The Robust InSAR Optimization Framework with Application to Monitoring Cities on Volcanoes

Yuanyuan Wang [(1)], Xiao Xiang Zhu [(1,2),*]

(1) Helmholtz Young Investigators Group "SiPEO"
Technische Universität München
Arcisstraße 21, 80333 Munich, Germany
wang@bv.tum.de

(2) Remote Sensing Technology Institute (IMF),
German Aerospace Center (DLR),
82234 Weßling, Germany. xiao.zhu@dlr.de

Abstract— **This paper introduces the Robust InSAR Optimization (RIO) framework to the multi-pass InSAR techniques, such as PSI, SqueeSAR and TomoSAR whose current optimal estimators were derived based on the assumption of Gaussian distributed stationary data, with seldom attention towards their robustness.**

The RIO framework effectively tackles two common problems in the multi-pass InSAR techniques: 1. treatment of images with bad quality, especially those with large uncompensated phase error, and 2. the covariance matrix estimation of non-Gaussian and non-stationary distributed scatterer (DS). The former problem is dealt with using a robust M-estimator which effectively down-weight the images that heavily violate the phase model, and the latter is addresses with a new method: the *Rank M-Estimator* (RME) by which the covariance is estimated using the *rank* of the DS. RME requires no flattening/estimation of the interferometric phase, thanks to the property of mean invariance of *rank*. The robustness of RME is achieved by using an M-estimator, i.e. amplitude-based weighing function in covariance estimation. The RIO framework can be easily extended to most of the multi-pass InSAR techniques.

Keywords: robust estimation, M-estimator, rank covariance matrix, D-InSAR, InSAR

I. INTRODUCTION

Human habitats in volcanic regions require constant monitoring. Such task on a continuous basis over large area is only achievable with InSAR methods so far. The common procedure for such non-urban area monitoring is exploiting both persistent scatterer (PS) [1] and distributed scatterer (DS) [2]–[4]. In general, millimeter accuracy of yearly linear deformation rate can be achieved.

However, such accuracy refers to the optimal estimators derived based on the assumption of Gaussian distributed data, i.e. PS is modeled as deterministic signal with additive complex circular Gaussian (CCG) noise of independent and identical distribution (i.i.d.), and DS is modeled as correlated zero mean CCG signal [5]. They are well justified in many cases, but in volcanic regions we may encounter fickle atmospheric phase delay and low number of images. They introduce non-Gaussian observation noises and outlier DS samples in the following way:

• Unpredictable atmospheric phase increase the difficulty of their estimation using current PS-based and weather data-based approaches. Uncompensated atmospheric phase introduce additional error to the phase signals of PS and DS.

• Outlier DS sample pixels become harder to detect using statistical tests, e.g. Kolmogorov–Smirnov (KS) test, as the number of coherent images decreases. The outliers considerably bias the covariance matrix estimation of DS. Furthermore, their stationarity cannot be guaranteed, due to spatially varying phase.

Obviously, the performance of the current techniques largely depends on the pre-processing step of atmospheric phase estimation and neighbourhood sample selection if DS is exploited. The Robust InSAR Optimization (RIO) framework [6] is designed not only but especially for dealing with such data. It introduces the following aspects in the current techniques:

• It replaces the maximum likelihood estimator (MLE) for nominal Gaussian distribution which minimizes the sum of the squared residuals with an M-estimator [7] which minimizes the sum of a customized function $\rho(x)$ of the residuals:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_i \rho\left\{\varepsilon_i\left(\mathbf{g} \mid \boldsymbol{\theta}, M\left(\hat{\mathbf{C}}\right)\right)\right\}, \tag{1.1}$$

where $\boldsymbol{\varepsilon}$ is the residual vector as a function of the observations vector $\mathbf{g}$, the parameters $\boldsymbol{\theta}$ to be retrieved and the assumed system model M with estimated covariance matrix $\hat{\mathbf{C}}$ (if used).

• Should DS be exploited, $\hat{\mathbf{C}}$ is replaced with the robust $\hat{\mathbf{C}}_{RME}$ proposed in this paper. $\hat{\mathbf{C}}_{RME}$ is robust against both outlier and non-stationary samples.

