Parameter Selection Criteria for Tomo-SAR Focusing

Gustavo Daniel Martín-del-Campo-Becerra, Sergio Alejandro Serafín-García, Andreas Reigber, *Fellow, IEEE*, and Susana Ortega-Cisneros

***Abstract*—The synthetic aperture radar (SAR) tomography (TomoSAR) inverse problem is commonly tackled in the context of the direction-of-arrival estimation theory. The latter allows achieving super resolution, along with ambiguity levels reduction, thanks to the use of parametric focusing methods, as multiple signal classification (MUSIC) and statistical regularization techniques, like the maximum-likelihood-inspired adaptive robust iterative approach (MARIA). Nevertheless, in order to correctly suit the considered signal model, MUSIC and most regularization approaches require an appropriate setting of the involved parameters. In both cases, the accuracy of the retrieved solutions depends on the right selection of the assigned values. Thus, with the aim of dealing with such an issue, this article addresses several parameter selection strategies, adapted specifically to the TomoSAR scenario. Parametric techniques as MUSIC solve the TomoSAR problem in a different manner as the regularization methods do, hence, each approach demands different methodologies for the proper estimation of their parameters. Consequently, we refer to the Kullback–Leibler information criterion for the model order selection of parametric techniques as MUSIC, whereas we rather explore the Morozov's discrepancy principle, the L-Curve, the Stein's unbiased risk estimate, and the generalized cross-validation to choose the regularization parameters. After the incorporation of these criteria to MUSIC and MARIA, respectively, their capabilities are first analyzed through simulations, and later on, utilizing real data acquired from an urban area.**

***Index Terms*—Generalized cross-validation, information criteria, L-Curve, maximum likelihood (ML), model order selection (MOS), synthetic aperture radar (SAR) and tomography (TomoSAR).**

NOMENCLATURE

List of Acronyms

AIC	Akaike information criterion.
BIC	Bayesian information criterion.
BMR	Bayes minimum risk.
CLS	Constrained least squares.
DOA	Direction of arrival.
ESPRIT	Estimation of signal parameters by rotational invariance techniques.
FD	Fréchet distance.
GCV	Generalized cross-validation.
KL	Kullback–Leibler.
MARIA	ML-inspired adaptive robust iterative approach.
MAP	Maximum a posteriori probability.
ML	Maximum likelihood.
MOS	Model order selection.
MSE	Mean squared error.
MUSIC	Multiple signal classification.
SAR	Synthetic aperture radar.
TomoSAR	SAR tomography.
TSVD	Truncated singular value decomposition.
PCA	Principal component analysis.
PLOS	Perpendicular to the line-of-sight.
PSP	Power spectrum pattern.
RMSE	Root MSE.
ROI	Region of interest.
SLC	Single-look complex.
SNR	Signal-to-noise ratio.
SURE	Stein's unbiased risk estimate.

Glossary of Notation

$\langle\cdot\rangle$	Averaging operator.
$\mathbf{D}(\mathbf{u})$	Diagonal matrix with vector $\mathbf{u}$ at the principal diagonal.
$\|\cdot\|$	Euclidean ℓ_2-norm.
$\mathrm{E}()$	Expectation operator.
$^+$	Hermitian conjugate (adjoin).
$\mathbf{I}$	Identity matrix.
$\{\mathbf{U}\}_{\mathrm{diag}}$	Main diagonal of matrix $\mathbf{U}$.
$\ln\{\cdot\}$	Natural logarithm.
$^{\mathrm{T}}$	Transpose.
$\mathrm{tr}\{\mathbf{U}\}$	Trace of matrix $\mathbf{U}$.

Manuscript received August 10, 2020; revised October 31, 2020; accepted November 30, 2020. Date of publication December 4, 2020; date of current version January 8, 2021. *(Corresponding author: Gustavo Daniel Martín-del-Campo-Becerra.)*

Gustavo Daniel Martín-del-Campo-Becerra and Andreas Reigber are with the Microwaves and Radar Institute, German Aerospace Center (DLR), 82230 Oberpfaffenhofen, Germany (e-mail: gustavo.martindelcampobecerra@dlr.de; andreas.reigber@dlr.de).

