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 $\boldsymbol{\gamma}(\sigma, L)$:

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

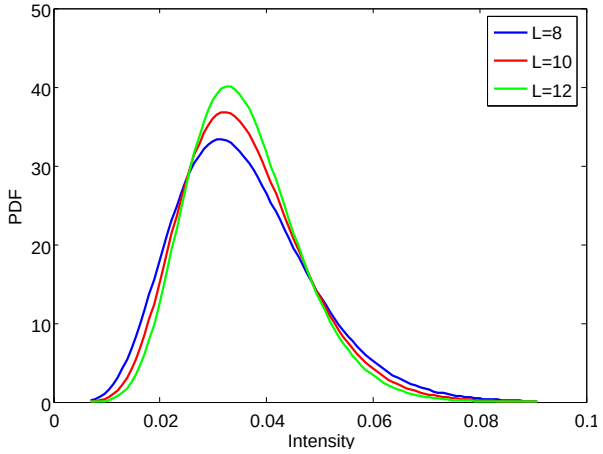

Fig. 1. Example of gamma distribution $\gamma(\sigma, L)$ parametrised with mean intensity $\sigma = 0.0358$ and number of looks $L = \{8, 10, 12\}$.

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.

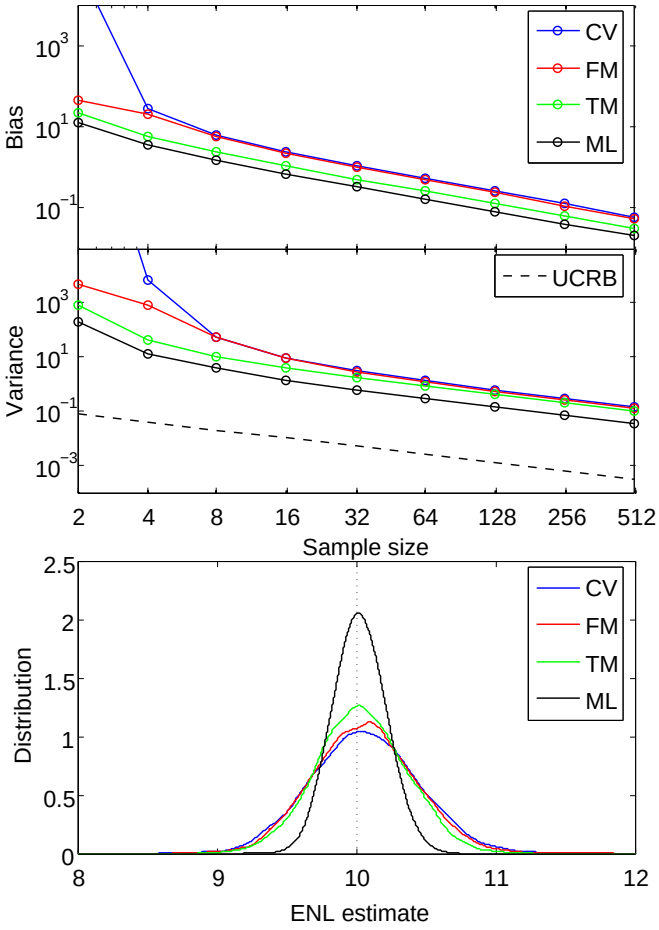

Fig. 2. Estimator bias (top) and variance (middle) as a function of sample size N_b, and the distribution of ENL estimates for $N_b = 512$ (bottom). Results shown for the CV, FM, TM, and ML estimator. The variance plot includes the unbiased Cramér-Rao Bound (UCRB). True $L = 10$ shown as dotted line.

