L(w-\bar{\mu})} e^{-Le^{(w-\bar{\mu})}}$$

Its characteristic function is obtained by taking the Fourier transform. Unfortunately, the required relation is not found in tables. This is commonly circumvented by showing that the evaluation amounts to calculating a Mellin transform. In effect, one applies (25) unknowingly.

To conclude, it is seen that in the case of the generalised gamma distribution, the second kind statistics approach allows us:

- to obtain a simpler expression for the second kind characteristic function than for the classical characteristic function.
- to estimate the distribution parameters more efficiently by inversion of Eqs. (32) and (33).
  - The shape parameter L is easily derived from the second order log-cumulant, even if no analytical formulation can be found, since the polygamma functions are monotonous and easy to invert numerically (Tabulation can also be used to save computation time). The variance of this estimator is evaluated in Section 5, and we will see that it is notably lower than the variance obtained with the method of moments estimator, as defined in (31).
  - After L is known, μ can be derived from the expression of the first order log-cumulant.
- to propose a "typical value" for use in image processing, lying between the mode and the mean, which realistically represents a sample if it can be regarded as homogeneous.

3.2 Rayleigh and Nakagami Distribution

We will now handle a problem specific to SAR imagery, namely the transformation of intensity data to amplitude data. Even if models have simple expressions for intensity data (the gamma distribution is known to all scientific communities), the images are quite often available as amplitude data, which will reveal new problems regarding parameter estimation. In this article, we will thus address the transformation from intensity data that follow the gamma distribution, to amplitude data with their resulting distribution.

The Nakagami distribution[3] is the name which in the radar literature has been associated with amplitude data that follow a gamma distribution when transformed into the intensity domain. It is thus a two parameter distribution: $\mathcal{RN}[\mu, L]$, given by:

$$\mathcal{RN}[\mu, L](u) = \frac{2}{\mu} \frac{\sqrt{L}}{\Gamma(L)} \left(\frac{\sqrt{L}u}{\mu}\right)^{2L-1} e^{-\left(\frac{\sqrt{L}u}{\mu}\right)^2}. \tag{36}$$

For $L = 1$, one retrieves the Rayleigh distribution:

$$\mathcal{RN}[\mu, L{=}1](u) = \frac{2}{\mu} \left(\frac{u}{\mu}\right) e^{-\left(\frac{u}{\mu}\right)^2}.$$

The fundamental relation between the Nakagami distribution (for amplitude) and the generalised gamma distribution (for intensity, i.e. squared amplitude) is obtained by a simple variable change, which can be written as:

$$\mathcal{RN}[\mu, L](u) = 2u\mathcal{G}[\mu_{\mathcal{G}}, L](u^2)\,. \tag{37}$$

By means of the following two Mellin transform properties [2]:

$$\mathcal{M}[u^a f(u)](s) = \mathcal{M}[f(u)](s+a)$$
$$\mathcal{M}[f(u^a)](s) = \frac{1}{a}\mathcal{M}[f(u)]\left(\frac{s}{a}\right)$$

3. It is important to return to Nakagami the paternity of this distribution described by two parameters: mean and shape parameter, which has often been wrongly associated with the generalised gamma distribution. The formalism was proposed in 1942 by Nakagami in an exhaustive study of the "m-distribution". It was not published in English until 1960 [12]. Of course, this distribution is for instance found in [13] as the result of transformations starting from the gamma distributions. However, it seems that Nakagami performed the first complete study.

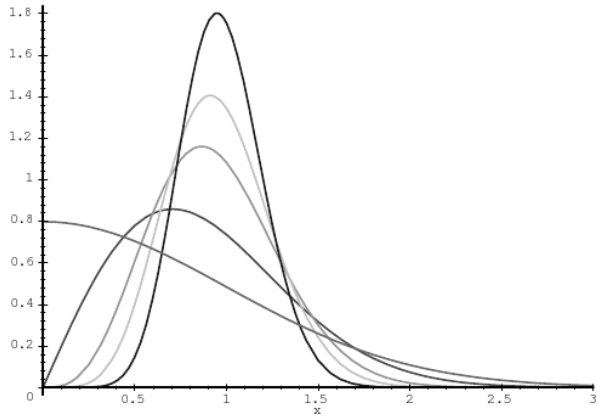

Fig. 2. Rayleigh-Nakagami distribution $\mathcal{RN}[\mu, L]$ with $\mu = 1$ and $L = 0.5, 1, 2, 3, 5$.

and knowing $\phi_{\mathcal{G},x}(s)$, the second kind characteristic function of the gamma distribution, the second kind characteristic function of the Nakagami distribution can be derived directly as:

$$\phi_{\mathcal{RN},x}(s) = \phi_{\mathcal{G},x}\left(\frac{s+1}{2}\right)$$

which, by inserting $\mu_{\mathcal{G}} = \mu^2$, allows us to write:

$$\phi_{\mathcal{RN},x}(s) = \mu^{s-1}\frac{\Gamma(\frac{s-1}{2}+L)}{L^{\frac{s-1}{2}}\Gamma(L)}\,.$$

This reasoning applies also elsewhere, regardless of the power to which u is raised in the change of variable. It is easily shown that for $v = u^\alpha$, we have

$$p_u(u) = \alpha u^{\alpha-1} p_v(u^\alpha)$$

and the second kind characteristic function of the random variable u is derived directly from the properties of the Mellin transform as:

$$\phi_u(s) = \phi_v\left(\frac{s+\alpha-1}{\alpha}\right)\,. \tag{38}$$

Note that this result would be useful for the analysis of the Weibull distribution [14], another well-known radar distribution, which we will not address in this article.

The classical moments of the Nakagami distribution follow directly from $\phi_{\mathcal{RN},x}(s)$:

$$m_1 = \frac{\Gamma\left(L+\frac{1}{2}\right)}{\sqrt{L}\Gamma(L)}\mu \qquad m_2 = \mu^2\,.$$

Take note of a peculiarity of this distribution: There is a very simple relation between the parameter μ and the second order moment, not the first order moment. On the other hand, the implicit expression of L obtained through the first order moment is very hard to handle. We cannot obtain a simple inversion formula (as in the gamma distribution case, where (31) gave L directly in terms of m_1 and m_2) to solve for L.

The mode of this pdf is

$$m_{\text{mode}} = \sqrt{\frac{2L-1}{2L}}\mu$$

The log-cumulants are derived directly from those of the gamma distribution as:

$$\begin{aligned}
\tilde{\kappa}_{\mathcal{RN},x(r)} &= \left.\frac{d^r \psi_{\mathcal{RN}}(s)}{ds^r}\right|_{s=1} \\
&= \left.\frac{d^r \log \phi_{\mathcal{RN}}(s)}{ds^r}\right|_{s=1} \\
&= \left.\frac{d^r \log \phi_{\mathcal{G}}\left(\frac{s+1}{2}\right)}{ds^r}\right|_{s=1} \\
&= \left(\frac{1}{2}\right)^r \left.\frac{d^r \log \phi_{\mathcal{G}}(s')}{ds'^r}\right|_{s'=1} \\
&= \left(\frac{1}{2}\right)^r \tilde{\kappa}_{\mathcal{G},x(r)}
\end{aligned}$$

From this we deduce:

$$\begin{aligned}
\tilde{\kappa}_{x(1)} &= \log(\mu) + \frac{1}{2}\Psi(L) - \frac{1}{2}\log(L) \\
\tilde{\kappa}_{x(2)} &= \frac{1}{4}\Psi(1, L)
\end{aligned}$$

and for all $r > 1$:

$$\tilde{\kappa}_{x(r)} = \left(\frac{1}{2}\right)^r \Psi(r-1, L)$$

More generally, it is shown for $v = u^\alpha$ that

$$\tilde{\kappa}_{p_u,x(r)} = \left(\frac{1}{\alpha}\right)^r \tilde{\kappa}_{p_v,x(r)}$$

In this case, L can be calculated directly if the second order log-cumulant is known. The problem we meet is of the same kind as for the gamma distribution, namely inversion of polygamma functions.

The log-mean is written

$$\tilde{\tilde{m}} = \mu\frac{e^{\frac{\Psi(L)}{2}}}{\sqrt{L}} \tag{39}$$

To conclude, the motivation of our approach is seen from the fact that the analytical expressions of the log-moments and the log-cumulants have a complexity comparable with the case of the gamma distribution, which is not the case in traditional statistics, where a simple relation between the first two moments and the shape parameter L cannot be obtained.

3.3 Inverse Gamma Distribution

The inverse gamma distribution is another two parameter distribution which is also a solution of the Pearson system (the type V solution). Is is expressed as

$$\mathcal{IG}[\nu, L](u) = \frac{1}{\Gamma(L)}\frac{1}{L\nu}\left(\frac{L\nu}{u}\right)^{L+1} e^{-\frac{L\nu}{u}}$$

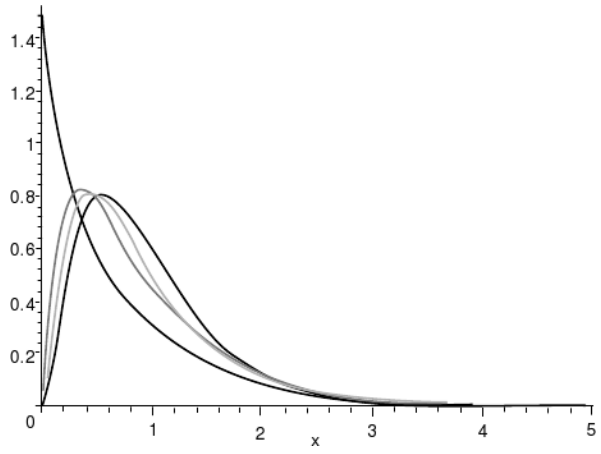

Fig. 3. $\mathcal{K}$ distribution: Fully developed speckle ($L = 3$) modulated multiplicatively by a Rayleigh distribution with $\mu = 1$ and $1, 3, 5$ and 10.

where $L \geq 0$ and $\nu > 0$. Its second kind characteristic function is written:

$$\phi_x(s) = \nu^{s-1} \frac{\Gamma(L+1-s)}{L^{1-s}\Gamma(L)} .$$

It is seen that the n-th order moments of the inverse gamma distribution are not defined for $n \geq L$. The inverse gamma distribution is thus an example of heavy tailed distributions. Its log-cumulants, that exist for all orders, are written:

$$\begin{aligned} \tilde{\kappa}_{x(1)} &= \log(\nu) - \Phi(L) + \log(L) \\ \tilde{\kappa}_{x(2)} &= \Psi(1, L) \\ \tilde{\kappa}_{x(r)} &= (-1)^r \Psi(r-1, L) \quad \forall r > 1 \end{aligned}$$

For even r, these are the same as those of the gamma distribution. For odd r, they are opposite (See the more general relation in (20)). As for the gamma distribution, from second order and upwards, the log-cumulants depend only on the shape parameter L. Note that the third order log-cumulant is positive, which is a sufficient condition for being heavy-tailed.

This distribution could also have been introduced as the inverse of the gamma distribution (cf. Section 2.4), which would make it possible to deduce the log-cumulants directly. However, it was important to recall that the inverse gamma distribution is a particular solution of the Pearson system and associated with its own share of work in the literature.

3.4 $\mathcal{K}$ Distribution

With the $\mathcal{K}$ distribution, we will show that second order statistics provide an estimation method for the parameters of a complex distribution by simple application of the results already achieved for the gamma distribution. The $\mathcal{K}$ distribution $\mathcal{K}[\mu, L, M]$ has three parameters and is defined as

$$\begin{aligned} \mathcal{K}[\mu, L, M](u) = \frac{1}{\Gamma(L)\Gamma(M)} \frac{2LM}{\mu} \left(\frac{LMu}{\mu}\right)^{\frac{M+L}{2}-1} \\ \times K_{M-L}\left[2\left(\frac{LMu}{\mu}\right)^{1/2}\right] \end{aligned} \tag{40}$$

where $K_n(\cdot)$ is the modified Bessel function of the second kind with order n. On this form, calculations of moments and log-moments require good knowledge of Bessel function properties as well as tables of transforms of Bessel functions.

In fact, the $\mathcal{K}$ distribution is the distribution which is followed by a random variable defined as the product of two independent variables that are both gamma distributed. Note that this definition made it possible for Lomnicki [15] to retrieve Eq. (40) using, already at this time, the Mellin transform.

More precisely, it is possible to define the $\mathcal{K}[\mu, L, M]$ distribution as a Mellin convolution of two gamma distributions [6]:

$$\mathcal{K}[\mu, L, M] = \mathcal{G}[1, L] \hat{\star} \mathcal{G}[\mu, M]$$

This definition greatly simplifies the calculations of the second kind characteristic function and thus the moments and log-cumulants. In effect, from the properties of the Mellin convolution (Section 2.4) and knowing the characteristics of the gamma distribution, one can write the second kind characteristic function of the $\mathcal{K}$ distribution like a product of the second kind characteristic functions of the gamma distributions $\mathcal{G}[1, L]$ and $\mathcal{G}[\mu, M]$:

$$\phi_x(s) = \mu^{s-1} \frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)} \frac{\Gamma(M+s-1)}{M^{s-1}\Gamma(L)} \tag{41}$$

which allows us to immediately deduce the classical moments m_1 and m_2 without using the definition of the $\mathcal{K}$ distribution, and thus without needing to know the properties of the Bessel function:

$$m_1 = \mu \qquad m_2 = \mu^2 \frac{L+1}{L} \frac{M+1}{M} .$$

In the same manner, we obtain directly the first two log-cumulants as the sum of the log-cumulants of the gamma distributions $\mathcal{G}[1, L]$ and $\mathcal{G}[\mu, M]$:

$$\tilde{\kappa}_{x(1)} = \log \mu + (\Psi(L) - \log(L)) + (\Psi(M) - \log M) \tag{42}$$

$$\tilde{\kappa}_{x(2)} = \Psi(1, L) + \Psi(1, M) \tag{43}$$

$$\tilde{\kappa}_{x(3)} = \Psi(2, L) + \Psi(2, M) \tag{44}$$

and we can show that for all $r > 1$:

$$\tilde{\kappa}_{x(r)} = \Psi(r-1, L) + \Psi(r-1, M) .$$

Finally, the normalised second order moment is written:

$$\tilde{M}_2 = \Psi(1, L) + \Psi(1, M) \tag{45}$$

and the log-mean:

$$\tilde{\tilde{m}} = \mu \frac{e^{\Psi(L)}}{L} \frac{e^{\Psi(M)}}{M} . \tag{46}$$

Also here, it is easy to derive the shape parameters L and M by virtue of the second and third order log-cumulants (Eqs. (43) and (44)). A simple numerical method is proposed and tested in [14]. The scale parameter is derived from the first order log-cumulant (Eq. (42)).