Sergio Alejandro Serafín-García and Susana Ortega-Cisneros are with the Center for Research and Advanced Studies (Cinvestav), National Polytechnic Institute (IPN), 45019 Zapopan, Jalisco, Mexico (e-mail: sergio.serafin@cinvestav.mx; susana.ortega@cinvestav.mx).

Digital Object Identifier 10.1109/JSTARS.2020.3042661

I. INTRODUCTION

WITHIN the context of the DOA estimation theory [1, Ch. 6], [2], the TomoSAR inverse problem is typically described via the linear equation of observation [3]–[6]

$$\begin{bmatrix}\mathbf{y}\end{bmatrix}_{L\times 1} = \begin{bmatrix}\mathbf{A}\end{bmatrix}_{L\times M}\begin{bmatrix}\mathbf{s}\end{bmatrix}_{M\times 1} + \begin{bmatrix}\mathbf{n}\end{bmatrix}_{L\times 1}. \quad (1)$$

The TomoSAR acquisition constellation is composed by L passes (flight tracks), each one with a different line of sight. One coregistered SAR image is acquired from each pass; afterward,

the imagery is coherently combined using SAR interferometric techniques. Assuming coregistration independent on height, these L passes are treated as a linear array. Accordingly, for a given azimuth-range position, vector $\mathbf{y}$ represents the set of the corresponding L processed signals; vector $\mathbf{s}$ gathers M samples of the continuous complex random reflectivity, taken at the PLOS elevation positions $\{z_m\}_{m=1}^{M}$; and vector $\mathbf{n}$ accounts for the additive noise. The $L \times M$ steering matrix $\mathbf{A}$ is the signal formation operator that maps $S \rightarrow Y$, the source Hilbert signal space S onto the observation Hilbert signal space Y.

Given an azimuth-range location within the illuminated scene, the main goal of TomoSAR is to retrieve the PSP in the PLOS height direction, depicted in a discrete form through vector $\mathbf{b} = \{b_m\}_{m=1}^{M} = \{\langle |s_m|^2 \rangle\}_{m=1}^{M}$, i.e., the second-order statistics of the complex reflectivity vector $\mathbf{s}$.

Using matched filtering for TomoSAR focusing attains a resolution in the PLOS height direction ρ_{PLOS}, which is constrained to the acquisition geometry [7], [8]. As a rule of thumb, the larger the tomographic aperture D_{Tomo} is, the finer the attainable resolution, as specified by [7, (8)]

$$\rho_{\mathrm{PLOS}} = \frac{\lambda r_1}{2D_{\mathrm{Tomo}}} \tag{2}$$

with λ standing for the carrier wavelength and with r_1 as the slant-range distance to a particular target. Still, the amount and distribution of the several passes must be carefully planned, taking the following into consideration.

1) In order to guarantee strong ambiguity rejection along the PLOS height range of interest, the baseline between passes must be sufficiently small. Assuming evenly distributed passes, the PLOS height of ambiguity V_{PLOS} is defined by [7, (9)]

$$V_{\mathrm{PLOS}} = \frac{\lambda r_1}{2\mathscr{d}} \tag{3}$$

in which $\mathscr{d}$ is the cross-range oriented baseline between passes.

2) Generally, when the passes are not uniformly distributed, it leads to higher ambiguity levels [8], [9].

The desired resolution specifies the length of the tomographic aperture D_{Tomo}, whereas the height of ambiguity V_{PLOS} defines the baseline $\mathscr{d}$ between passes. The baseline $\mathscr{d}$ between passes, necessary to avoid ambiguities, depends on the total PLOS height extent to be examined. Yet, in practical scenarios, the number of passes is constrained to the revisit time that leads to temporal decorrelation issues and to the feasibility of the individual missions to perform the numerous passes, limiting the achievable resolution.

With the aim of easing these limitations, the usage of super resolved DOA-inspired techniques is a common practice. These methods are not constrained to the resolution of the acquisition geometry, as they do not involve a conventional inversion of the spectrum [10]. Particularly, parametric focusing techniques as MUSIC [1], [11], and ESPRIT [1], and iterative statistical regularization methods, like the ones based on ML [3]–[6], [10], improve significantly the resolution in the PLOS height direction, performing suppression of artifacts, and reduction of the ambiguity levels.