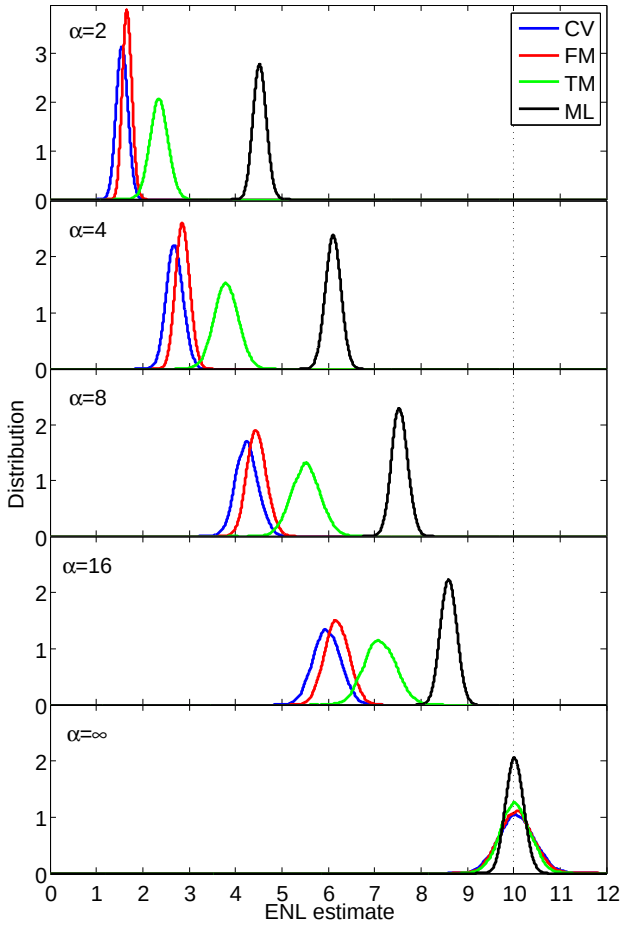

Fig. 3. Distribution estimates for the CV, FM, TM, and ML estimator, calculated from single class multilook polarimetric K-distributed data random generated with a fixed $\mathbf{\Sigma}$ and $\alpha = \{2, 4, 8, 16, \infty\}$.

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-

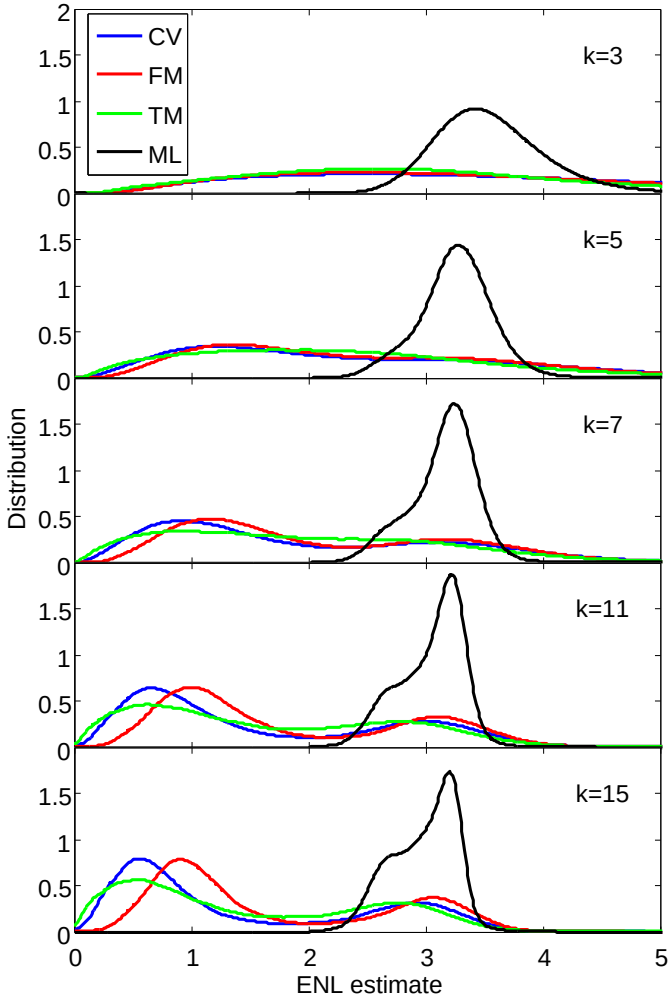

Fig. 4. Distribution estimates for the CV, FM, TM, and ML estimator calculated from the AIRSAR image of Flevoland. No speckle filter applied. ENL estimated for window sizes of $k = \{3, 5, 7, 11, 15\}$.