This method is much easier than the traditional method of moments, which results in a third degree equation. Note also that maximum likelihood estimation cannot be applied for this distribution [16].

4 Applications

Second kind statistics prove easy to put into practice in the framework of fundamental probability distribution defined on $\mathbb{R}^+$. Except for the introduction of the gamma function and its logarithmic derivatives (the polygamma functions), the obtained expressions contain no difficult terms. On the contrary, they are simple and easy to comprehend.

We will now look at several innovative applications of this model:

- A new approach to analysis of the three parameter distributions used to model SAR imagery
- The case of additive mixtures of gamma distributions, for which the traditional approaches lead to expressions that are very hard to handle
- The case of positive α-stable distributions, used for instance by Pierce to characterise sea clutter [17], for which it is difficult to estimate the parameters. (The analytical expression of the pdf for such heavy-tailed distributions is generally not known.)
- Finally, another example of the α-stable distribution proposed by Kuruoğlu and Zerubia [18], which can be seen as a generalisation of the Rayleigh distribution.

We will start by pointing out a method classically used to characterise these distributions: the use of the parameters of asymmetry β_1 (skewness) and peakedness β_2 (kurtosis).

4.1 (β_1, β_2) Diagram

Traditionally, the skewness and kurtosis are used to characterise distributions belonging to the Pearson system. These two coefficients are written in terms of the second, third and fourth order moment:

$$\beta_1 = \frac{M_3^2}{M_2^3}$$
$$\beta_2 = \frac{M_4}{M_2^2}$$

The curves obtained for the Pearson system are shown in Figure 4 in their classical representation. The characteristic point of $(\beta_1 = 0, \beta_2 = 3)$ corresponds to the Gaussian case (It is invariant with respect to variance).

Because of the choice of squaring the third order central moment in β_1, this coefficient is not able to distinguish between distributions that have skewness of the same magnitude but with opposite sign. Thus, it is not possible to separate between "standard" distributions and heavy-tailed distributions. Hence, this diagram seems to be badly adapted to the distributions defined on $\mathbb{R}^+$.

4.2 Characterisation of Texture Distributions in the $(\tilde{\kappa}_3\tilde{\kappa}_2)$ Diagram

The (β_1, β_2) diagram is founded on the calculation of traditional centred moments and aims at comparing distributions against the reference Gaussian distribution, for which the skewness is zero and the kurtosis is directly related to the variance (σ). It is then natural to propose a similar approach, founded on the functions of second kind statistics. We propose in this section a new method: the (κ_3, κ_2) diagram, that is, the representation of third order log-cumulants versus second order log-cumulants (that are always positive or zero for pdfs defined on $\mathbb{R}^+$).

In this diagram, the origin corresponds to the homothetic distribution. Because of the asymptotic behaviour of the gamma distribution and the inverse gamma distribution at $L \to \infty$, these distributions are represented by curves joining at the origin. As noted, the gamma

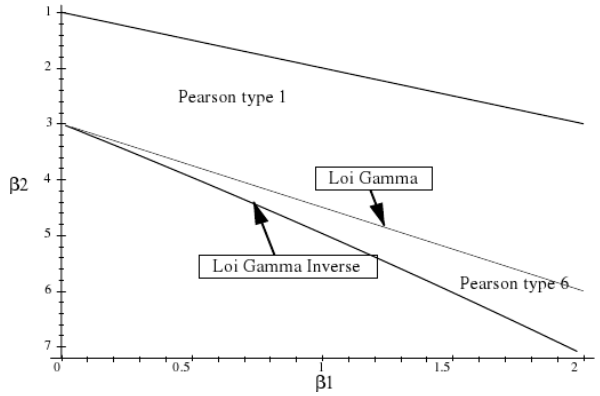

Fig. 4. The Pearson system displayed in a (β_1, β_2) diagram.

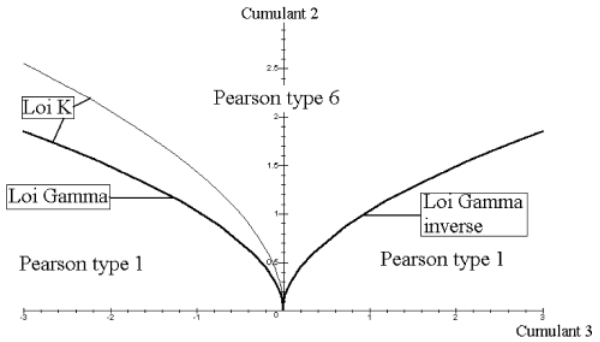

Fig. 5. The Pearson system and the $\mathcal{K}$ distribution displayed in a (κ_3, κ_2) diagram.

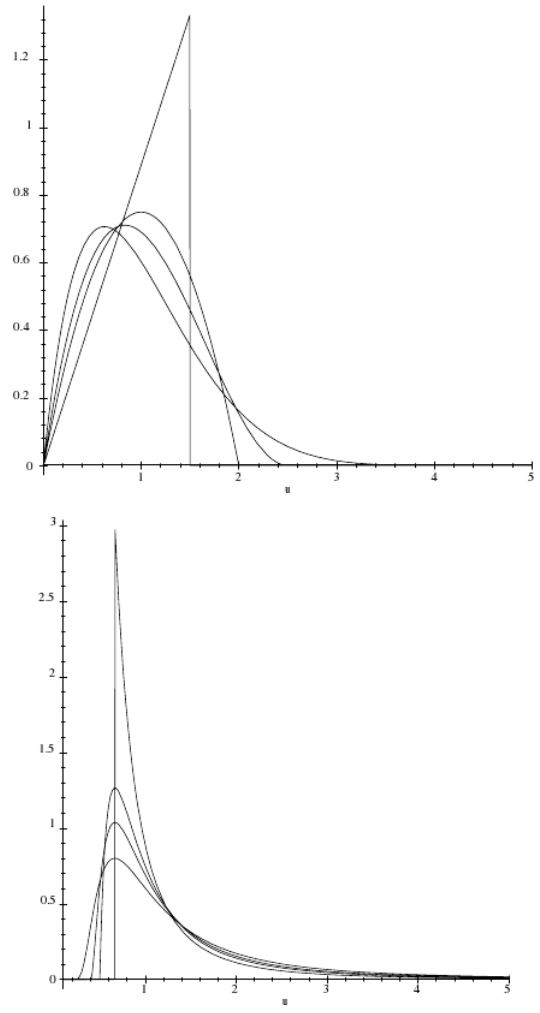

Fig. 6. The Pearson type I distribution (top) and the inverse Pearson type I distribution (bottom).

distribution has negative values for the third-order log-cumulants, while the heavy-tailed inverse gamma distribution has positive values. Note that it is easy to show that the log-normal distribution, for which all log-cumulants of order $n > 2$ is zero, occupies the second axis.

Figure 5 places the gamma distribution and the inverse gamma distribution in the diagram, together with the $\mathcal{K}$ distribution (which occupies a surface above the gamma distribution, limited upwards by a curve defined by the distribution $\mathcal{K}[L, L]$) as well as the Pearson distributions of type I (standard and inverse) and type VI. We will see in the following section that the inverse Pearson distributions of type I find their natural place in this diagram, but offer some theoretical surprises.

4.3 An Original Approach to Characterisation of Three-Parameter Distributions Used for SAR Imagery

Knowing the two elementary two parameter distributions (the gamma distribution and the inverse gamma distribution), it falls natural to make use of these as basic generating functions to obtain a kind of grammar by using elementary operations like the Mellin convolution and the inverse Mellin convolution (One could also have used the Mellin correlation instead of the Mellin convolution while inverting one of the distributions). Assume that we have two distributions p_A and p_B with respective second kind characteristic functions ϕ_A and ϕ_B and log-cumulants $\tilde{\kappa}_A$ and $\tilde{\kappa}_B$. Applying a Mellin convolution or an inverse Mellin convolution will correspond to forming the product or ratio of their second kind characteristic functions, and the sum or difference of the log-cumulants, respectively (See Table 1).

The characteristic functions of the distributions obtained by direct or inverse convolution of the two normalised distributions: the gamma distribution ($\mathcal{G}[1, L]$) or the inverse gamma distribution ($\mathcal{G}[1, M]$), are included in Table 2. From these expressions, and by consulting tables of the Mellin transform (and also using the properties of the transform), it is possible to retrieve the analytical expressions of these distributions without further calculations [14]. Furthermore, while considering only the second and third order second kind cumulants, Table 3 summarises the result obtained by direct or inverse convolution of the two normalised distributions: the gamma distribution ($\mathcal{G}[1, L]$) or the inverse gamma distribution ($\mathcal{G}[1, M]$). Recall that the second and third order second kind cumulants only depend on the shape parameter.

The distributions traditionally used in processing of SAR data are found in this table. These are

- The $\mathcal{K}$ distribution
- The solutions of the Pearson system [6] corresponding to the distributions defined on $\mathbb{R}^+$, that is, in addition to the gamma and inverse gamma distribution, also the type I solutions (also known as the beta distribution) and the type VI solutions (known as the Fisher distribution).

Moreover, uncommon distributions are generated by this algebraic method. It provides:

- A new distribution which is effectively the inverse Pearson distribution of type I, denoted $\mathcal{IPI}[\xi, L, M]$:

$$\mathcal{IPI}[\xi, L, M] = \frac{M}{L\xi}\frac{\Gamma(M)}{\Gamma(L)\Gamma(M-L)}\left(\frac{L\xi}{Mu}\right)^{L+1}\left(1-\frac{L\xi}{Mu}\right)^{M-L-1} \tag{47}$$

  with $u \geq \frac{L\xi}{M}$ and $M \geq L+1$.

This expression is derived simply by means of the Mellin transform tables [2], since the distribution can be expressed by the relation:

$$\mathcal{IPI}[\xi, L, M] = \mathcal{IG}[\xi, L] \,\hat{\star}^{-1}\, \mathcal{IG}[1, M]$$

whose characteristic function is

$$\xi^{s-1}\frac{\Gamma(L+1-s)}{L^{1-s}\Gamma(L)}\frac{M^{1-s}\Gamma(M)}{\Gamma(M+1-s)}$$

TABLE 2
Second kind characteristic functions of the different distributions obtained by Mellin convolution (direct and inverse) of the gamma distribution $\left(\phi(s)=\mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}\right)$ and inverse gamma distribution $\left(\phi(s)=\mu^{s-1}\frac{\Gamma(L+1-s)}{L^{1-s}\Gamma(L)}\right)$. The distributions whose names are typeset with boldface correspond to new analytical expressions. The second and third order log-cumulants are included in Table 3.

$\hat{\star}\nearrow$	$\mathcal{G}[1,M]$	$\mathcal{IG}[1,M]$
$\mathcal{G}[\mu,L]$	$\mathcal{K}$ distribution $\mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}\frac{\Gamma(M+s-1)}{M^{s-1}\Gamma(M)}$	Pearson VI $\mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}\frac{\Gamma(M+1-s)}{M^{1-s}\Gamma(M)}$
$\mathcal{IG}[\mu,L]$	Pearson VI $\mu^{s-1}\frac{\Gamma(L+1-s)}{L^{1-s}\Gamma(L)}\frac{\Gamma(M+s-1)}{M^{s-1}\Gamma(M)}$	**$\mathcal{IK}$ distribution** $\mu^{s-1}\frac{\Gamma(L+1-s)}{L^{1-s}\Gamma(L)}\frac{\Gamma(M+1-s)}{M^{1-s}\Gamma(M)}$

$\hat{\star}^{-1}\nearrow$	$\mathcal{G}[1,M]$	$\mathcal{IG}[1,M]$
$\mathcal{G}[\mu,L]$	Pearson I $\mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}\frac{M^{s-1}\Gamma(M)}{\Gamma(M+s-1)}$	**Bessel** $\mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}\frac{M^{1-s}\Gamma(M)}{\Gamma(M+1-s)}$
$\mathcal{IG}[\mu,L]$	**Bessel** $\mu^{s-1}\frac{\Gamma(L+1-s)}{L^{1-s}\Gamma(L)}\frac{M^{s-1}\Gamma(M)}{\Gamma(M+s-1)}$	**Inverse Pearson I** $\mu^{s-1}\frac{\Gamma(L+1-s)}{L^{1-s}\Gamma(L)}\frac{M^{1-s}\Gamma(M)}{\Gamma(M+1-s)}$

TABLE 3
Second and third order log-cumulants of the different distributions obtained by Mellin convolution (direct and inverse) of the gamma distribution and the inverse gamma distribution (cf. Table 2). To simplify the presentation, only the second and third order log-cumulants are included in the table. The distributions whose names are typeset in boldface correspond to new analytical expressions.