In order to correctly suit the considered signal model, MUSIC, ESPRIT, and most regularization approaches (e.g., MARIA [3], [4]) require an appropriate setting of the involved parameters. The right performance of parametric techniques as MUSIC and ESPRIT, depends on the proper specification of the model order [1, Appendix C], [12]. This type of methods assumes that the scene is composed by a finite number of point-type like backscattering sources, which is expected to be known *a priori*. Conversely, regularization methods as MARIA entail the right selection of the regularization parameters to assure retrieving well-conditioned solutions.

All aforementioned parametric and regularization approaches must guarantee "good-fitted" reconstructions. A solution is said to be good fitted, when it fits correctly enough with the position and density of the targets within the field backscattered toward the sensor. Consequently, the values assigned to the corresponding parameters must avoid the next issues [13].

1) "Underfitting," meaning that the retrieved PSP may not be well adapted to the particularities of the input signals. Some of the actual targets might be suppressed, as they are taken by noise.
2) "Overfitting," meaning that the residual variation (i.e., noise) is considered as part of the recovered signal. The retrieved PSP may display targets where in reality there are none, causing false detections.

Accordingly, the main contribution of this article is the adaptation of different parameter selection criteria to the TomoSAR scenario, and their subsequent incorporation into the treated parametric and regularization focusing techniques. Parametric methods as MUSIC and ESPRIT solve the TomoSAR problem in a different manner as the regularization methods do, hence, each approach demands different methodologies for the proper estimation of their parameters. Therefore, in the case of such parametric techniques, this work assesses a MOS rule that is based on ML and on the KL information criterion, which is also the basis for tools as AIC and BIC [12]. On the other hand, we explore four different criteria for the right selection of regularization parameters: The Morozov's discrepancy principle [14], the L-Curve [14], [15], SURE [16], [17], and GCV [17], [18]. For demonstration and evaluation purposes, the addressed MOS rule is incorporated into MUSIC [1], [11], whereas we refer to MARIA [3], [4], to analyze the other methodology.

The remaining of the article is organized as follows: The TomoSAR signal model is described in Section II; Section III presents the TomoSAR-adapted MOS rule; Section IV reviews the strategies for choosing the regularization parameters, namely, the Morozov's discrepancy principle, the L-Curve, SURE, and GCV; Section V analyzes the addressed criteria through numerical examples, making use of MUSIC to study the treated MOS rule, and MARIA to study the different tools for selecting the regularization parameters; Section VI presents a comparison between the best versions (according to the chosen parameter selection strategies) of MUSIC and MARIA via experimental results, gotten from an urban test site; finally Section VII concludes this work

II. TomoSAR Signal Model

For each azimuth-range position within the illuminated area, given the data recordings $\mathbf{y} = \{y_l\}_{l=1}^{L}$, the steering matrix $\mathbf{A}$ and some prior knowledge on the problem (e.g., about the statistics of the signal and noise), the nonlinear TomoSAR inverse problem consists in retrieving an estimate $\hat{\mathbf{b}}$ of the actual PSP vector $\mathbf{b} = \{\langle |s_m|^2 \rangle\}_{m=1}^{M}$.

The covariance matrix

$$\mathbf{Y} = \frac{1}{J}\sum_{j=1}^{J} \mathbf{y}_{(j)}\mathbf{y}_{(j)}^{+} \tag{4}$$

is commonly employed in order to increase accuracy in presence of signal-dependent (multiplicative) noise [3], [19, Ch. 18], and in order to handle the multiple nondeterministic sources. Here, J indicates the amount of independent realizations (looks, snapshots) of the signal acquisitions. Recall that TomoSAR is customarily treated as an ergodic process, meaning that its statistical properties are deduced from a single random realization. Thus, multilooking is technically accomplished through the averaging of adjacent values, e.g., via Boxcar filtering.