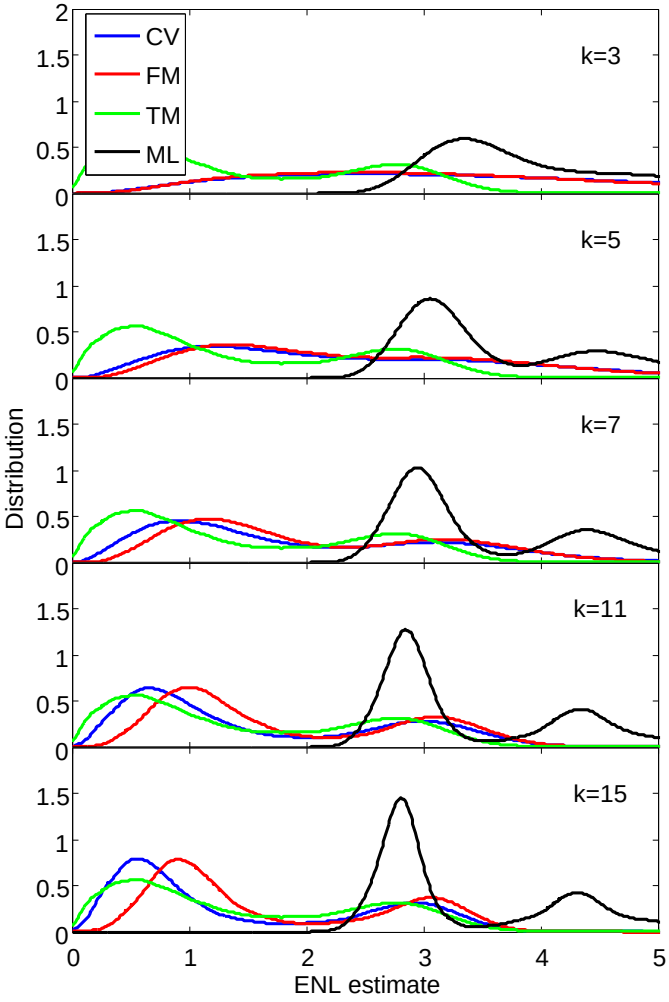

Fig. 5. Distribution estimates for the CV, FM, TM, and ML estimator calculated from the AIRSAR image of San Francisco. No speckle filter applied. ENL estimated for window sizes of $k = \{3, 5, 7, 11, 15\}$.