II. M-ESTIMATOR FOR PHASE PARAMETERS RETRIEVAL

A. M-estimator basics

M-estimators are a class of well-known robust estimators. It stands for MLE-type estimator, which allows minimizing a customized loss function $\rho(x)$ of the residuals to resist outliers without pre-processing the data such as outlier trimming. Let $f_{\mathbf{g}}(\mathbf{g})$ be a generic likelihood function of $\mathbf{g}$. Choosing $\rho = -\ln f_{\mathbf{g}}(\mathbf{g})$ gives the MLE for $f_{\mathbf{g}}(\mathbf{g})$. The MLE under the Gaussian distribution assumption corresponds to an M-estimator with $\rho(x) = x^2$. For a linear system, its M-estimator can in general be solved by iteratively re-weighted least square, with the weights of each observation being

$$w(\varepsilon_i) = \rho'(\varepsilon_i)/\varepsilon_i . \tag{2.1}$$

978-1-4799-6652-3/15/$31.00~

The M-estimator is a trade-off between efficiency and robustness. However, it can still maintain high efficiency under the nominal model by properly choosing the loss function.

B. M-estimator for PS parameters retrieval

One of the commonly used estimators for PS phase history parameters $\boldsymbol{\theta}$ is the periodogram [1]:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \left| \frac{1}{N} \sum_{n=1}^{N} g_n \exp\left(-j\varphi_n(\boldsymbol{\theta})\right) \right| \right\} \tag{2.2}$$

The g_n and φ_n are the complex pixel value and the modeled phase of the nth image, respectively. Usually the amplitude of g_n is dropped in the estimation. Although not explicitly stated in the original paper [1], Equation (2.2) is actually the MLE under the assumption of additive i.i.d. CCG noise [8].

Assuming Gaussian noise for PS is well justified. However, the uncompensated phase error especially atmospheric phase renders the PS no longer Gaussian. Therefore, we propose the following estimator to deal with possible large phase error:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \rho\left(\mathrm{Re}\left[\varepsilon_i(\boldsymbol{\theta})\right]/\sigma_R\right) + \rho\left(\mathrm{Im}\left[\varepsilon_i(\boldsymbol{\theta})\right]/\sigma_I\right), \tag{2.3}$$

where the residual $\varepsilon_i(\boldsymbol{\theta})$ is $g_i - \exp\left(-j\varphi_i(\boldsymbol{\theta})\right)$, and $\mathrm{Re}[\bullet]$, $\mathrm{Im}[\bullet]$ are the real and imaginary parts operators. Due to the random and impulsive phase error which changes the distribution of real and imaginary part differently, the circularity of the complex distribution is no longer assumed, i.e. the standard deviations σ_R and σ_I are not necessarily identical.

An appropriate $\rho(x)$ function should be chosen according to the noise distribution after atmospheric and other phase correction, which depends on the performance of the phase correction steps. If such information is not available, a good alternative is the integral of the Tukey's biweight function:

$$\rho(x) = \begin{cases} -\left(c^2 - x^2\right)^3 / 6c^4 + c^2/6, & |x| < c \\ c^2/6, & \text{elsewhere} \end{cases}, \tag{2.4}$$

where $c \in \mathbb{R}^+$ is a tuning parameter beyond which the loss is constant. The lower the value of c, the more robust the estimator, and vice versa. Choosing c = 4.586 gives 95% efficiency at normal distribution [9].

Equation (2.3) is solved iteratively, where σ_R and σ_I are updated each iteration. Their initial values are obtained from an initial solution of $\boldsymbol{\theta}$ which is not critical for convex $\rho(x)$. However, a robust initial estimation is recommended for non-convex $\rho(x)$ like the integral of Tukey's biweight which gives multiple minima. The initial estimation, however, does not need to be efficient. To name one, the least trimmed square which minimizes the sum of k smallest squared residuals. The initial estimates of the standard deviations as well as at each iteration should be obtained using a robust estimator, for example the median absolute deviation:

$$\hat{\sigma} = 1.483 \bullet \mathrm{median}\left(\left|\boldsymbol{\varepsilon} - \mathrm{median}(\boldsymbol{\varepsilon})\right|\right). \tag{2.5}$$

With such combination, convergence is achieved just in a few iterations.