Matrix $\mathbf{A}$ in (1) gathers M steering vectors, each one of dimension L. The steering vectors $\{\mathbf{a}_m\}_{m=1}^{M}$ contain the interferometric phase information associated to a source located along the PLOS elevation positions $\{z_m\}_{m=1}^{M}$, above the reference focusing plane. For a specified elevation position z, the related steering vector is given by [20], [21],

$$\mathbf{a}(z) = \left[\, 1 \; \exp\{\mathrm{j}k_{z2}z\} \; \cdots \; \exp\{\mathrm{j}k_{zL}z\} \,\right]^{\mathrm{T}}, \tag{5}$$

in which

$$\left\{ k_{zl} = \left(\frac{4\pi}{\lambda}\right)\left(\frac{d_l}{r_1 \sin\theta}\right) \right\}_{l=2}^{L} \tag{6}$$

is the two-way vertical wavenumber between the master track and the lth acquisition position. The slant-range distance to a particular target is defined by r_1, whereas $\{d_l\}_{l=2}^{L}$ is the cross-range oriented baseline between the master position and the lth acquisition position; with λ standing for the carrier wavelength and θ standing for the incidence angle.

The TomoSAR problem is intrinsically ill-conditioned, since it does not accomplish two of the three Hadamard conditions [19, Ch. 15], [22], to be considered as well posed.

1) The *existence* condition is not achieved. The PSP is a continuous function $b(z)$, in which z refers to the PLOS height direction. Since the data recordings $\mathbf{y} = \{y_l\}_{l=1}^{L}$ are finite dimensional, there is not enough information to reconstruct the whole PSP function $b(z)$. Therefore, due to practical reasons, the PSP is rather represented in a discrete form through vector $\mathbf{b} = \{b_m\}_{m=1}^{M}$, constructed using a sufficiently large number of samples M, taken from $b(z)$ at the PLOS height positions $\{z_m\}_{m=1}^{M}$.
2) The *uniqueness* condition is not accomplished. Normally, the number of samples M is much larger than the number of data recordings L. From a limited small quantity of measurements, a large number of samples representing the PSP are to be estimated; this means that an infinite number of solutions to recover $\hat{\mathbf{b}}$ exist. Additionally, the unavoidable presence of additive and signal-dependent noise, along with the inherent imprecisions of any problem model, adds statistical uncertainty.

Hence, by making some appropriate assumptions and/or by imposing some form of constraints, the different focusing techniques must guarantee well-conditioned solutions to the nonlinear TomoSAR inverse problem.

Parametric methods as MUSIC and ESPRIT parameterize vector $\mathbf{b}$. They assume that vector $\mathbf{b}$ is composed by a certain amount of point-type like backscattering sources, usually much smaller than the number of acquisitions L [1, Ch. 5]. In this way, the TomoSAR problem is reduced to the problem of selecting an *integer-valued* parameter, which describes the number of signals impinging on the array [1, Appendix C]. For the case of MUSIC, explained more in detail in Section III, the *integer-valued* parameter (so-called model order) refers to the number of eigenvalues (and the corresponding eigenvectors) of matrix $\mathbf{Y}$ in (4), which belong to the signal subspace.

On the other hand, most regularization techniques replace the ill-conditioned TomoSAR inverse problem with a well-posed optimization problem. For such aim, the regularization methods addressed in this article consider the signal model defined in the other related studies [3]–[6]: The complex random Gaussian zero-mean vectors $\mathbf{n}$, $\mathbf{s}$, and $\mathbf{y}$ in (1) are characterized by their corresponding correlation matrices

$$\mathbf{R}_{\mathbf{n}} = \mathrm{E}\left(\mathbf{n}\mathbf{n}^{+}\right) = N_0\mathbf{I}, \tag{7}$$

$$\mathbf{R}_{\mathbf{s}} = \mathrm{E}\left(\mathbf{s}\mathbf{s}^{+}\right) = \mathbf{D}\left(\mathbf{b}\right), \tag{8}$$

and

$$\mathbf{R}_{\mathbf{y}} = \mathrm{E}\left(\mathbf{y}\mathbf{y}^{+}\right) = \mathbf{A}\mathbf{R}_{\mathbf{s}}\mathbf{A}^{+} + \mathbf{R}_{\mathbf{n}}, \tag{9}$$

where N_0 is the power spectral density of the white noise power [20], and vector $\mathbf{b}$ defines the backscattering power for a specified azimuth-range position, the so-called PSP.