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

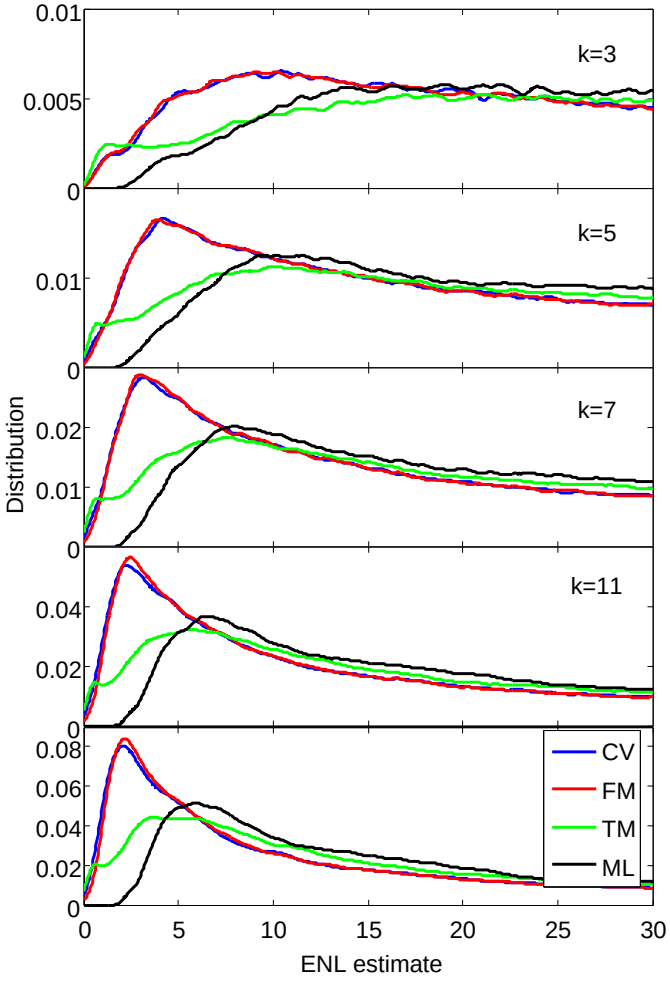

Fig. 6. Distribution estimates for the CV, TM, TM, and ML estimator calculated from the AIRSAR image of Flevoland. Modified Lee filter with window size 7×7 applied. ENL estimated for window sizes of $k = \{3, 5, 7, 11, 15\}$.

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

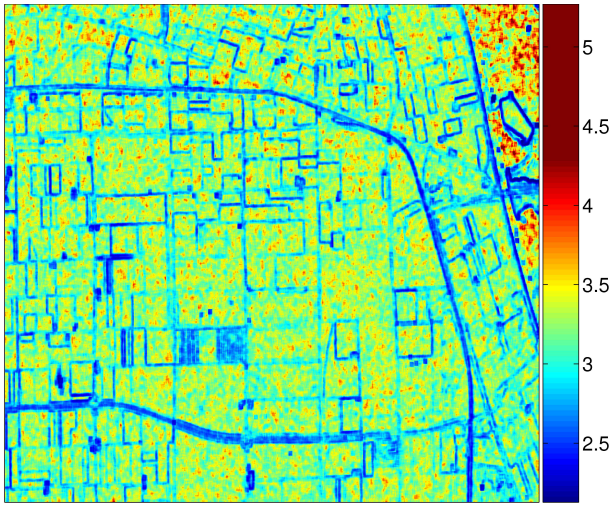

Fig. 7. Local ENL estimates obtained with the ML estimator and window size $k=7$ for the AIRSAR image of Flevoland. No speckle filter applied.

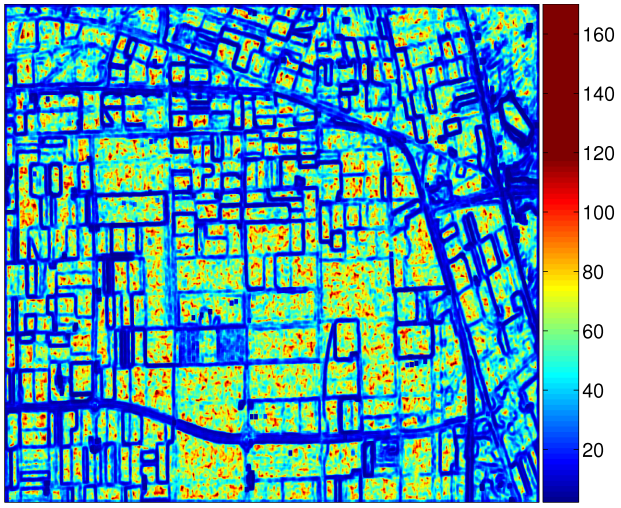

Fig. 9. Local ENL estimates obtained with the ML estimator and window size $k=7$ for the AIRSAR image of Flevoland. Modified Lee filter with window size $\ell=7$ applied.