C. M-estimator for DS parameters retrieval

If stationarity is assumed for the DS and its neighbourhood, one can treat the DS neighbourhood as a single PS by averaging all the DS pixels in the neighbourhood, like SqueeSAR. The robustified estimator is simply the same as Equation (2.3).

However, stationarity is not assumes in our considerations. We aim at a full inversion of each DS pixel. The original MLE introduced in [10] is recalled here:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \mathbf{g}^H \boldsymbol{\Phi}(\boldsymbol{\theta}) \left|\hat{\mathbf{C}}\right|^{-1} \boldsymbol{\Phi}(\boldsymbol{\theta})^H \mathbf{g} \right\}, \tag{2.6}$$

where $\boldsymbol{\Phi}(\boldsymbol{\theta})$ is the diagonal matrix containing the modelled phase of $\mathbf{g}$, and $|\bullet|$ is the matrix element-wise absolute value operator. Seemingly, the robustified estimator for DS phase history parameters follows the same expression as Equation (1.1), except the residuals $\boldsymbol{\varepsilon}(\boldsymbol{\theta})$ is whitened with a covariance matrix estimate, better with a robust version, such as the $\hat{\mathbf{C}}_{RME}$ which will be covered in Section III:

$$\boldsymbol{\varepsilon}(\boldsymbol{\theta}) = \left|\hat{\mathbf{C}}\right|^{-1/2} \boldsymbol{\Phi}(\boldsymbol{\theta})^H \mathbf{g}. \tag{2.7}$$

The robustified DS estimator can be written into an iteratively reweighted expression with k being the iteration index:

$$\hat{\boldsymbol{\theta}}_{k+1} = \arg\min_{\boldsymbol{\theta}} \left\{ \boldsymbol{\varepsilon}^H(\boldsymbol{\theta}) \mathbf{W}\left(\boldsymbol{\varepsilon}\left(\hat{\boldsymbol{\theta}}_k\right)\right) \boldsymbol{\varepsilon}(\boldsymbol{\theta}) \right\}, \tag{2.8}$$

where the diagonal weighting matrix $\mathbf{W}$ is calculated using Equation (2.1) based on the previous estimates.

However, one should be aware that the marginal distribution of the phase of zero mean CCG multivariate is uniform, as well as its whitened version. The atmospheric phase, presumably also uniformly distributed over time, applies no change to the DS observations statistically. In another word, the robust loss function is blind to such phase contamination on a single-look DS observation.

Therefore, the corrected weighting on the contaminated observations has to be introduced to the estimator. The weights should be calculated posteriorly, i.e. calculated based on the expected residuals $\bar{\boldsymbol{\varepsilon}}$ of the whole DS neighbourhood. The expected residual must be robustly estimated, due to the possible outliers in the neighbourhood:

$$\bar{\varepsilon}_i = \left(\sum_{m=1}^{M} w_i^m \right)^{-1} \sum_{m=1}^{M} w_i^m \varepsilon_i\left(\hat{\boldsymbol{\theta}}^m\right), \tag{2.9}$$

where the superscript m denotes the sample number in the neighbourhood, and w_i^m is a robust weight, e.g. Tukey's biweight. The complete final estimator should be as follows:

$$\hat{\boldsymbol{\theta}}_{k+1}^m = \arg\min_{\boldsymbol{\theta}^m} \left\{ \boldsymbol{\varepsilon}^H\left(\boldsymbol{\theta}^m\right) \mathbf{W}\left(\bar{\boldsymbol{\varepsilon}}_k\right) \boldsymbol{\varepsilon}\left(\boldsymbol{\theta}^m\right) \right\}. \tag{2.10}$$

Its computation should start with a selected DS neighbourhood which jointly determines a single weighting matrix. This very same matrix is used for the parameters retrieval of each single-look DS observation vector in the neighbourhood. The weighting matrix is then updated according to all the estimates in the neighbourhood.