Assuming that the entries of vector $\mathbf{s}$ are uncorrelated, simplifies the mathematical developments that led to MARIA in [3] and [4]. This work refers to the same mathematical expressions; therefore, we make the same assumption. Consequently, matrix $\mathbf{R}_{\mathbf{s}}$ in (8) is modeled as a diagonal matrix, as described by the unconditional model in [21, (10)]. Also, contrary to [23] and [24], the MOS rule presented next in Section III, makes use of the same signal model, specified by the correlation matrix $\mathbf{R}_{\mathbf{y}}$ in (9).

In principle, all regularization techniques addressed in this article, along with the respective parameter selection criteria, are applicable to all those inverse ill-posed problems, described by the signal model stated previously. This means that the scope of this work is not only limited to researchers interested in TomoSAR, as it also includes those researchers interested in signal processing.

III. Model Order Selection

As mentioned before, spectral analysis-based parametric techniques as MUSIC [1], [11], and ESPRIT [1], require the selection of an *integer-valued* parameter (so-called model order) that correctly specifies the data model [12]. For the linear case, the

ML method of parameter estimation is usually employed for this purpose, being the basis of MOS rules as AIC and BIC. Such MOS rules suggest using the log-likelihood function in [1, (C.2.1)]

$$\hat{\boldsymbol{\varphi}} = \underset{\boldsymbol{\varphi}}{\operatorname{argmax}} \left\{ \ln \left\{ p\left(\mathbf{y} \left| \boldsymbol{\varphi} \right.\right) \right\} \right\}. \tag{10}$$

Here, vector $\boldsymbol{\varphi}$ contains the values that correctly specify the data model, utilized to properly recover the data vector $\mathbf{y}$; whereas vector $\hat{\boldsymbol{\varphi}}$ gathers the corresponding estimated values. The dimension of $\boldsymbol{\varphi}$ defines the model order.

However, the log-likelihood function $\ln\{p(\mathbf{y}|\boldsymbol{\varphi})\}$ in (10) is not applicable to the case under study, since the TomoSAR problem is nonlinear, as it consists on retrieving an estimate $\hat{\mathbf{b}}$ of the PSP. Therefore, the ML method of parameter estimation must be properly adapted to the TomoSAR scenario. With this aim, we refer first to the ML solution to the previously stated nonlinear TomoSAR inverse problem, which is the basis of the TomoSAR-adapted MOS rule introduced afterwards.

A. Maximum Likelihood

Vectors $\mathbf{y}$, $\mathbf{s}$, and $\mathbf{n}$ in (1) are customarily modeled as statistical vectors, as they are constructed from physical phenomena with an intrinsic statistical nature. A statistical vector represents, in theory, an infinite number of different realizations of a process. Hence, vector $\mathbf{y}$ is explicitly characterized through its *probability density function* (pdf), given by [3], [10],

$$p\left(\mathbf{y}\right) = \pi^{-L} \det{}^{-1}\left\{\mathbf{R}_{\mathbf{y}}\right\} \cdot \exp\left\{-\left(\mathbf{y}^{+}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{y}\right)\right\}. \tag{11}$$

Recall that $\mathbf{y}$ is a L-dimensional complex random Gaussian zero-mean vector.

Assuming *a priori* knowledge about the pdf of the noise vector $\mathbf{n}$ and the PSP vector $\mathbf{b}$, the MAP approach provides a statistically optimal solution to the TomoSAR inverse problem, via [2, Ch. 8]

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmax}} \left\{ \ln \left\{ p\left(\mathbf{b}|\mathbf{y}\right) \right\} \right\} \tag{12}$$

with

$$\ln\left\{p\left(\mathbf{b}|\mathbf{y}\right)\right\} = \ln\left\{p\left(\mathbf{y}|\mathbf{b}\right)\right\} + \ln\left\{p\left(\mathbf{b}\right)\right\} - \ln\left\{p\left(\mathbf{y}\right)\right\}, \tag{13}$$

according to the Bayes formula, and since $\ln\{\cdot\}$ is a monotonically increasing function.