$\hat{\star}\nearrow$	$\mathcal{G}[1,M]$	$\mathcal{IG}[1,M]$	$\hat{\star}^{-1}\nearrow$	$\mathcal{G}[1,M]$	$\mathcal{IG}[1,M]$
$\mathcal{G}[1,L]$	$\mathcal{K}$ distribution	Pearson VI	$\mathcal{G}[1,L]$	Pearson I	**Bessel**
	$\Psi(1,L)+\Psi(1,M)$	$\Psi(1,L)+\Psi(1,M)$		$\Psi(1,L)-\Psi(1,M)$	$\Psi(1,L)-\Psi(1,M)$
	$\Psi(2,L)+\Psi(2,M)$	$\Psi(2,L)-\Psi(2,M)$		$\Psi(2,L)-\Psi(2,M)$	$\Psi(2,L)+\Psi(2,M)$
$\mathcal{IG}[\mu,L]$	Pearson VI	**$\mathcal{IK}$ distribution**	$\mathcal{IG}[\mu,L]$	**Bessel**	**Inverse Pearson I**
	$\Psi(1,L)+\Psi(1,M)$	$\Psi(1,L)+\Psi(1,M)$		$\Psi(1,L)-\Psi(1,M)$	$\Psi(1,L)-\Psi(1,M)$
	$-\Psi(2,L)+\Psi(2,M)$	$-\Psi(2,L)-\Psi(2,M)$		$-\Psi(2,L)-\Psi(2,M)$	$-\Psi(2,L)+\Psi(2,M)$

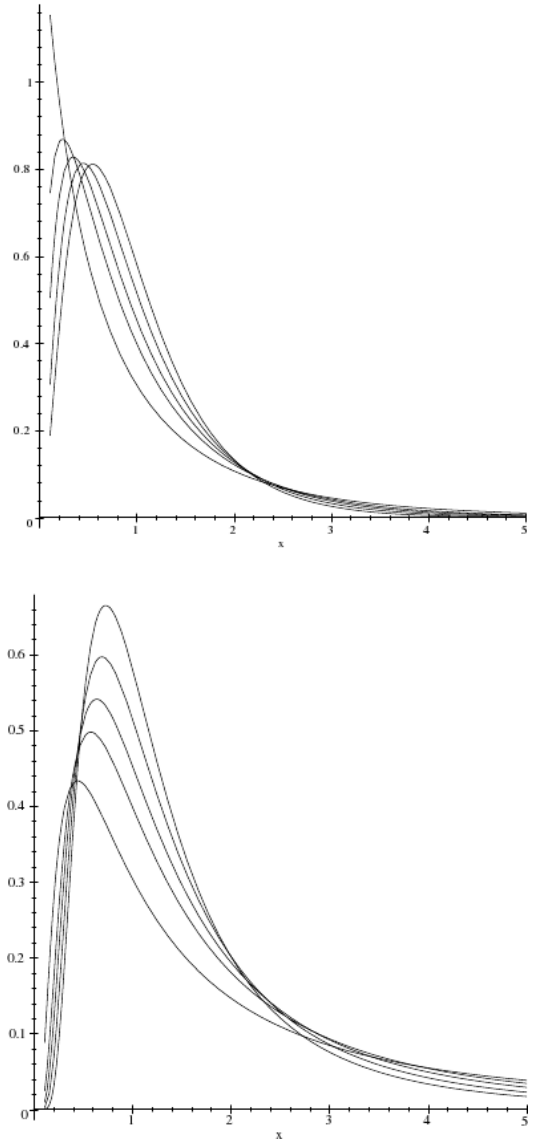

Fig. 7. The $\mathcal{K}$ distribution (top) and the inverse $\mathcal{K}$ distribution (bottom).

where the last expression is found in the Mellin transform tables. Figure 6 allows a comparison between the $\mathcal{IP}$ distribution and the $\mathcal{IP}I$ distribution for the same set of parameters. Recall that the Pearson type I distribution is expressed as:

$$\mathcal{IP}[\xi, L, M] = \frac{L}{M\xi}\frac{\Gamma(M)}{\Gamma(L)\Gamma(M-L)}\left(\frac{Lu}{M\xi}\right)^{L-1}\left(1-\frac{Lu}{M\xi}\right)^{M-L-1}$$

with $u \leq \frac{M}{L\xi}$ and $M \geq L+1$.

Curiously, this distribution is never mentioned in the classical works [9], [19], whereas they characterise the gamma distribution and the inverse gamma distribution separately. Moreover, the inverse Pearson type I distribution has the peculiar property of being localised, in the (β_1, β_2) diagram, between the gamma distribution and the inverse gamma distribution, that is, exactly where the Pearson type VI solution is found. Indeed, the case $M \to \infty$ corresponds to to the inverse gamma distribution and the zone corresponding to the inverse Pearson type I distribution cannot have ambiguities, whereas the (κ_3, κ_2) diagram separate well between the solutions of the Pearson system.

- The inverse $\mathcal{K}$ distribution, which is also uncommon, is expressed as:

$$\mathcal{IK}[\mu, L, M](u) = \frac{1}{L\Gamma(L)M\Gamma(M)}\frac{2}{\mu}\left(\frac{LM\mu}{u}\right)^{\frac{M+L}{2}+1} \times K_{M-L}\left[2\left(\frac{LM\mu}{u}\right)^{\frac{1}{2}}\right] \quad (48)$$

Figure 7 allows a comparison between the $\mathcal{K}$ distribution and the inverse $\mathcal{K}$ distribution for the same set of parameters. As for the $\mathcal{K}$ distribution, the modelling through the Mellin convolution makes it easy to show that the inverse $\mathcal{K}$ distribution tends to an inverse gamma distribution as one of the shape parameters (L or M) goes to to infinity. One thus has a three parameter distribution which is heavy-tailed.

- The combinations $\mathcal{G}\hat{\star}^{-1}\mathcal{IG}$ and $\mathcal{IG}\hat{\star}^{-1}\mathcal{G}$ have known analytical solutions that include Bessel functions. However, they are not probability distributions, as the condition $p_x(u) \geq 0$ is not satisfied.

We see that insightful interpretations can be made based on the second and third order log-cumulants, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$. By the simple analysis of these entities, we can effectively get an idea about the flexibility of a certain distribution compared to the generalised gamma distribution, its inverse, and the other distributions that cover the log-cumulant space: the $\mathcal{K}$ distribution and its inverse, and the solutions of the Pearson system. The analysis of the second and third order log-cumulants can also account for more complex models. We shall see that the same diagram can be used to analyse an additive mixture of gamma distributions, and propose an original and simple practical method to determine the model parameters.

4.4 Additive Mixture of Gamma Distributions

Additive mixtures of gamma distributions are important practical modelling tools (in particular for SAR imagery). Contrarily to the Gaussian case, a unimodal pdf is generally obtained, except when the (two) initial distributions are very different (see Figure 8). However, we will show that there exists a simple solution to determine the parameters of the mixture by analysing this problem aided by second and third order log-cumulants.

Consider the following mixture of gamma distributions:

$$\lambda\mathcal{G}[\mu, L] + \lambda'\mathcal{G}[\mu', L]$$

with $\lambda \geq 0$, $\lambda' \geq 0$ and $\lambda + \lambda' = 1$. In this model, L has the same value for the two gamma distributions.

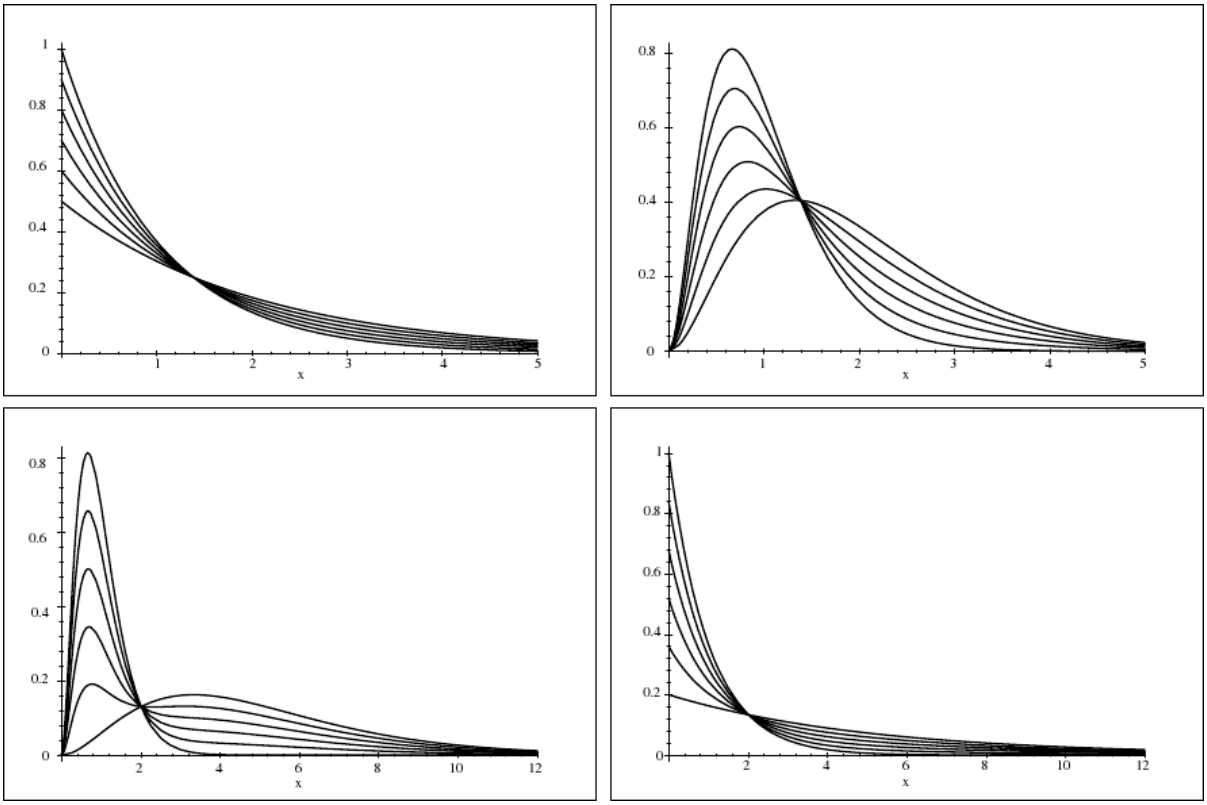

Fig. 8. Examples of additive mixtures of gamma distributions. The left column shows distributions with $L = 1$, and the right column with $L = 3$. In the first row, $\rho = 2$ ($\mu = 1$ and $\mu' = 2$). In the second row, $\rho = 5$ ($\mu = 1$ and $\mu' = 5$). λ takes the values $0, 0.2, 0.4, 0.6, 0.8$ and 1.

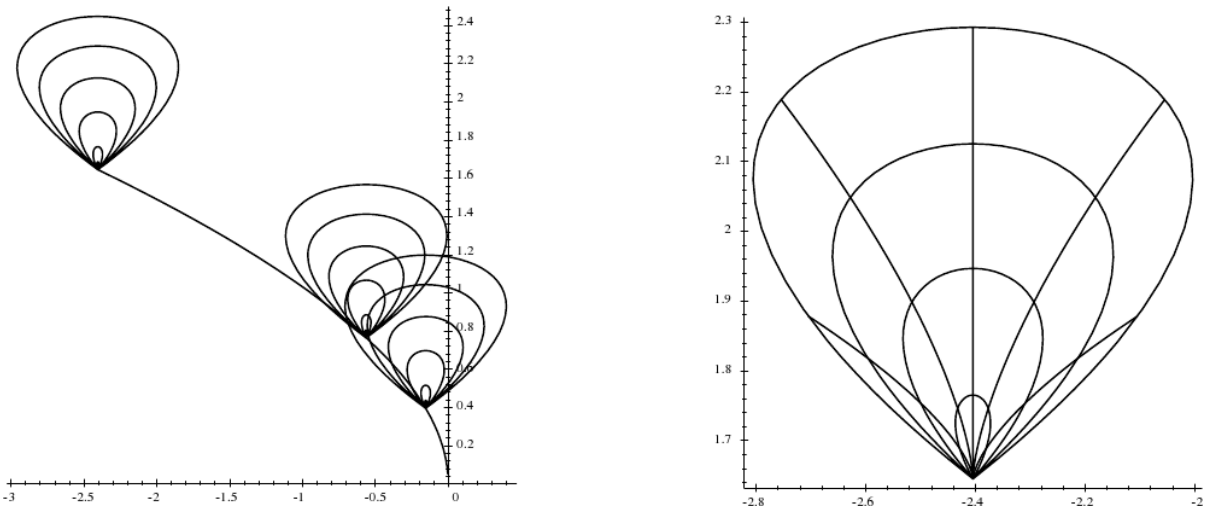

Fig. 9. (κ_2, κ_3) diagram for a mixture of gamma distributions described by the parameters λ (mixing proportion) and ρ (ratio of component means). To the left, for λ varied between 0 and 1, three diagrams are traced out corresponding to several values of ρ for three values of L ($L = 1, 3$ and 5). In the same plot, the gamma distribution is represented by a line which spans $L \in [1, \infty]$. To the right, for a fixed value of $L = 1$, one separately varies ρ between 0 and 5 (five curves, with λ taking the values $0.1, 0.3, 0.5, 0.7$ and 0.9) and λ between 0 and 1 (four closed curves, with ρ taking values $2, 3, 4$ and 5), placing both diagrams in the same figure.