III. THE RANK M-ESTIMATOR OF COVARIANCE

A. Rank basics

Using ranks of the samples instead the samples themselves show its robustness in assessing the statistical dependence between two random variables, and more importantly, its invariance to the non-linear relation between the two random variables. For univariate real random variable, rank refers to the integer ranking of each realization of this variable. In [11], the centered rank of the sample x_m in the univariate data set $\{x_1, x_2, ..., x_M\}$ is defined as $\hat{r}_m = \frac{1}{M}\sum_{m=1}^{M} \text{sign}(x_i - x_m)$, where sign($x$) = -1,0, 1 for x<0, x=0, and x>0. This is can be easily extended to the real multivariable case:

$$\hat{\mathbf{r}}_m = \frac{1}{M}\sum_{m=1}^{M} \text{sign}(\mathbf{x}_i - \mathbf{x}_m) = \frac{1}{M}\sum_{m=1}^{M} (\mathbf{x}_i - \mathbf{x}_m)/\|\mathbf{x}_i - \mathbf{x}_m\| \tag{2.11}$$

for the sample set $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\}$ where $\|\bullet\|$ is the Euclidean norm, and $\mathbf{x}_m$ is a vector of observations such as a time series of pixels in multi-pass InSAR techniques, despite they are not complex number here.

B. Rank of SAR complex multivariate

Equation (2.11) defines the rank of additive real random variable which is not directly useful for the SAR complex multivariate with multiplicative phase signal. Therefore, we define the rank for SAR pixels as follows:

$$\hat{\mathbf{r}}_m = \frac{1}{J}\sum_{j=1}^{J} \mathbf{g}_m \bullet \mathbf{g}_j^* / \left\| \mathbf{g}_m \bullet \mathbf{g}_j^* \right\|, \tag{2.12}$$

where $\mathbf{g}_j$ is the direct neighbourhood of $\mathbf{g}_m$, and the $\bullet$ denotes the element-wise product. Note its difference to the original definition where the subtraction is replaced by multiplication of complex conjugate. For this reason, the deterministic phase has disappeared.

C. Rank M-estimator of covariance

The RME is the M-estimate of the rank covariance matrix. Based on the previous section, we can define the rank covariance matrix which is simply the sample covariance matrix calculated using the rank of each sample of $\mathbf{g}$:

$$\hat{\mathbf{C}} = \frac{1}{M}\sum_{m=1}^{M} \hat{\mathbf{r}}_m \hat{\mathbf{r}}_m^H . \tag{2.13}$$

The M-estimate of the above equation, i.e. RME of the covariance, is the iteratively reweighted version of itself [12]:

$$\hat{\mathbf{C}}_{k+1\,RME} = \frac{1}{M}\sum_{m=1}^{M} w\left(\varepsilon_{\substack{m \\ k}}\right) \hat{\mathbf{r}}_m \hat{\mathbf{r}}_m^H , \tag{2.14}$$

with the weighting function $w(\square)$ analogous to Equation (2.1), k refers to the iteration index, and the squared residual $\varepsilon_m^2 = \hat{\mathbf{r}}_m^H \hat{\mathbf{C}}^{-1} \hat{\mathbf{r}}_m$ (k dropped for simplicity) [6]. It can be proven that the element wise square root of $\left|\hat{\mathbf{C}}_{RME}\right|$ approaches $\left|\hat{\mathbf{C}}_{MLE}\right|$ asymptotically under CCG for using one direct neighbourhood [6]. Therefore, element-wise square root on $\left|\hat{\mathbf{C}}_{RME}\right|$ is required in prior of using it.

IV. ROBUSTNESS ANALYSIS

A. Comparison of robust covariance estimators

The original MLE and the RME are compared under three different scenarios: 1. multivariate CCG, 2. multivariate complex t-distribution (CT), and 3. non-stationary multivariate CT. The data are simulated to have ten acquisitions, with each acquisition being 1000 samples. Each sample vector is simulated by de-whitening a ten-variate zero mean i.i.d. CCG (for scenario 1) or CT (for scenario 2, 3) vector with a predefined coherence matrix. In the third scenario, linear fringes with ten different fringe frequencies randomly picked within [0 π/10] are added to the phases of the ten acquisitions, respectively.

We also predefined two different coherence matrices, one exponentially decaying, and the other constant coherence of 0.5 between acquisitions. That corresponds to the subfigure (a) and (b) of Figure 1. Please refer to Figure 1 for its detailed description. The subplots (1, 1) in both subfigures shall be regarded as their reference coherence matrices, respectively, because the MLE is the optimal estimator under CCG.