Nevertheless, for the case under study, the pdf $p(\mathbf{b})$ is generally unknown. Thus, setting $p(\mathbf{b}) \approx const$ and ignoring those terms that do not comprise $\mathbf{b}$ in (13), we rather refer to the log-likelihood function

$$\ln\left\{p\left(\mathbf{y}|\mathbf{b}\right)\right\} = -\ln\left\{\det\left\{\mathbf{R}_{\mathbf{y}}\right\}\right\} - \mathbf{y}^{+}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{y} \tag{14}$$

where the terms that do not contain $\mathbf{b}$ have been ignored. The ML solution to the TomoSAR inverse problem is given by [3], [4],

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmax}} \left\{ \ln \left\{ p\left(\mathbf{y}|\mathbf{b}\right) \right\} \right\}. \tag{15}$$

Consider then that $\mathbf{b} = \mathbf{b}(\boldsymbol{\varphi})$, such that $\mathbf{b}$ entails having a vector $\boldsymbol{\varphi}$ that correctly specifies the data model. In this way, we can make use of the log-likelihood function in (14) to derive the TomoSAR-adapted MOS rule introduced below.

Previous related studies [23, (10)], [24, (14)], propose approximating (14) through the eigendecomposition of the sample covariance matrix $\mathbf{Y}$ in (4); leading to various eigenvalues-based MOS rules, which are based on the KL information criterion [12]. Yet, the novelty of this work is approximating (14) differently.

We refer to the signal model defined in Section II, characterized by the correlation matrix $\mathbf{R}_{\mathbf{y}} = \mathbf{R}_{\mathbf{y}}\left(\mathbf{b}\right) = \mathbf{A}\mathbf{D}(\mathbf{b})\mathbf{A}^{+} + N_0\mathbf{I}$. Consequently, the log-likelihood function in (14) becomes solution dependent. The actual PSP vector $\mathbf{b}$ is unknown, meaning that an estimate $\hat{\mathbf{b}}$ is necessary to construct the correlation matrix $\mathbf{R}_{\mathbf{y}}(\hat{\mathbf{b}})$. The more accurate $\hat{\mathbf{b}}$ is, the better the approximation to the true log-likelihood function.

We infer that after trying several hypotheses $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$, the retrieval of the most accurate estimate $\hat{\mathbf{b}}$ entails the appropriate selection of the model order. In this way, vector $\boldsymbol{\varphi}$ can be indirectly estimated. Note that the dimension of $\hat{\boldsymbol{\varphi}}_n$ changes among the hypotheses $n = 1, \ldots, N$; meaning that each $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$ refers to a different model order.

As exemplified in [1, (C.3.6)], we emphasize the dependence of the pdf in (14) on vector $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$ in (10) through

$$\hat{p}_n\left(\mathbf{y}\left|\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right.\right) \triangleq \hat{p}_n\left(\mathbf{y}, \hat{\boldsymbol{\varphi}}_n\right);\ n = 1, \ldots, N. \tag{16}$$

The expression at the right hand side denotes the pdf of $\mathbf{y}$ under $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$, while the expression at the left hand side denotes the pdf of $\mathbf{y}$ under $\{\hat{\mathbf{b}}(\hat{\boldsymbol{\varphi}}_n)\}_{n=1}^{N}$ for the hypotheses $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$, which are assumed mutually exclusive.

With these clarifications, we can deduce that the treated MOS tool is *a posteriori*. Given a parametric focusing technique (e.g., MUSIC or ESPRIT), different hypotheses $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$ are considered; then, the respective several estimates $\{\ln\{\hat{p}_n(\mathbf{y}|\hat{\mathbf{b}}(\hat{\boldsymbol{\varphi}}_n))\}\}_{n=1}^{N}$ are retrieved; and finally, after comparing all of them, the most suitable model order is selected, according to a criterion defined next.