The model can be rewritten by defining the variable ρ such that $\mu' = \rho\mu$, which makes it possible to write the mixture in the following form:

$$\lambda\mathcal{G}[\mu, L] + (1-\lambda)\mathcal{G}[\rho\mu, L] \tag{49}$$

The mixture is then defined by a gamma distribution $\mathcal{G}[\mu, L]$ (corresponding to only one of the mixture components) and two parameters describing the mixture: λ and ρ. The second kind characteristic function is written:

$$\phi(s) = (\lambda + (1-\lambda)\rho^{s-1})\mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}$$

Based on this expression, calculation of the log-cumulants can be carried out directly, giving the following expressions:

$$\begin{aligned}
\tilde{\kappa}_{x(1)} &= \Psi(L) - \log L + \log\mu + (1-\lambda)\log\rho \\
\tilde{\kappa}_{x(2)} &= \Psi(1, L) + \log(\rho)^2\lambda(1-\lambda) \\
\tilde{\kappa}_{x(3)} &= \Psi(2, L) + \log(\rho)^3\lambda(1-\lambda)(2\lambda-1)
\end{aligned}$$

We remark that starting from the second order, the log-cumulants do not depend on μ, and they have the same values as the standard gamma distribution for the limiting values $\lambda = 0$ and $\lambda = 1$.

We assume that the entity L is known (L can be perceived as a function of the instrument, thus it will be known by the processor). This leads to:

$$\begin{aligned}
\overline{\tilde{\kappa}_{x(2)}} &= \tilde{\kappa}_{x(2)} - \Psi(1, L) \\
\overline{\tilde{\kappa}_{x(3)}} &= \tilde{\kappa}_{x(3)} - \Psi(2, L)
\end{aligned}$$

Parameters λ and ρ are then given by the solutions of quadratic equation, which gives:

$$\begin{aligned}
\lambda &= \frac{1}{2}\left(1 \pm \frac{\overline{\tilde{\kappa}_{x(3)}}}{\sqrt{4\overline{\tilde{\kappa}_{x(2)}}^3 + \overline{\tilde{\kappa}_{x(3)}}^2}}\right) \\
\rho &= e^{\frac{\sqrt{4\overline{\tilde{\kappa}_{x(2)}}^3 + \overline{\tilde{\kappa}_{x(3)}}^2}}{\overline{\tilde{\kappa}_{x(2)}}}}
\end{aligned}$$

The evolution of the different parameters in the $(\tilde{\kappa}_3, \tilde{\kappa}_2)$ diagram is shown in Figure 9. It is interesting to notice that the shape of these curves does not depend on L.

If the obtained results are compared with those found in the literature, it is noticed that this approach relies on one assumption only: knowledge of the parameter L, while analyses of mixtures by classical methods require the additional knowledge of μ [20].

4.5 Positive α-Stable Distributions

We will now apply the methodology proposed in this article to the case of a 'heavy-tailed distribution, for which neither the analytical form of the pdf nor moments from a certain order and upwards are known. This prohibits the method of moments.

A positive α-stable distribution has a pdf characterised by two parameters: α and γ. It cannot in general be defined, other than by its characteristic function $\Phi(\nu)$, which is written (according to Pierce [17]) as:

$$\Phi(\nu) = e^{-\gamma|\nu|^\alpha\left(1 + j\,\mathrm{sgn}(\nu)\tan\left(\frac{\alpha\pi}{2}\right)\right)}$$

with

$$\mathrm{sgn}(\nu) = \begin{cases} 1, & \nu > 0 \\ 0, & \nu = 0, \\ -1, & \nu < 0 \end{cases} \qquad 0 < \alpha < 1, \quad \gamma > 0\,.$$

Except for certain particular values of α, the analytical expression of the distribution is not known.

One can nevertheless express the moments of this distribution (including fractional ones) as:

$$m_\nu = \frac{\gamma^{\frac{\nu}{\alpha}}\sin(\pi\nu)\Gamma(\nu+1)\left(1 + \left(\tan\left(\frac{\pi\alpha}{2}\right)\right)^2\right)^{\frac{\nu}{2\alpha}}}{\alpha\sin\left(\frac{\pi\nu}{\alpha}\right)\Gamma\left(1+\frac{\nu}{\alpha}\right)} \tag{50}$$

These moments are only defined for $\nu < \alpha < 1$, which means that even the first moment is not defined. This is evidently a heavy-tailed distribution.

It is nevertheless possible, by an analytical continuation, to derive the second order characteristic function, which is written:

$$\phi(s) = \frac{\gamma^{\frac{s-1}{\alpha}}\sin(\pi(s-1))\Gamma(s)\left(1 + \left(\tan\left(\frac{\pi\alpha}{2}\right)\right)^2\right)^{\frac{s-1}{2\alpha}}}{\alpha\sin\left(\frac{\pi(s-1)}{\alpha}\right)\Gamma\left(1+\frac{s-1}{\alpha}\right)}$$

It is seen that this function is well defined in a complex neighbourhood around the value $s = 1$. The existence theorem in Section 2.5 thus confirms that the log-moments and log-cumulants of all orders exist, whereas the moments of orders $\nu \geq \alpha$ are not defined.

Even though the analytical form is rather complicated, it is still possible to obtain simple expressions of the log-cumulants. Note that these expressions are only analytical continuations because the gamma functions in the derivatives of $\phi(s)$ have discontinuities at $s = 1$, a value at which they must be evaluated when calculating the log-cumulants. It is then necessary to study the limit at $s = 1$ in order to obtain the analytical expressions. The following results were established with assistance of mathematical computation software Maple, as a result of lengthy analytical developments:

$$\begin{aligned}
\tilde{\kappa}_1 &= \frac{(1-\alpha)\Psi(1)}{\alpha} + \frac{-\log\left(\cos\left(\frac{\pi\alpha}{2}\right)\right)}{\alpha} + \frac{\log\gamma}{\alpha} \\
\tilde{\kappa}_2 &= \frac{(1-\alpha^2)}{\alpha^2}\Psi(1,1) \\
\tilde{\kappa}_3 &= \frac{\alpha^3 - 1}{\alpha^3}\Psi(2,1)
\end{aligned} \tag{51}$$

These expressions, that eventually appear as rather simple, illustrate well the strength of our new approach. The two parameters of the distribution are easily derived since:

- The parameter α can be estimated from the second order log-cumulant as:

$$\alpha = \sqrt{\frac{\Psi(1,1)}{\Psi(1,1) + \tilde{\kappa}_2}}$$

- When α is known, γ can be obtained from the first log-cumulant as:

$$\gamma = e^{\alpha\tilde{\kappa}_1 - (1-\alpha)\Psi(1) + \log\left(\cos\left(\frac{\pi\alpha}{2}\right)\right)}$$

- By combining the previous expressions, one can also write γ in terms of $\tilde{\kappa}_1$ and $\tilde{\kappa}_1$:

$$\gamma = \exp\left(\sqrt{\frac{\Psi(1,1)}{\Psi(1,1)-\tilde{\kappa}_2}}\tilde{\kappa}_1 - \left[1-\sqrt{\frac{\Psi(1,1)}{\Psi(1,1)-\tilde{\kappa}_2}}\right]\Psi(1) + \log\left[\cos\left(\frac{\pi}{2}\sqrt{\frac{\Psi(1,1)}{\Psi(1,1)-\tilde{\kappa}_2}}\right)\right]\right)$$

We note that the Mellin transform sheds, on the theoretical side, a new and recent light on the heavy-tailed distributions [21].

4.6 A Generalisation of the Rayleigh Distribution

Another example drawn from the α-stable distributions inspired Kuruoğlu and Zerubia to propose a generalisation of the Rayleigh distribution [18]. The pdf has two parameters (α and γ) and its analytical expression is given by the following integral equation:

$$p(u) = u\int_0^\infty ve^{-\gamma v^\alpha} J_0(uv)dv \tag{52}$$

where J_0 is the Bessel function of the second kind. This distribution falls into the heavy-tailed category, since its moments are not defined from a certain order and upwards, with this order given as: $\min(\alpha, 2)$.

To calculate its second kind characteristic function, two approaches are possible:

- The first search for the Mellin transform of this expression [14] led to the following result:

$$\phi(s) = \frac{2^s\Gamma\left(\frac{s+1}{2}\right)\gamma^{\frac{s-1}{\alpha}}\Gamma\left(\frac{1-s}{\alpha}\right)}{\Gamma\left(\frac{1-s}{2}\right)\alpha} \tag{53}$$

- A second approach consists of rewriting (52) on the form of a Mellin correlation:

$$p(u) = u\left(J_0(u)\hat{\otimes}\left(e^{-\gamma u^\alpha}\right)\right)$$

  By using the property in (22), (53) is immediately retrieved.

It can be noted that at $s = 1$, the second kind characteristic function of this distribution goes in the limit to the value 1, since

$$\lim_{s\to 1}\frac{\Gamma\left(\frac{1-s}{\alpha}\right)}{\Gamma\left(\frac{1-s}{2}\right)} = \frac{\alpha}{2}$$

It is thus a valid pdf.

In the vicinity of $s = 1$, this function is defined for $s < 1 + \min(\alpha, 2)$ and for $s > -1$. It is thus well defined in a vicinity of $s = 1$, hence it is legitimate to calculate its log-moments and log-cumulants. Note that the case

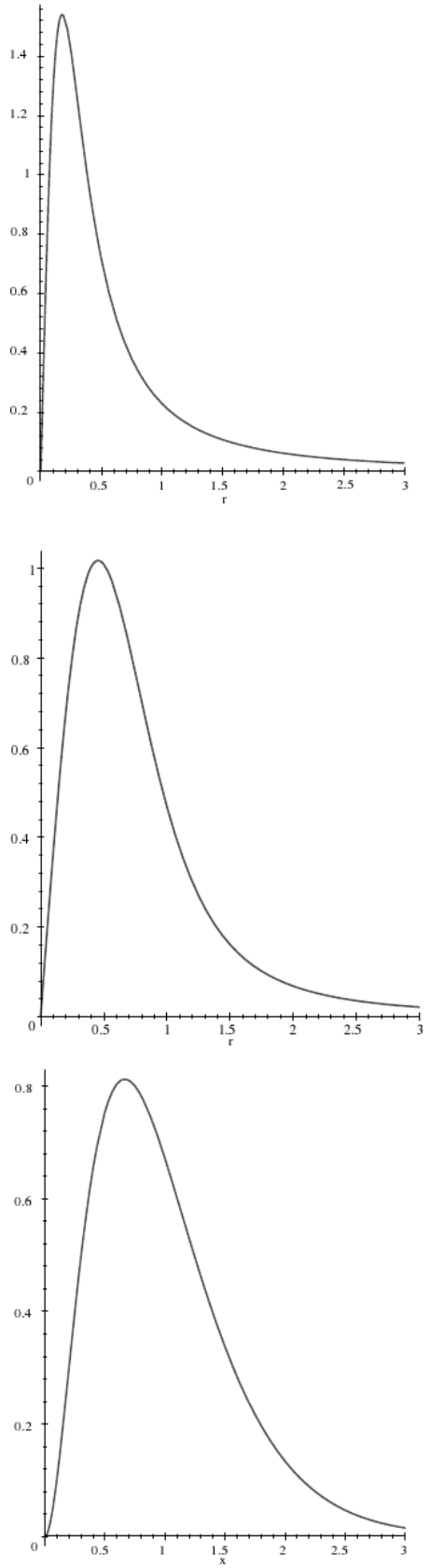

Fig. 10. Generalised Rayleigh distribution [18]: $\alpha = 1$, $\alpha = 1.5$ and $\alpha = 2$.

$\alpha = 2$ gives the Rayleigh distribution and the case $\alpha = 1$ gives the distribution

$$p(u) = \frac{\gamma u}{(u^2+\gamma^2)^{\frac{3}{2}}} .$$

Figure 10 shows this distribution for $\alpha = 1$, $\alpha = 1.5$ and $\alpha = 2$ (the Rayleigh distribution).

Although the analytical form is rather complicated, it is possible to formulate the first and second order log-cumulants of the probability distribution. Again, the expressions obtained are analytical continuations because of discontinuities in the gamma functions. Also in this case, Maple was used to establish following expressions:

$$\tilde{\kappa}_1 = -\Psi(1)\frac{1-\alpha}{\alpha} + \log\left(2\gamma^{\frac{1}{\alpha}}\right)$$
$$\tilde{\kappa}_2 = \frac{\Psi(1,1)}{\alpha^2} .$$

The equation system obtained is easy to handle. The distribution parameters are easily retrieved from the first two log-cumulants:

$$\alpha = \sqrt{\frac{\Psi(1,1)}{\tilde{\kappa}_2}}$$
$$\gamma = e^{\alpha\tilde{\kappa}_1 - \Psi(1)(1-\alpha)} .$$

It is verified that for $\alpha = 2$, the log-cumulants of the Rayleigh distribution are retrieved (with $\mu = 2\sqrt{\gamma}$):

$$\tilde{\kappa}_1 = \frac{1}{2}\Psi(1) + \log(2\sqrt{\gamma})$$
$$\tilde{\kappa}_2 = \frac{1}{4}\Psi(1,1) .$$

5 PARAMETER ESTIMATION

The proposal of a new methodology to evaluate the parameters of a probability distribution requires us to compare the results with those obtained by traditional methods in a realistic setting where N samples are available. In order to decide which method is the preferred one, it is important to establish the variance of the estimators. The goal of this section is to carry out an exhaustive comparison for a schoolbook example: the gamma distribution, for which the method of moments, the method of log-moments, and the method of maximum likelihood are applicable.