Both estimators give correct estimate under CCG. The RME has minor fluctuation at low coherence due to the element-wise square root operation (can be seen in Figure 1(a)), and experiences slightly higher variance (can be seen in Figure 1(b)). The efficiency RME is always lower than 100% at the nominal distribution (CCG). This is a trade-off between robustness and efficiency. The MLE fails at the second scenario, where the samples are contaminated by outliers. The coherence is usually overestimated due to the large amplitude of the outliers. It also failed at the third scenario where it underestimates the covariance due to non-stationary samples. Yet the RME is mean invariant and robust against outlier, which keeps its performance at all conditions.

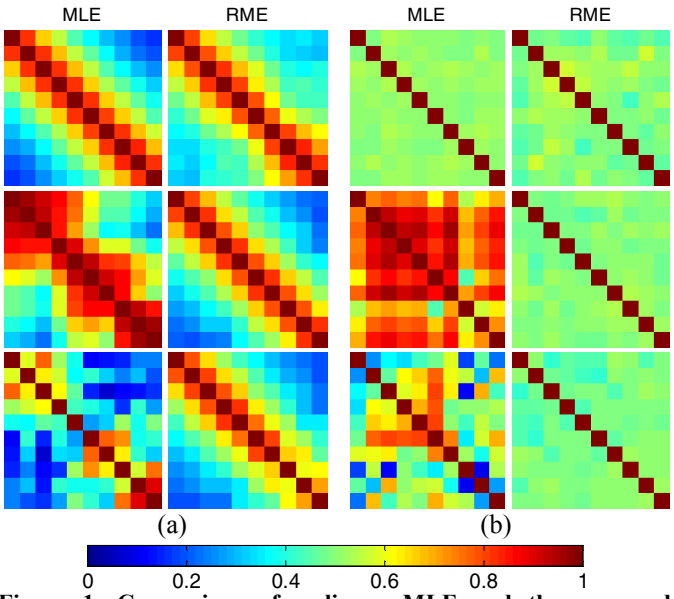

Figure 1. Comparison of ordinary MLE and the proposed covariance RME for three different scenarios: 1st row: complex circular Gaussian, 2nd row: complex t-distribution with one degree of freedom, and 3rd row: non-stationary complex t-distribution with one degree of freedom. Subfigure (a) and (b) correspond to reference covariance matrices being exponentially decaying, and constant, respectively. For either subfigure, 1st

column: MLE, 2nd column: rank M-estimator with t-distribution weighting.

B. Efficiencies of robust PS and DS estimators

The efficiencies of the proposed PS and DS estimators are quantitatively compared with their original MLEs using Monte Carlo simulation. The common parameters for the simulation are set to be similar to those of TerraSAR-X data. The wavelength λ = 0.031 m, and range distance R = 700 km, acquisitions number N = 20, with spatial baselines $\mathbf{b}_\perp$ evenly spaced in [-100 +100] m and temporal baselines **t** randomly sampled in [-1 +1] year with uniform distribution. The phase history parameters are set to be elevation s = 20 m and linear deformation rate v = 15 mm/year.

The PS observation vector is simulated by adding zero mean i.i.d. CCG noise with certain SNR to the deterministic PS signal. The SNR is kept same for each acquisition. A DS vector is simulated using the same procedure described in the previous section, i.e. de-whitening a multivariate zero mean i.i.d. CCG using a coherence matrix. The coherence is kept constant between any two acquisitions. To complete the DS simulation, the deterministic interferometric phase is added to the phases of the DS vector. 1000 samples of such DS vector are generated for covariance matrix estimation. These are the contamination-free PS and DS observations.

Contamination is then introduced by: 1. adding uniformly distributed random phase to 40% of the acquisitions. This applies for both PS and DS. Especially for DS, the same phase is added to the samples of the same acquisition. This simulates the temporally random, spatially stationary atmospheric phase; and 2. using zero mean i.i.d. CT with one degree of freedom in the DS simulation, which generates outliers in DS samples.