B. Kullback–Leibler Information Criterion

The KL information function [1, Appendix C], [12],

$$D\left(p, \hat{p}\right) = \int p\left(\mathbf{y}\right) \ln\left\{\frac{p\left(\mathbf{y}\right)}{\hat{p}\left(\mathbf{y}\right)}\right\} d\mathbf{y} \tag{17}$$

measures the discrepancy between the true pdf $p(\mathbf{y})$ and the pdf of the data model $\hat{p}(\mathbf{y})$. The ideal choice for $\hat{p}(\mathbf{y})$ in (17) would be the modeled likelihood $p(\mathbf{y}|\mathbf{b}(\boldsymbol{\varphi}))$ in (15), nonetheless, as explained previously, this function is not available, since $\mathbf{b}(\boldsymbol{\varphi})$ is unknown and needs to be estimated. Therefore, we refer to $\{\hat{p}_n(\mathbf{y}|\hat{\mathbf{b}}(\hat{\boldsymbol{\varphi}}_n))\}_{n=1}^{N}$ instead.

The AIC and BIC MOS rules, introduced in [12], consist in minimizing the KL discrepancy in (17), given the several hypotheses $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$. In the case of AIC, this is equivalent to maximizing the relative KL information [12]

$$I\left(p, \hat{p}_n\right) \cong \mathrm{E}_{\mathbf{y}}\left(\ln\left\{\hat{p}_n\left(\mathbf{y}\left|\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right.\right)\right\} - M\right);$$
$$n = 1, \ldots, N. \tag{18}$$

Here, $\mathrm{E}_{\mathbf{y}}(\cdot)$ denotes the expectation with respect to the pdf of $\mathbf{y}$. Subsequently, the relative KL information in (18) is estimated in an unbiased manner via minimizing

$$\mathrm{AIC}\left(\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right) = -2\cdot\ln\left\{\hat{p}_n\left(\mathbf{y}\left|\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right.\right)\right\} + 2M; \quad n = 1,\ldots,N; \tag{19}$$

which defines the AIC MOS rule. Conversely, the BIC MOS rule selects the hypothesis that maximizes the relative KL information in (18) through the approximation [12]

$$\hat{I}\left(p,\hat{p}_n\right) \cong \ln\left\{\hat{p}_n\left(\mathbf{y}\left|\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right.\right)\right\} - \frac{M}{2}\ln\{L\};\ n = 1,\ldots,N; \tag{20}$$

or equivalently minimizes

$$\mathrm{BIC}\left(\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right) = -2\cdot\ln\left\{\hat{p}_n\left(\mathbf{y}\left|\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right.\right)\right\} + M\cdot\ln\{L\}; \quad n = 1,\ldots,N. \tag{21}$$

Under the log-likelihood function in (10), M in (19) and (21), respectively, represents the model order, namely, the dimension of vector $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$ [1, Appendix C]. Nonetheless, for the considered log-likelihood function in (14), with $\mathbf{R}_{\mathbf{y}}(\hat{\mathbf{b}})$ as in (9), M refers to the dimension of vector $\hat{\mathbf{b}}(\hat{\boldsymbol{\varphi}}_n)$, which remains constant. This means that both expressions in (19) and (21), are reduced to

$$\mathrm{KL}\left(\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right) = -2\cdot\ln\left\{\hat{p}_n\left(\mathbf{y}\left|\hat{\mathbf{b}}\left(\hat{\boldsymbol{\varphi}}_n\right)\right.\right)\right\};\ n = 1,\ldots,N; \tag{22}$$

since the addition of a constant value does not change the minimizer.

In the previous studies [12], [23], [24], the second term of AIC and BIC prevents selecting model orders larger than the most proper one, with the aim of avoiding overfitting. However, in this work, both expressions are reduced to (22), which does not have a penalty term. The latter may translate into choosing models with a higher order. Nevertheless, take into account that the models with relatively larger orders tend to perform better when the data generating mechanism is more complex than the models used to fit [1, Appendix C]. This means that, under such circumstance, the model orders that are relatively higher than the most proper one, tend to retrieve also good-fitted solutions.

We assume that TomoSAR falls into the case aforementioned, since most of the perturbations in the observed data are due to the signal formation; recall that the TomoSAR problem is intrinsically ill-conditioned as $M \gg L$. Subsequently, as argued in Section V, since KL in (22) does not have a penalty term, we suggest preventing overfitting by restricting the upper limit of the range of possible hypotheses $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N}$, excluding the highest orders. Conversely, restricting the lower limit prevents underfitting.