5.1 Traditional Methods: Method of Moments Estimation and Maximum Likelihood Estimation

5.1.1 Experimental Framework

Assume that a probability distribution is described by p parameters: α_j, $j \in [1,p]$. Estimation of the parameters describing this distribution is commonly performed by the two following approaches:

- The *method of moments* (MoM) consists of calculating the sample moments of order 1 to p in order to obtain a system of p equations in p unknown pdf parameters. If N samples are available, x_i, $i \in [1,N]$, the rth order sample moment is expressed simply as

$$m_r = \frac{1}{N}\sum_{i=1}^{N} x_i^r .$$

  In order to determine p parameters, it is necessary that all moments up to order p exist, which can pose a problem for instance for distributions with heavy tails. It is also possible to use fractional moments (like FLOM [3]), or lower (and even negative) order moments, whose possible existence is justified in section 2.2 [4], to obtain an equation system which can be solved. Note, however, that the expressions sometimes prove impossible to invert analytically, and the system may also be difficult to invert numerically.
- The *maximum likelihood* approach consists of regarding the N samples x_i as N independent realisations of the distribution which one seeks to estimate, so that they maximise the expression

$$\prod_{i=1}^{N} p_x(x_i)$$

  or, equivalently,

$$\sum_{i=1}^{N} \log(p_x(x_i)) .$$

  With these expressions representing a maximum, calculation of partial derivatives for each parameter then makes it possible to obtain another system of p equations in p unknowns:

$$\frac{\partial\left(\sum_{i=1}^{N}\log(p_x(x_i))\right)}{\partial\alpha_j} = 0 . \qquad (54)$$

  The solution relies on the existence of the partial derivatives, which can pose a problem, as for the $\mathcal{K}$ distribution [16].

5.1.2 Estimator Variance

With several applicable methods available, we must compare them to select the one which is likely to give the user the most reliable results. A natural approach is to seek the method which provides minimum variance for the estimator of a given parameter, knowing that one has a finite number of N samples.

It is known that for the distributions of the exponential family, maximum likelihood estimators attain the Cramer-Rao boundary. Provided that the p partial derivatives in Eq. (54) can be calculated analytically, and that the system of equations can be solved, one obtains p estimators whose variances are minimal. However, many existing distributions (such as the $\mathcal{K}$ distribution) do not have analytical expressions for all partial derivatives,

which then renders the method of maximum likelihood inadequate.

In this case, the use of the method of moments is justified, even if the estimator variance thus obtained is higher. The variance of estimators obtained by the method of moments can be derived through an approach suggested by Kendall [19]. Let m_1 and m_2 be the estimates of the first two moments, and $g(m_1, m_2)$ a function depending only on these two entities. We seek to calculate the variance of the function $g(m_1, m_2)$ by linearising it and writing it as a first-order expansion around the values of the theoretical moments, $m_{0,1}$ and $m_{0,2}$:

$$g(m_1,m_2) = g(m_{0,1},m_{0,2}) + (m_1 - m_{0,1})\frac{\partial g}{\partial m_1}(m_{0,1},m_{0,2}) + (m_2 - m_{0,2})\frac{\partial g}{\partial m_2}(m_{0,1},m_{0,2})\,.$$

After having verified that the $\partial g/\partial m_i$ are not both zero in the point $(m_{0,1}, m_{0,2})$, the variance of g is established as the quadratic error between $g(m_1, m_2)$ and $g(m_{0,1}, m_{0,2})$ due to the following formula [19, Eq. (10.12)]:

$$\begin{aligned}
&\mathrm{Var}\{g(m_1,m_2)\}\\
&= \mathrm{E}\left\{[g(m_1,m_2) - g(m_{0,1},m_{0,2})]^2\right\}\\
&= \mathrm{E}\left\{\left[(m_1 - m_{0,1})\frac{\partial g}{\partial m_1}(m_{0,1},m_{0,2})\right.\right.\\
&\qquad \left.\left.+(m_2 - m_{0,2})\frac{\partial g}{\partial m_2}(m_{0,1},m_{0,2})\right]^2\right\}\\
&= \frac{\partial g}{m_1}(m_{0,1},m_{0,2})^2\,\mathrm{Var}\{m_1\}\\
&\quad + \frac{\partial g}{m_2}(m_{0,1},m_{0,2})^2\,\mathrm{Var}\{m_2\}\\
&\quad + 2\frac{\partial g}{\partial m_1}(m_{0,1},m_{0,2})\frac{\partial g}{\partial m_2}(m_{0,1},m_{0,2})\\
&\qquad \times \mathrm{Cov}\{m_1,m_2\}\,.
\end{aligned} \tag{55}$$

The method can obviously be generalised to functions utilising moments m_i of order i. The definition of the covariance matrix allows us to write:

$$\mathrm{Var}\{m_i\} = \frac{1}{N}(m_{2i} - m_i^2)$$
$$\mathrm{Cov}\{m_i,m_j\} = \frac{1}{N}(m_{i+j} - m_i m_j)$$

5.2 Method of Log-Moments

We propose in this article a new method for analysis of pdfs defined on $\mathbb{R}^+$ based on log-moments and log-cumulants. We will see in this section how to implement it and how to calculate the variance of the estimators obtained.

5.2.1 Description

The method of log-moments (MoLM) consists of calculating estimates of log-moments and log-cumulants in order to obtain a system of p equations in p unknowns (the parameters of the pdf). Assume that we have N samples $x_i, i \in [1, N]$ from the distribution to be estimated. The estimate of the pth order log-moment is expressed simply as

$$\tilde{m}_p = \frac{1}{N}\sum_{i=1}^{N}\log x_i^p\,.$$

To determine p parameters, it is necessary to check in advance that the log-moments up till order p exist. This is in general true, as stated by the theorem of existence, which has been verified for the distributions generally used in signal and image processing.

5.2.2 Estimator Variance

Since we use a logarithmic scale, the criterion of the quadratic error (applied in Eq. (55)), is replaced by another criterion which we will call "normalised quadratic error", E_{nq}, which is in fact the quadratic error calculated on a logarithmic scale:

$$\mathrm{E}_{nq} = \mathrm{E}\left\{\left(\log\left(\frac{x}{y}\right)\right)^2\right\}.$$

In the same spirit, we introduce the second kind variance and covariance, $\tilde{\mathrm{Var}}$ and $\tilde{\mathrm{Cov}}$, on the form

$$\begin{aligned}
\tilde{\mathrm{Var}}\{\tilde{m}_i\} &= \mathrm{E}\left\{\left[(\log x)^i - \tilde{m}_i\right]^2\right\} = \frac{1}{N}\left(\tilde{m}_{2i} - \tilde{m}_i^2\right)\\
\tilde{\mathrm{Cov}}\{\tilde{m}_i,\tilde{m}_j\} &= \mathrm{E}\left\{\left[(\log x)^i - \tilde{m}_i\right]\left[(\log x)^j - \tilde{m}_j\right]\right\}\\
&= \frac{1}{N}(\tilde{m}_{i+j} - \tilde{m}_i\tilde{m}_j)
\end{aligned}$$

where $\tilde{m}_i$ is the ith order log-moment.

With this new approach, and taking the preceding step as starting point, let the function g be expressed in terms of the first two estimated log-moments as $g(\tilde{m}_i, \tilde{m}_2)$. Then g can be expanded around the first two theoretical log-moments, $\tilde{m}_{0,1}$ and $\tilde{m}_{0,2}$, as

$$g(\tilde{m}_1,\tilde{m}_2) = g(\tilde{m}_{0,1},\tilde{m}_{0,2}) + (\tilde{m}_1 - \tilde{m}_{0,1})\frac{\partial g}{\partial \tilde{m}_1}(\tilde{m}_{0,1},\tilde{m}_{0,2}) + (\tilde{m}_2 - \tilde{m}_{0,2})\frac{\partial g}{\partial \tilde{m}_2}(\tilde{m}_{0,1},\tilde{m}_{0,2})\,.$$

After verifying that the $\partial g/\partial \tilde{m}_i$ are not both zero in $(\tilde{m}_{0,1}, \tilde{m}_{0,2})$, the variance of g is established by the same formula applied in the previous section.

$$\begin{aligned}
&\mathrm{Var}\{g(\tilde{m}_1,\tilde{m}_2)\} = \mathrm{E}\left\{[g(\tilde{m}_1,\tilde{m}_2) - g(\tilde{m}_{0,1},\tilde{m}_{0,2})]^2\right\}\\
&= \mathrm{E}\left\{\left[(\tilde{m}_1 - \tilde{m}_{0,1})\frac{\partial g}{\partial \tilde{m}_1}(\tilde{m}_{0,1},\tilde{m}_{0,2})\right.\right.\\
&\qquad \left.\left.+(\tilde{m}_2 - \tilde{m}_{0,2})\frac{\partial g}{\partial \tilde{m}_2}(\tilde{m}_{0,1},\tilde{m}_{0,2})\right]^2\right\}\\
&= \frac{\partial g}{\tilde{m}_1}(\tilde{m}_{0,1},\tilde{m}_{0,2})^2\tilde{\mathrm{Var}}\{\tilde{m}_1\} + \frac{\partial g}{\tilde{m}_2}(\tilde{m}_{0,1},\tilde{m}_{0,2})^2\tilde{\mathrm{Var}}\{\tilde{m}_2\}\\
&\quad + 2\frac{\partial g}{\partial \tilde{m}_1}(\tilde{m}_{0,1},\tilde{m}_{0,2})\frac{\partial g}{\partial \tilde{m}_2}(\tilde{m}_{0,1},\tilde{m}_{0,2}) \times \tilde{\mathrm{Cov}}\{\tilde{m}_1,\tilde{m}_2\}\,.
\end{aligned}$$

As in the previous, this method can obviously be generalised to functions of the moments $\tilde{m}_i$ of unspecified order i.

5.3 The Gamma Distribution Case

We will use the gamma distribution as an example to compare the available methods. This distribution is not heavy tailed, thus the method of moments can be used, as well as the method of log-moments. The partial derivatives with respect to the parameters are known, which makes it possible to apply maximum likelihood estimation.

5.3.1 Variance of the Gamma Distribution Parameter Estimators with the Method of Moments

The method of moments (MoM) utilises the first two moments to deduce estimates of L and μ (Eqs. (30) and (31)). The method of Kendall, presented in Section 5.1.2, gives the following variance for the estimators of μ and L:

$$\mathrm{Var}_{MoM}\{\hat{\mu}\} = \frac{\mu^2}{NL} \tag{56}$$

$$\mathrm{Var}_{MoM}\{\hat{L}\} = \frac{2L(L+1)}{N} \tag{57}$$

5.3.2 Variance of the Gamma Distribution Parameter Estimators with the Method of Log-Moments

The parameter L is derived from (33) as

$$\Psi(1, L) = \tilde{\kappa}_{x(2)}$$

which can be rewritten as a function of $(\tilde{m}_1, \tilde{m}_2)$:

$$\Psi(1, L) = \tilde{m}_2 - \tilde{m}_1^2 .$$

One then carries out the limited expansion proposed in the previous, which requires the use of implicit differentiation. Although the expression brings into play the first to fourth order log-moments, the result can be simplified and we obtain:

$$\mathrm{Var}_{MoLM}\{\hat{L}\} = \frac{1}{N}\frac{\Psi(3, L) + 2\Psi(1, L)^2}{\Psi(2, L)^2} . \tag{58}$$

Figure 11 (left panel) presents the ratio of the standard deviation for the variance of MoLM estimate of $\hat{L}$ to the standard deviation for the variance of the MoM estimate of $\hat{L}$. The whole motivation for using the new method is evident for low values of L, where the improvement approaches 30%. When the variance of the different estimators is fixed, this results in the same amount of shrinking of the analysis window, and therefore a better spatial localisation of the estimate.

For the parameter μ, the calculation is much more elaborate, and we finally arrive at the following expression, whose interpretation is not simple, but which can easily be implemented numerically:

$$\begin{aligned}\mathrm{Var}_{MoLM}\{\hat{L}\} = &-\frac{1}{N}\frac{\mu^2}{L^2\Psi(2, L)^2} \\ &\times \big[2\Psi(1, L)L\Psi(3, L) - \Psi(1, L)^2L^2\Psi(3, L) \\ &+ 4\Psi(1, L)^3L - 2\Psi(1, L)^4L^2 - 2L\Psi(2, L)^2 \\ &+\Psi(1, L)L^2\Psi(2, L)^2 - 2\Psi(1, L)^2 - \Psi(3, L)\big]\end{aligned}$$

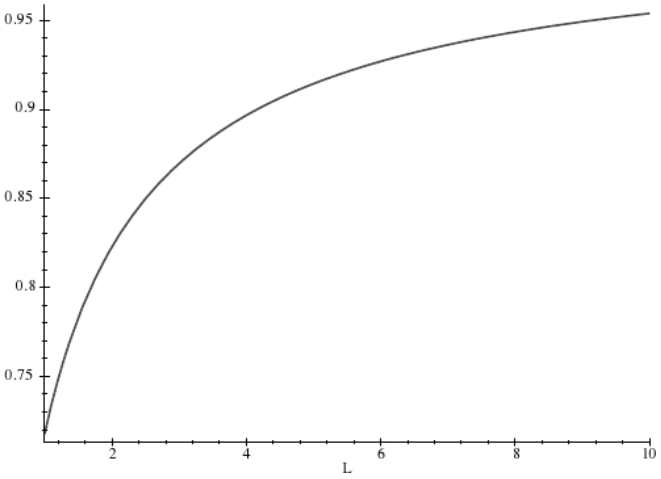

Fig. 12. Gamma distribution: Comparison of the variance of the estimator for μ by the method of log-moments with the method of moments. The curve represents the ratio of the standard deviations for values of L between 1 and 10.

Also Figure 12 presents the ratio of the standard deviation of $\hat{\mu}$ calculated by the MoLM to the standard deviation of $\hat{\mu}$ calculated by the MoM.

It can be noted that the MoM provides better results for low values of L. Recall moreover that this is also the maximum likelihood estimator and thus attains minimum variance (i.e. the Cramer-Rao bound).