In the robust estimators, we use the integral of Tukey's biweight as the loss function. The parameter c is set to 4.586. The Monte Carlo simulation is repeated 10,000 times. Figure 2 shows the standard deviation of the estimates of elevation and linear deformation rate w.r.t. different SNRs. The SNR of a DS refers to the correlated signal w.r.t. the decorrelated part, and can be directly related to the coherence by the formula $\gamma = \left(1 + SNR^{-1}\right)^{-1}$. Please refer to Figure 2 for its detailed description. All the simulation results suggest that the proposed robust estimators greatly outperform the original MLE when as much as 40% the observations are contaminated. Yet the efficiency of the robust estimator at nominal distribution is close to the original MLE. The relative efficiency of the robust estimator and the MLE defined as $\eta_{RE} = \sigma^2_{MLE} / \sigma^2_{Robust}$ at nominal distribution is about 70% for the robust PS estimator, and 80% for the DS estimator. Therefore, the robust estimators are also capable of handling uncontaminated observations, without too much loss of the efficiency.

V. CONCLUSION

This paper introduced an easily extendable robust InSAR optimization framework which encloses the current state-of-the-art estimators with a robust kernel and implants a robust covariance estimator. The new covariance estimator, unlike most current methods, is invariant to the non-zero fringe frequency. The proposed PS and DS estimators are both robust against non-Gaussian outlier and efficient at Gaussian noise.

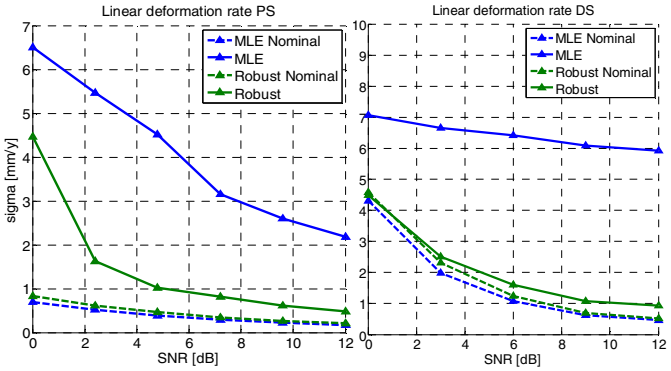

Figure 2. Comparison of the standard deviation of linear deformation rate estimates of the robust estimators and the ordinary MLEs. Subfigure (a) refers to the PS estimators, and (b) is for DS estimators. In each subfigure, the green curves are the results of the proposed robust estimators, and the blue ones are the original MLE; the dashed curves are the estimators performing on observations with nominal distribution, i.e. CCG, and the solid ones are results for contaminated data. The blue dashed curves in all the subfigures are the MLEs under CCG, and hence they are the optimal estimates.

REFERENCES

[1] A. Ferretti, et al., "Permanent scatterers in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, 39(1), pp. 8–20, Jan. 2001.

[2] A. Ferretti, et al., "A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR," *IEEE Trans. Geosci. Remote Sens.*, 49(9), pp. 3460–3470, Sep. 2011.

[3] Y. Wang, et al., "Retrieval of Phase History Parameters from Distributed Scatterers in Urban Areas Using Very High Resolution SAR Data," *ISPRS J. Photogramm. Remote Sens.*, 73, pp. 89–99, Sep. 2012.

[4] G. Fornaro, et al., "CAESAR: An Approach Based on Covariance Matrix Decomposition to Improve Multibaseline-Multitemporal Interferometric SAR Processing," *IEEE Trans. Geosci. Remote Sens.*, 53(4), pp. 2050–2065, Apr. 2015.

[5] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inverse Probl.*, vol. 14, no. 4, p. R1, 1998.

[6] Y. Wang and X. X. Zhu, "Robust Estimators in Multi-pass SAR Interferometry," *IEEE Trans. Geosci. Remote Sens.*, 2014.

[7] P. J. Huber, *Robust Statistics*. John Wiley & Sons, 1981.

[8] D. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *Inf. Theory IEEE Trans. On*, vol. 20, no. 5, pp. 591–598, 1974.

[9] W. N. Venables and B. D. Ripley, *Modern applied statistics with S*. Springer, 2002.

[10] F. De Zan and F. Rocca, "Coherent processing of long series of SAR images," in *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International*, 2005, vol. 3, pp. 1987–1990.

[11] C. Croux, et al., "Sign and rank covariance matrices: statistical properties and application to principal components analysis," in *Statistical data analysis based on the L1-norm and related methods*, Springer, 2002, pp. 257–269.

[12] E. Ollila and V. Koivunen, "Influence functions for array covariance matrix estimators," in *Statistical Signal Processing, 2003 IEEE Workshop on*, 2003, pp. 462–465.