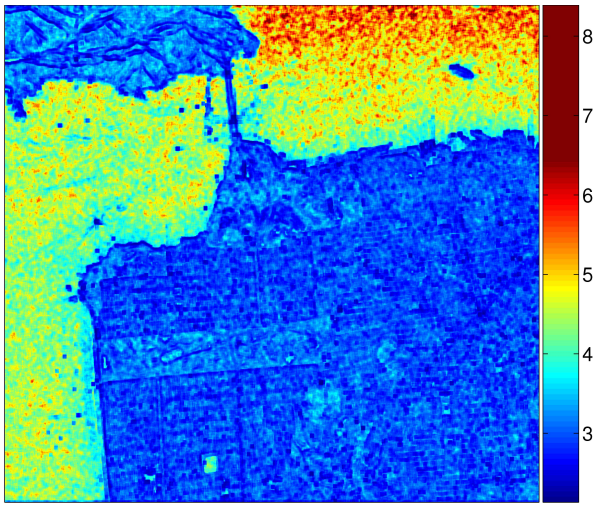

Fig. 8. Local ENL estimates obtained with the ML estimator and window size $k=7$ for the AIRSAR image of San Francisco. No speckle filter applied.

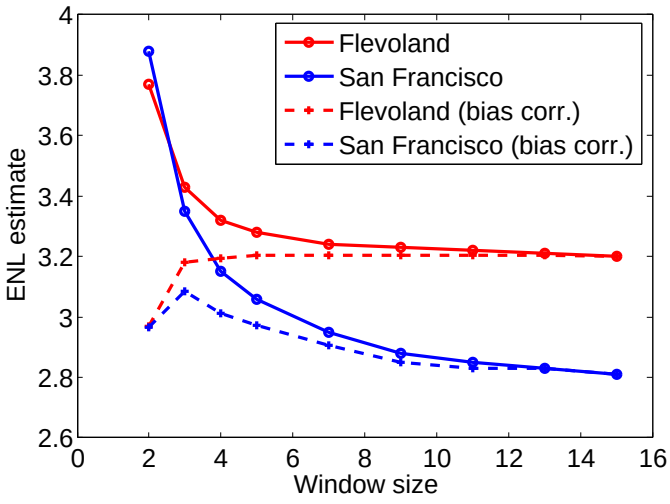

Fig. 10. ENL estimates obtained with the ML estimator as a function of window size k for the AIRSAR images of Flevoland and San Francisco, with and without bias correction. No speckle filter applied.

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.

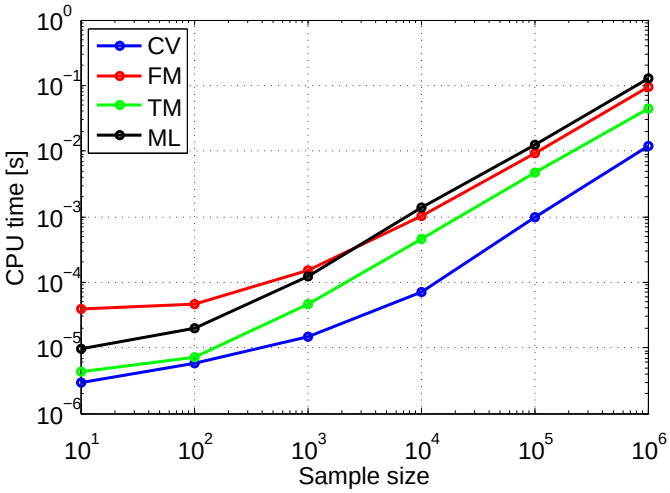

Fig. 11. Computational complexity of the CV, FM, TM, and ML estimator measured in CPU time per estimate calculation as function of sample size.

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.