5.3.3 Variance of the Gamma Distribution Parameter Estimators with the Method of Lower Order Moments

The existence of the second kind characteristic function for values of s lower than 1 justifies the use of the method of lower order moments (MoLOM), i.e. negative ones. In the case of the gamma distribution, it is known that the lower order moments exist for $\nu > -L$. For a given value of ν it is verified that $\nu > -L$ and using the three moments μ_ν, $\mu_{\nu+1}$ and $\mu_{\nu+2}$ it is easy to show that $\hat{\mu}$ and $\hat{L}$ can be derived from the relation:

$$\hat{\mu} = \frac{\hat{m}_{\nu+1}}{\hat{m}_\nu}(1+\nu) - \nu\frac{\hat{m}_{\nu+2}}{\hat{m}_{\nu+1}}$$

$$\hat{L} = \frac{1}{\hat{m}_\nu\left(\frac{\hat{m}_{\nu+2}}{\hat{m}_{\nu+1}}\right) - 1} - \nu$$

For $\nu = 0$, this reduces to the MoM (Eqs. (30) and (31)).

The variances of the estimators for μ and L can be established by the method of Kendall, used in Section 5.3.1 (for the MoM). For L, the following expression is obtained:

$$\begin{aligned}\mathrm{Var}_{MoLOM}\{\hat{L}\} = &\frac{1}{N}\frac{\Gamma(L)\Gamma(2\nu+L)}{\Gamma(\nu+L)^2} \\ &\times \big(2L(L+1) + \nu[4L(\nu+2) + (\nu+4)(\nu+1)^2]\big)\end{aligned} \tag{59}$$

The problem with this relation is that it has a minimum for ν, which cannot be expressed explicitly as a function of L. The optimal values of ν must be calculated numerically. Table 4 gives some values of ν as a function

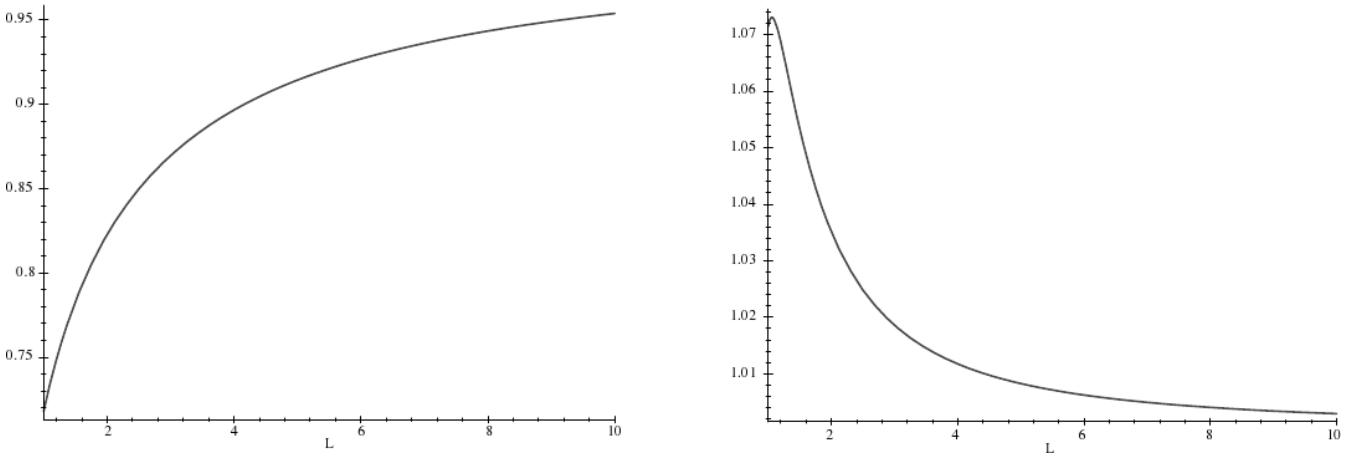

Fig. 11. To the left, a comparison of the variance of the estimator for L by the method of log-moments and the method of moments. To the right, a comparison of the variance of the estimators of L by the method of log-moments and the method of lower order moments (only for the value of $\nu = -0.35$). The curves represent the standard deviation for values of L between 1 and 10.

TABLE 4
Gamma distribution estimated with the method of lower order moments. Optimal values of the parameter ν that minimises the variance of $\hat{L}$ as a function of L

L	ν_{opt}
1	-0.35
2	-0.44
3	-0.56
4	-0.59

of L. When information about L is absent, the choice of $\nu = -0.35$ seems to be a good compromise.

Figure 11 (right panel) presents the ratio of the standard deviation of $\hat{L}$ calculated by the MoLM to the standard deviation of $\hat{L}$ calculated by the MoLOM with $\nu = -0.35$. It is interesting to note that the MoLOM is slightly better than the MoLM. Nevertheless, if one wants to fully utilise this method, then one must know L to be able to choose the optimal value of ν. As the difference is altogether rather small, we promote the MoLM because it does not require us to determine a parameter in order to make optimal use of the method.

However, it is easily shown that minimum variance is obtained for $\nu = 0$, which is less than astonishing since this value corresponds to the maximum likelihood estimator.

5.3.4 Summary

We propose to summarise these results by posting in Table 5 the optimal window dimension for these three methods when we seek to reach an error of 10% for the estimate of the shape parameter L (i.e. the standard deviation is 10% of the value to be estimated). For the parameter μ, only the method of moments (which coincides with the maximum likelihood estimator) is used.

TABLE 5
Number of samples (and examples of the analysis window) needed to estimate the parameters L and μ of a gamma distribution with 10% error. The methods used are, for the shape parameter L, the method of moments (MoM), the method of lower order moments (MoLOM) with $\nu = -0.35$, and the method of log-moments (MoLM). The Cramer-Rao bound (CRB) is calculated by the means of the Fisher information matrix. For μ, only the MoM is used.

Gamma distribution					
L	$\hat{L}$				$\hat{\mu}$
	MoM	MoLOM	MoLM	CRB	MoM
1	400 20×20	179 13×13	206 14×14	155 12×12	100 10×10
2	300 17×17	189 14×14	203 14×14	172 13×13	50 7×7
3	267 16×16	194 14×14	202 14×14	180 13×13	33 6×6
5	240 15×15	197 14×14	201 14×14	187 14×14	20 4×4
10	220 14×14	199 14×14	200 14×14	194 14×14	10 3×3

We first remark that for an identical relative error (10%), the estimate of L requires much more samples

TABLE 6
Number of samples (and examples of the analysis windows) needed to estimate the parameters σ and μ of a Gaussian distribution with 10% error.

Gaussian distribution		
σ	$\hat{\sigma}$	$\hat{\mu}$
1	200 14×14	100 10×10
0.707	200 14×14	50 7×7
0.577	200 14×14	33 6×6
0.447	200 14×14	20 4×4
0.316	200 14×14	10 3×3

than the estimate of μ. Secondly, the estimate of μ requires much less samples when L is large, i.e. the distribution is localised. Lastly, it is interesting to note a characteristic feature of the MoLM: It requires about the same number of samples for all values of L, whereas the MoM requires a much higher number of samples when L is small. From this, two remarks can be made:

- It can be shown that the variance of $\hat{L}$ for the MoLM (Eq. (58)) is almost quadratic in L:

$$\frac{\Psi(3,L)+2\Psi(1,L)^2}{\Psi(2,L)^2} \simeq 2L^2$$

  Thus, if a constant relative error is sought, the number of samples is independent of L.
- It is interesting to analyse the same problem for the Gaussian distribution $\mathcal{N}[\mu,\sigma^2]$. It is easy to show that the variances of the estimators of μ and σ do not depend on σ for the MoM. They are written:

$$\mathrm{Var}_{\mathcal{N}}(\mu) = \frac{\sigma^2}{N}$$
$$\mathrm{Var}_{\mathcal{N}}(\sigma) = \frac{\sigma^2}{2N}$$

  By choosing Gaussian distributions with $\mu = 1$, the values of σ become comparable and an identical criterion, and the required window sizes can be calculated. These are included in Table 6.

It is seen that the MoM needs a constant number of samples to estimate the shape parameter σ, an analogy to the property of the MoLM for the gamma distribution.

To achieve this analysis, we calculate Fisher's information matrix for the gamma distribution[4]:

$$\begin{bmatrix} \frac{L}{\mu^2} & 0 \\ 0 & \Psi(1,L)-\frac{1}{L} \end{bmatrix},$$

which allows the calculation of the Cramer-Rao bound, given in Table 5.

5.4 The Mixture of Gamma Distribution Case

The analytical calculation of the variance of the estimators in the mixture of gamma distributions case described in Section 4.4 does not pose any problem, except for the apparent complexity of the expressions obtained, whose length prohibits us from including them in a publication. Another possibility would be to assess them by numerical evaluation.

Table 7 presents the standard deviations of λ and ρ if the analysis is carried out in neighbourhood of 100 samples (a 10×10 window), for various values of ρ and λ.

Table 8 presents the optimal dimension of a square window which guarantees a maximum of 10% estimation error (where the error is defined as the ratio of the standard deviation to the estimated value). Note that for $L = 1$, a large window size is required, which is not surprising when recalling Figure 8 showing that a mixture of gamma distributions is generally unimodal.

6 Conclusions

Second kind statistics seem to be an innovative and powerful tool for the study of distributions defined on $\mathbb{R}^+$. The analytical formulation of the log-moments and the log-cumulants is indeed particularly simple and easy to exploit. At least, this is true for the examples presented in this article, whereof some, such as the mixture distributions, are not commonplace. Moreover, the variance of the estimators thus defined approaches the minimal values reached by the maximum likelihood method, while avoiding some of the analytical pitfalls. This approach shows great potential in certain applications in SAR image processing (such as the characterisation of an optimal homomorphic filter [11]). One can reasonably question why this approach, in all its simplicity, has not been proposed before. Several reasons can be called upon:

- The first is based on the observation that a Mellin transform of a pdf is only a Fourier transform of the same pdf taken on a logarithmic scale. Even if this step is perfectly justified on the theoretical level, all the possible advantages of moving into the Mellin domain remain hidden, such as the use of existing tables of known Mellin transforms, or the direct use of the log-moments and log-cumulants that produce better estimates of the distribution parameters.

4. The diagonal form of this matrix justifies *a posteriori* the analytical expression of the gamma distribution that we chose, which differs slightly from the one found in reference book like [9], [19].

TABLE 7
Standard deviation (SD) for the estimates of λ and ρ in the case of a mixture of gamma distributions. The size of the analysis window is 10×10. Since the SD is inversely proportional to the square root of the number of samples in the analysis window, the table can serve to determine the optimal window after a maximum error has been set.

$\rho = 2$			$\rho = 5$			$\rho = 10$		
λ	SD_λ	SD_ρ	λ	SD_λ	SD_ρ	λ	SD_λ	SD_ρ
.1	.553	3.874	.1	.071	2.638	.1	.043	3.867
.2	.617	1.967	.2	.083	1.606	.2	.052	2.481
.3	.676	1.235	.3	.092	1.183	.3	.059	1.918
.4	.730	.784	.4	.099	.925	.4	.063	1.591
.5	.780	.455	.5	.105	.749	.5	.065	1.380

TABLE 8
Optimal window size for a relative error of 10% in λ and ρ.

$\rho = 2$			$\rho = 5$			$\rho = 10$		
λ	SD_λ	SD_ρ	λ	SD_λ	SD_ρ	λ	SD_λ	SD_ρ
.1	553×553	194×194	.1	71×71	53×53	.1	43×43	39×39
.2	309×309	98×98	.2	41×41	32×32	.2	26×26	25×25
.3	225×225	62×62	.3	31×31	24×24	.3	20×20	19×19
.4	182×182	39×39	.4	25×25	19×19	.4	16×16	16×16
.5	156×156	23×23	.5	21×21	15×15	.5	13×13	14×14

- The analysis of the product model, which is reserved for particular processes like coherent imaging, has not received the same strong attention as the additive signal model. The philosophy adopted for the study of the product model has too often consisted of transformation into logarithmic scale, in order to use the known tools for the additive model. This reductional step quickly pose problems, as it requires large control of the analytical expressions thus obtained. It is probably the reason why non-experts have written off other distributions than the gamma distribution and the inverse gamma distribution, such as the $\mathcal{K}$ distribution, for instance.
- Finally, the Mellin transform has been completely ignored. Its applications has been confined to certain specialised applications, which has unfortunately prevented diffusion of the method beyond the field of study (e.g., radar and sonar signals, number theory, ultrasound propagation in heterogeneous media, the Fourier-Mellin transform in image processing). Even if certain pieces of work, old [15] as well as recent [22], [21] ones, have shown its applicability in the field of probability, its use has been very restricted. Therefore, few people know the fundamental properties, or even the exact definition.

The unfortunate consequence of the confidentiality is that few research groups have worked on the subject. Therefore, powerful and sufficiently general numerical implementations of the analytical transform are still missing. These would make it possible to consider numerical deconvolutions of the probability distributions described by a Mellin convolution, and thus to recover significant parameters of a SAR scene [23].

REFERENCES

[1] M. Métivier, *Notions Fundamentales de la Théorie des Probabilités*. Dunod, 1972.

[2] S. Colombo, *Les Transformations de Mellin et de Hankel*. Centre National de la Recherche Scientifique, 1959.

[3] M. Shao and C.-L. Nikias, "Signal processing with fractional lower order moments," *Proc. of the IEEE*, vol. 81, no. 7, pp. 986–1010, Jul. 1993.

[4] J.-M. Nicolas, A. Maruani, and R. Badeau, "Les moments d'ordre inferieur: Principles et application au filtrage des images RSO," in *RFIA'00*, Paris, France, Jan. 2000.

[5] J.-M. Nicolas, M. Sigelle, C. Thuillier, and F. Tupin, "Images de radar à ouverture synthétique: Transformée de Mellin et multirésolution," in *GRETSI'97*, Grenoble, France, Sep. 1997.

[6] J.-M. Nicolas and A. Maruani, "Speckle well modeled by Mellin transform," in *PIERS'98*, Nantes, France, Jul. 1998.

[7] H. Cartan, *Théorie Élémentaire des Fonctions Analytiques d'Une ou Plusieurs Variables Complexes*. Hermann, 1961.

[8] E. Lukacs, *Developments in Characteristic Functions Theory*. Charles Griffin, 1983.

[9] N. Johnson and S. Kotz, *Continuous Univariate Distributions 1*. John Wiley & Sons, 1970.
[10] F. Oberhettinger, *Tables of Mellin Transform*. Springer, 1974.
[11] J.-M. Nicolas, "Filtrage homomorphique optimal RSO (Radar à Synthèse d'Ouverture)," in *GRETSI'01*, Toulouse, France, Sep. 2001.
[12] W. C. Hoffmann, *Statistical Methods in Radio Wave Propagation*. Pergamon Press, 1960.
[13] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946.
[14] J.-M. Nicolas, "Introduction aux statistique de deuxième espèce: Applications aux lois d'images RSO," ENST, Tech. Rep. 2002D001, 2002.
[15] Z. A. Lomnicki, "On the distribution of products of random variables," *J. Royal Statistics Soc. Ser. B*, vol. 29, no. 3, pp. 513–524, 1967.
[16] C.-J. Oliver, "Optimum texture estimators for SAR clutter," *J. Phys. D.: Appl. Phys.*, vol. 26, pp. 1824–1835, 1993.
[17] R.-D. Pierce, "Application of the positive alpha-stable distribution," in *IEEE Signal Processing Workshop on Higher-Order Statistics*. IEEE, 1997, pp. 420–424.
[18] E. Kuruoğlu and J. Zerubia, "Modelling SAR images with a generalisation of the Rayleigh distribution," INRIA, Tech. Rep. 4121, 2001.
[19] A. Stuart and J. Keith, Eds., *Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory*, 5th ed. Griffin, 1987.
[20] M.-J. Wasilewski, "Sur certaines propriétés de la distribution Gamma généralisée," *Revue de Statistique Apliquées*, vol. 15, no. 1, pp. 95–105, 1967.
[21] A. Tagliani, "Numerical inversion of the Mellin transform on the real line for heavy-tailed probability density functions," *Appl. Math. and Computation*, vol. 130, no. 2-3, pp. 525–536, 2002.
[22] ——, "Recovering a probability density function from its Mellin transform," *Appl. Math. and Computation*, vol. 118, no. 2-3, pp. 151–159, 2001.
[23] J.-M. Nicolas and A. Maruani, "Numerical Mellin transform applied to texture classification on SAR images," in *PIERS'99*, Taiwan, Mar. 1999.

Bibliography

[Abdelfattah and Nicolas, 2006] R. Abdelfattah and J.-M. Nicolas. *Interferometric SAR coherence magnitude estimation using second kind statistics*. *IEEE Trans. Geosci. Remote Sens.*, **44**(7, part 2): 1942–1953, July 2006.

[Achim et al., 2006] A. Achim, E. E. Kuruoğlu and J. Zerubia. *SAR image filtering based on the heavy-tailed Rayleigh model*. *IEEE Trans. Image Process.*, **15**(9): 2686–2693, September 2006.

[Anfinsen et al., 2009] S. N. Anfinsen, T. Eltoft and A. P. Doulgeris. *A relaxed Wishart model for polarimetric SAR data*. In *Proc. 4th Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '09)*, volume ESA SP-668. Frascati, Italy, 8 pp., April 2009.

[Barakat, 1986] R. Barakat. *Weak-scatterer generalization of the K-density function with application to laser scattering in atmospheric turbulence*. *J. Opt. Soc. Am. A*, **3**(4): 401–409, April 1986.

[Bartle, 1995] R. G. Bartle. *The Elements of Integration and Lebesgue Measure*. John Wiley & Sons, New York, USA, Wiley Classics Library edition, 1995.

[Benboudjema et al., 2007] D. Benboudjema, F. Tupin, W. Pieczynski, M. Sigelle and J.-M. Nicolas. *Unsupervised segmentation of SAR images using triplet Markov fields and Fisher noise distributions*. In *Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS'07*, volume 1, pp. 3891–3894. Barcelona, Spain, 23-27 July 2007.

[Bertrand et al., 1990] J. Bertrand, P. Bertrand and J.-P. Ovarlez. *Discrete Mellin transform for signal analysis*. In *Proc. IEEE Int. Conf. Acoustics Speech Signal Process., ICASSP'90*, pp. 1603–1606. Albuquerque, USA, 3-6 April 1990.

[Bertrand et al., 2000] J. Bertrand, P. Bertrand and J.-P. Ovarlez. *The Mellin transform*. In *The Transform and Applications Handbook*, edited by A. D. Poularikas, chapter 11. CRC Press, Boca Raton, USA, second edition, 2000.

[Blacknell, 1994] D. Blacknell. *Comparison of parameter estimators for K-distribution*. *IEE Proc. Radar, Sonar, Navig.*, **141**(1): 45–52, February 1994.

[Bombrun and Beaulieu, 2008] L. Bombrun and J.-M. Beaulieu. *Fisher distribution for texture modeling of polarimetric SAR data*. *IEEE Geosci. Remote Sens. Lett.*, **5**(3): 512–516, July 2008.

[Born and Wolf, 1999] M. Born and E. Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, Cambridge, UK, 7th edition, 1999.

[Bujor et al., 2004] F. Bujor, E. Trouvé, L. Valet, J.-M. Nicolas and J.-P. Rudant. *Application of log-cumulants to the detection of spatiotemporal discontinuities in multitemporal SAR images*. *IEEE Trans. Geosci. Remote Sens.*, **42**(10): 2073–2084, October 2004.

[Butzer and Jansche, 1997] P. L. Butzer and S. Jansche. *A direct approach to the Mellin transform*. *J. Fourier Anal. Appl.*, **3**(4): 325–376, July 1997.

[Chen and Liu, 2008] G. Chen and X. Liu. *Wavelet-based SAR image despeckling using Cauchy pdf modeling*. In *IEEE Radar Conf. 2008*, pp. 1–5. Rome, Italy, 26-30 May 2008.

[Clausi and Yue, 2004] D. A. Clausi and B. Yue. *Comparing co-occurence probabilities and Markov random fields for texture analysis of SAR sea ice imagery*. *IEEE Trans. Geosci. Remote Sens.*, **42**(1): 215–228, January 2004.

[Cloude, 2010] S. R. Cloude. *Polarisation: Applications in Remote Sensing*. Oxford University Press, Oxford, UK, 2010.

[Cohen, 1993] L. Cohen. *The scale representation*. *IEEE Trans. Signal Process.*, **41**(12): 3275–3292, December 1993.

[Conradsen et al., 2003] K. Conradsen, A. A. Nielsen, J. Schou and H. Skriver. *A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data*. *IEEE Trans. Geosci. Remote Sens.*, **41**(1): 4–19, January 2003.

[Cook, 1981] I. D. Cook, Jr. *The H-Function and Probability Density Functions of Certain Algebraic Combinations of Independent Random Variables with H-Function Prabability Distribution*. Ph.D. thesis, University of Texas at Austin, Austin, USA, May 1981.

[Cumming and Wong, 2005] I. G. Cumming and F. H. Wong. *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*. Artech House, Norwood, USA, 2005.

[De Grandi et al., 2009] G. D. De Grandi, R. M. Lucas and J. Kropacek. *Analysis by wavelet frames of spatial statistics in SAR data for characterizing structural properties of forest*. *IEEE Trans. Geosci. Remote Sens.*, **47**(2): 494–507, February 2009.

[Debnath and Bhatta, 2007] L. Debnath and D. Bhatta. *Integral Transforms and their Applications*, chapter 8. Chapman & Hall/CRC, Boca Raton, USA, second edition, 2007.

[Delignon et al., 2001] Y. Delignon, R. Fjørtoft and W. Pieczynski. *Compound distributions for radar images*. In *Proc. Scand. Conf. Image Anal.*, pp. 741–748. Bergen, Norway, 11-14 June 2001.

[Dolan, 1964] B. A. Dolan. *The Mellin transform for moment-generation and for the probability density of products and quotients of random variables*. *Proc. IEEE*, **52**(12): 1745–1746, December 1964.

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

[Doulgeris et al., 2008] A. P. Doulgeris, S. N. Anfinsen and T. Eltoft. *Classification with a non-Gaussian model for PolSAR data*. *IEEE Trans. Geosci. Remote Sens.*, **46**(10): 2999–3009, October 2008.

[Eltoft, 2006] T. Eltoft. *Modeling the amplitude statistics of ultrasonic images*. *IEEE Trans. Med. Imag.*, **25**(2): 229–240, February 2006.

[Eltoft et al., 2006] T. Eltoft, T. Kim and T.-W. Lee. *Multivariate scale mixture of Gaussians modeling*. In *Independent Component Analysis and Blind Signal Separation*, edited by J. Rosca, D. Erdogmus, J. C. Principe and S. Haykin, volume 3889 of *Lecture Notes in Computer Science*, pp. 799–806. Springer-Verlag, Berlin, Germany, 2006.

[Epstein, 1948] B. Epstein. *Some applications of the Mellin transform in statistics*. *Ann. Math. Statist.*, **19**(3): 370–379, September 1948.

[Fikioris, 2006] G. Fikioris. *Integral evaluation using the Mellin transform and generalized hypergeometric functions: Tutorial and applications to antenna problems*. *IEEE Trans. Antennas Propag.*, **54**(12): 3895–3907, December 2006.

[Fikioris, 2007] G. Fikioris. *Mellin transform method for integral evaluation: Introduction and applications to electromagnetics*. In *Synthesis Lectures on Computational Electromagnetics*, edited by C. A. Balanis, volume 2, pp. 1–67. Morgan & Claypool, January 2007.

[Flajolet et al., 1995] P. Flajolet, X. Gourdon and P. Dumas. *Mellin transforms and asymptotics: Harmonic sums*. *Theoretical Computer Science*, **144**(1/2): 3–58, June 1995.

[Franceschetti et al., 1995] G. Franceschetti, V. Pascazio and G. Schirinzi. *Iterative homomorphic technique for speckle reduction in synthetic-aperture radar imaging*. *J. Opt. Soc. Am. A*, **12**(4): 686–692, April 1995.

[Freeman and Durden, 1998] A. Freeman and S. L. Durden. *A three-component scattering model for polarimetric SAR data*. *IEEE Trans. Geosci. Remote Sens.*, **36**(3): 963–973, May 1998.

[Freeman and Saatchi, 2004] A. Freeman and S. S. Saatchi. *On the detection of Faraday rotation in linearly polarized L-band SAR backscatter signatures*. *IEEE Trans. Geosci. Remote Sens.*, **42**(8): 1607–1616, August 2004.

[Freitas et al., 2005] C. C. Freitas, A. C. Frery and A. H. Correia. *The polarimetric G distribution for SAR data analysis*. *Environmetrics*, **16**(1): 13–31, February 2005.

[Galetti et al., 2008] M. Galetti, D. H. O. Bebbington, M. Chandra and T. Börner. *Measurement and characterization of entropy and degree of polarization of weather radar targets*. *IEEE Trans. Geosci. Remote Sens.*, **46**(9): 3196–3207, October 2008.

[Galland et al., 2009] F. Galland, J.-M. Nicolas, H. Sportouche, M. Roche, F. Tupin and P. Réfrégier. *Unsupervised synthetic aperture radar image segmentation using Fisher distributions*. *IEEE Trans. Geosci. Remote Sens.*, **47**(8): 2966–2972, August 2009.

[Gierull and Sikaneta, 2002] C. H. Gierull and I. C. Sikaneta. *Estimating the effective number of looks in interferometric SAR data*. *IEEE Trans. Geosci. Remote Sens.*, **40**(8): 1733–1742, August 2002.

[Glickman, 2000] T. S. Glickman, ed. *AMS Glossary of Meteorology*. American Meteorology Society, Boston, USA, second edition, 2000.

[Goodman, 1975] J. W. Goodman. *Statistical properties of laser speckle patterns*. In *Laser Speckle and Related Phenomena*, edited by J. Dainty, volume 9 of *Topics in Applied Physics*, pp. 9–75. Springer-Verlag, New York, USA, 1975.

[Goodman, 2000] J. W. Goodman. *Statistical Optics*. John Wiley & Sons, New York, USA, Wiley Classics Library edition, 2000.

[Goodman, 2007] J. W. Goodman. *Speckle Phenomena in Optics: Theory and Applications*. Ben Roberts, Greenwood Village, USA, 2007.

[Goodman, 1963] N. R. Goodman. *Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)*. *Ann. Math. Statist.*, **34**(1): 152–177, March 1963.

[Harant et al., 2009] O. Harant, L. Bombrun, M. Gay, R. Fallourd, E. Trouvé and F. Tupin. *Segmentation and classification of polarimetric SAR data based on the KummerU distribution*. In *Proc. 4th Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry (POLinSAR '09)*, volume ESA SP-668. Frascati, Italy, 6 pp., April 2009.

[Hoekman, 1991] D. H. Hoekman. *Speckle ensemble statistics of logarithmically scaled data*. *IEEE Trans. Geosci. Remote Sens.*, **29**(1): 180–182, January 1991.

[IEEE, 2009] IEEE. *IEEE Standard for Letter Designations for Radar-Frequency Bands*. IEEE Standard 510-2002(R2009). IEEE Press, New York, USA, 2009.

[Jakeman and Pusey, 1976] E. Jakeman and P. N. Pusey. *A model for non-Rayleigh sea echo*. *IEEE Trans. Antennas Propag.*, **24**(6): 806–814, November 1976.

[Jakeman and Tough, 1987] E. Jakeman and R. Tough. *Generalized k distribution: a statistical model for weak scattering*. *J. Opt. Soc. Am. A*, **4**(9): 1764–1772, September 1987.

[Kaiser, 1996] G. Kaiser. *Wavelet filtering with the Mellin transform*. *Appl. Math. Lett.*, **9**(5): 69–74, September 1996.

[Kersten et al., 2005] P. R. Kersten, J.-S. Lee and T. L. Ainsworth. *A comparison of change detection statistics in POLSAR images*. In *Proc. IEEE Int. Geosc. Remote Sens. Symp., IGARSS'05*, volume 7, pp. 4836–4839. IEEE, Seoul, South Korea, July 2005.

[Kotlarski, 1965] I. Kotlarski. *On the generalized Mellin transform of a complex random variable and its application*. *Ann. Math. Stat.*, **36**(5): 1459–1467, October 1965.

[Kreithen and Hogan, 1991] D. E. Kreithen and G. G. Hogan. *Statistical analysis of Ka-band sea clutter*. In *Proc. IEEE OCEANS '91*, volume 2, pp. 1217–1222. IEEE, Honolulu, USA, October 1991.

[Langley et al., 2009] K. Langley, P. Lacroix, S.-E. Hamran and O. Brandt. *Sources of backscatter at 5.3 GHz from a superimposed ice and firn area revealed by multi-frequency GPR and cores*. *J. Glaciology*, **55**(190): 373–383, April 2009.

[Lee, 2005] H.-C. Lee. *Introduction to Color Imaging Science*, chapter 8. Cambridge University Press, Cambridge, UK, 2005.

[Lee and Pottier, 2009] J.-S. Lee and E. Pottier. *Polarimetric Radar Imaging: From Basics to Applications*. Number 143 in Optical Science and Engineering. CRC Press, Boca Raton, USA, 2009.

[Lee et al., 1994] 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*, volume 4, pp. 2179–2181. Pasadena, USA, August 1994.

[Lomnicki, 1967] Z. A. Lomnicki. *On the distribution of products of random variables*. *J. Royal Stat. Soc. B*, **29**(3): 513–524, 1967.

[Lopès et al., 2008] A. Lopès, R. Garello and S. Le Hégarat-Mascle. *Speckle models*. In *Processing of Synthetic Aperture Radar Images*, edited by H. Maître, chapter 5, pp. 87–142. John Wiley & Sons, London, UK, 2008.

[López-Martínez and Fàbregas, 2005] C. López-Martínez and X. Fàbregas. *Polarimetric SAR speckle noise model*. *IEEE Trans. Geosci. Remote Sens.*, **41**(10): 2232–2242, oct 2005.

[Marichev, 1982] O. I. Marichev. *Handbook of Integral Transforms of Higher Transcendental Functions: Theory and Algorithmic Tables*. Mathematics and Its Applications. Ellis Northwood, Chichester, USA, 1982. Translated from Russian by L. W. Longdon.

[Marzano and Weinman, 2008] F. S. Marzano and J. A. Weinman. *Inversion of spaceborne X-band synthetic aperture radar measurements for precipitation remote sensing over land*. *IEEE Trans. Geosci. Remote Sens.*, **46**(11): 3472–3487, November 2008.

[Massonnet and Souyris, 2008] D. Massonnet and J.-C. Souyris. *Imaging with Synthetic Aperture Radar*. EPFL Press, Lausanne, Switzerland, 2008.

[Mathai, 1978] A. M. Mathai. *Some results on functions of matrix argument*. *Math. Nachr.*, **84**(1): 171–177, 1978.

[Mathai, 1981] A. M. Mathai. *Distribution of the canonical correlation matrix*. *Ann. Inst. Statist. Math.*, **33, part A**: 35–43, 1981.

[Mathai, 1997] A. M. Mathai. *Jacobians of Matrix Transformations and Functions of Matrix Arguments*, chapter 6. World Scientific, New York, USA, 1997.

[Mathur and Krishna, 1977] B. Mathur and S. Krishna. *On multivariate fractional integration operators*. *Indian J. Pure Appl. Math.*, **8**: 1078–1082, 1977.

[Moser and Serpico, 2006] G. Moser and S. B. Serpico. *Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude imagery*. *IEEE Trans. Geosci. Remote Sens.*, **44**(10): 2972–2982, October 2006.

[Moser and Serpico, 2009] G. Moser and S. B. Serpico. *Unsupervised change detection from multichannel SAR data by Markovian data fusion*. *IEEE Trans. Geosci. Remote Sens.*, **47**(7): 2114–2128, July 2009.

[Moser et al., 2006a] G. Moser, J. Zerubia and S. B. Serpico. *Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation*. *IEEE Trans. Geosci. Remote Sens.*, **44**(1): 188–200, January 2006a.

[Moser et al., 2006b] G. Moser, J. Zerubia and S. B. Serpico. *SAR amplitude probability density function estimation based on a generalized Gaussian model*. *IEEE Trans. Image Process.*, **15**(6): 1429–1442, June 2006b.

[Mott, 2007] H. Mott. *Remote Sensing with Polarimetric Radar*. John Wiley & Sons, Hoboken, USA, 2007.

[Nair, 1939] U. S. Nair. *The application of the moment function in the study of distribution laws in statistics*. *Biometrika*, **30**(3/4): 273–294, January 1939.

[Nelson, 1995] D. Nelson. *The Mellin-wavelet transform*. In *Proc. IEEE Int. Conf. Acoustics Speech Signal Process., ICASSP'95*, pp. 1101–1104. Detroit, USA, 8-12 May 1995.

[Nicolas, 2002] 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*, **19**(3): 139–167, 2002. In French.

[Nicolas, 2003] J.-M. Nicolas. *A Fisher-MAP filter for SAR image processing*. In *Proc. IEEE Int. Geosc. Remote Sens. Symp., IGARSS'03*, volume 3, pp. 1996–1998. Toulouse, France, 21-25 July 2003.

[Nicolas, 2006] J.-M. Nicolas. *Application de la transformée de Mellin: Étude des lois statistiques de l'imagerie cohérente*. Technical Report 2006D010, Ecole Nationale Supérieure des Télécommunications, Paris, France, 2006. In French.

[Oliver and Quegan, 2004] C. Oliver and S. Quegan. *Understanding Synthetic Aperture Radar Images*. SciTech Publishing, Raleigh, USA, second edition, 2004.

[Oliver, 2000] C. J. Oliver. *Rain forest classification based on SAR texture*. *IEEE Trans. Geosci. Remote Sens.*, **38**(2): 1095–1104, March 2000.

[Ovarlez et al., 1992] J.-P. Ovarlez, J. Bertrand and P. Vertrand. *Computation of affine time-frequency distributions*. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process., ICASSP'92*, pp. 2909–2912. San Francisco, USA, 23-26 March 1992.

[Poularikas, 1999] A. D. Poularikas, ed. *The Handbook of Formulas and Tables for Signal Processing*, chapter 18. CRC Press, Boca Raton, USA, 1999.

[Rignot and Chellappa, 1993] E. Rignot and R. Chellappa. *Maximum a posteriori classification of multifrequency, multilook, synthetic aperture radar intensity data*. *J. Opt. Soc. Am. A*, **10**(4): 573–582, April 1993.

[Ruffing and Fleischer, 1985] B. Ruffing and J. Fleischer. *Spectral correlation of partially or fully developed speckle patterns generated by rough surfaces*. *J. Opt. Soc. Am. A*, **2**(10): 1637–1643, October 1985.

[Ruth and Gilbert, 1994] D. M. Ruth and J. E. Gilbert. *The Mellin transform in signal analysis*. Technical Report ARL-TR-94-4, University of Texas at Austin, Austin, USA, March 1994.

[Séry and Lopès, 1997] F. Séry and A. Lopès. *Statistical properties of speckle and full polarimetric filters in SAR*. In *Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS'97*, volume 2, pp. 761–763. Singapore, 3-8 August 1997.

[Solbø and Eltoft, 2004] S. Solbø and T. Eltoft. *Homomorphic wavelet-based statistical despeckling of SAR iimages*. *IEEE Trans. Geosci. Remote Sens.*, **42**(4): 711–721, April 2004.

[Springer and Thompson, 1966] M. D. Springer and W. E. Thompson. *The distribution of products of independent random variables*. *SIAM J. Appl. Math.*, **14**(3): 511–526, May 1966.

[Springer and Thompson, 1970] M. D. Springer and W. E. Thompson. *The distribution of products of beta, gamma and Gaussian random variables*. *SIAM J. Appl. Math.*, **18**(4): 721–737, June 1970.

[Stacy, 1962] E. W. Stacy. *A generalization of the gamma distribution*. *Ann. Math. Statist.*, **33**(3): 1187–1192, September 1962.

[Stacy and Mihram, 1965] E. W. Stacy and G. A. Mihram. *Parameter estimation for a generalized gamma distribution*. *Technometrics*, **7**(3): 349–358, August 1965.

[Subrahmaniam, 1970] K. Subrahmaniam. *On some applications of Mellin transforms: Dependent random variables*. *SIAM J. Appl. Math.*, **19**(4): 658–662, December 1970.

[Szpankowski, 2001] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*, chapter 9. John Wiley & Sons, New York, USA, 2001.

[Tison et al., 2004] 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.*, **42**(10): 2046–2057, October 2004.

[Tuceryan and Jain, 1994] M. Tuceryan and A. K. Jain. *Texture analysis*. In *Handbook of Pattern Recognition & Computer Vision*, edited by C. Chen, L. Pau and P. Wang, chapter 2.1, pp. 235–276. World Scientific Publishing Company, Singapore, 1994.

[Valade and Nicolas, 2004] C. Valade and J.-M. Nicolas. *Homomorphic wavelet transform and new subband statistics models for SAR image compression*. In *Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS'04*, volume 1, pp. 285–288. Anchorage, USA, 20-24 September 2004.

[van den Bos, 1994] A. van den Bos. *A Cramér-Rao lower bound for complex parameters*. *IEEE Trans. Signal Process.*, **42**(10): 2859–2859, October 1994.

[Vasile et al., 2009] G. Vasile, J.-P. Ovarlez and F. Pascal. *Estimation and segmentation in non-Gaussian POLSAR clutter by SIRV stochastic processes*. In *Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS'09*, volume 3, pp. 963–966. Cape Town, South Africa, 12-17 July 2009.

[Wagner et al., 1987] R. F. Wagner, M. F. Insana and D. G. Brown. *Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound*. *J. Opt. Soc. Am. A*, **4**(5): 910–922, May 1987.

[Weisstein, 2010a] E. W. Weisstein. *Beta function*. From *MathWorld*–A Wolfram Web Resource. URL=http://mathworld.wolfram.com/BetaFunction.html, Visited at 1. Feb. 2010, 2010a.

[Weisstein, 2010b] E. W. Weisstein. *Confluent hypergeometric function of the second kind*. From *MathWorld*–A Wolfram Web Resource. URL=http://mathworld.wolfram.com/ConfluentHypergeometricFunctionofthe-SecondKind.html, Visited at 1. Feb. 2010, 2010b.

[Weisstein, 2010c] E. W. Weisstein. *Gamma function*. From *MathWorld*–A Wolfram Web Resource. URL= http://mathworld.wolfram.com/GammaFunction.html, Visited at 1. Feb. 2010, 2010c.

[Weisstein, 2010d] E. W. Weisstein. *Modified Bessel function of the second kind*. From *MathWorld*–A Wolfram Web Resource. URL=http://mathworld.wolfram.com/ModifiedBesselFunctionoftheSecond-Kind.html, Visited at 1. Feb. 2010, 2010d.

[Weisstein, 2010e] E. W. Weisstein. *Polygamma function*. From *MathWorld*–A Wolfram Web Resource. URL=http://mathworld.wolfram.com/Polygamma-Function.html, Visited at 1. Feb. 2010, 2010e.

[Yamaguchi et al., 2005] Y. Yamaguchi, T. Moriyama, M. Ishido and H. Yamada. *Four-component scattering model for polarimetric SAR image decomposition*. *IEEE Trans. Geosci. Remote Sens.*, **43**(8): 1699–1706, August 2005.

[Yu, 1998] Y. Yu. *Textural-partially correlated polarimetric K-distribution*. In *Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS'98*, volume 4, pp. 2098–2100. Seattle, USA, 6-10 July 1998.

[Zou et al., 2000] Q. Zou, Y. Pi and G. Liu. *Polarimetric speckle reduction by using multi-texture maximum likelihood method*. In *Proc. IEEE Int. Geosci. Remote Sens. Symp., IGARSS'00*, volume 5, pp. 2287–2289. Honolulu, USA, 24-28 July 2000.

About the author — *Stian Normann Anfinsen* was born in 1975 in Bodø, Norway, and grew up in the town of Harstad. In November 1998, he received the M.Sc. degree in communications, control and digital signal processing from the University of Strathclyde, Glasgow, UK. In March 2000, he graduated with the degree of Cand.Scient. (M.Sc. equivalent) in physics from the University of Tromsø, Tromsø, Norway. He worked as a software engineer for satellite data ground station provider Kongsberg Spacetec AS in Tromsø from 2001 to 2005. He then took up the position as Research Fellow at the University of Tromsø, Department of Physics and Technology, working towards the Ph.D. degree. In May 2010, he defended the current Ph.D. thesis, entitled "Statistical Analysis of Multilook Polarimetric Radar Images with the Mellin Transform". The content reflects his research interest, which are in the areas of statistical signal processing, pattern recognition, earth observation and satellite remote sensing.

ISBN 978-82-8236-017-3