C. Multiple Signal Classification

The capabilities of KL in (22) are assessed in Section V; for such a purpose, we refer to MUSIC [1], [11]. Let $[\varrho_1 \geq \varrho_2 \geq \ldots \geq \varrho_L]$ denote the set of eigenvalues, arranged in decreasing order, of the sample covariance matrix $\mathbf{Y}$ in (4) of dimension $L \times L$. The corresponding eigenvectors are split into two subsets: $\mathbf{Q} = [\,\mathbf{q}_1\ \mathbf{q}_2\ \ldots\ \mathbf{q}_n\,]$ is related to the n eigenvalues in the signal subspace, whereas $\mathbf{G} = [\,\mathbf{g}_1\ \mathbf{g}_2\ \ldots\ \mathbf{g}_{L-n}\,]$ is related to the remaining $L-n$ eigenvalues in the noise subspace. For such a model, the MUSIC spectral estimator is defined via [1, (4.5.15)]:

$$\left\{\hat{b}_m = \frac{1}{\mathbf{a}_m^{+}\mathbf{G}\mathbf{G}^{+}\mathbf{a}_m}\right\}_{m=1}^{M} \tag{23}$$

with $\mathbf{G}\mathbf{G}^{+}$ as the so-called noise subspace covariance matrix. Note that n refers to the model order, which is to be selected using KL in (22).

Vectors $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N=L-1}$ in (22) contain the eigenvalues in the signal subspace (arranged in decreasing order) as presumed for the different hypothesis $n = 1,\ldots,L-1$. The corresponding model orders, for each one of these vectors, are given by $\{n\}_{n=1}^{N=L-1}$. For this particular case, the index of the hypotheses $\{\hat{\boldsymbol{\varphi}}_n\}_{n=1}^{N=L-1}$ corresponds to the dimension of the respective vector. The upper limit is set to $N = L-1$, since we have to make sure of having at least one eigenvector in $\mathbf{G}$ to be able to compute $\mathbf{G}\mathbf{G}^{+}$ in (23).

Table I presents the explicit implementation of MUSIC, incorporating the MOS tool in (22). We consider all possible assignments $n = 1,\ldots,\ L-1$, however, this range can be modified by the user. As many other super resolution techniques, MUSIC does not preserve radiometric accuracy; therefore, in order to keep $\{\hat{\mathbf{b}}_n\}_{n=1}^{N=L-1}$ within the same numerical range as $\mathbf{y}$ when computing KL in (22), it is advisable to normalize $\{\hat{\mathbf{b}}_n\}_{n=1}^{N=L-1}$ with respect to $\mathrm{tr}\{\mathbf{Y}\}$.

TABLE I
INCORPORATION OF MOS INTO MUSIC FOR THE TOMOSAR CASE

$\{\hat{\mathbf{b}}_n\}_{n=1}^{N=L-1} = \mathbf{0}$
$\mathbf{h} = \mathbf{0}$

for $n = 1,\ldots,(L-1)$
  $\mathbf{G}(n) = [\mathbf{g}_1 \quad \cdots \quad \mathbf{g}_{L-n}]$
  for $m = 1,\ldots,M$
    $\hat{b}_{n_m} = \dfrac{1}{\mathbf{a}_m^{+}\mathbf{G}\mathbf{G}^{+}\mathbf{a}_m}$
  $h_n = \mathrm{KL}(\hat{\mathbf{b}}_n)$

$ord = \mathrm{argmin}(\mathbf{h})$
$\hat{\mathbf{b}}_{\mathrm{MUSIC}} = \hat{\mathbf{b}}_{ord}$

return $\hat{\mathbf{b}}_{\mathrm{MUSIC}}$, ord

IV. Regularization Parameter Selection

Regularization approaches are widely used to solve linear problems as the one given in (1), see [14] and [22], and the references therein. The main idea is to replace the ill-conditioned problem with a well-posed optimization problem. The retrieval of well-conditioned solutions (in the Hadamard sense [22]) is

accomplished by incorporating known properties of the solution into the solver and by providing smoothing into the solution.

In previous related studies [3]–[6], [25], [26], different statistical regularization techniques have been extended to cope with nonlinear ill-posed inverse problems, which is the case for TomoSAR. The methodology is easy to implement; however, the correct setting of regularization parameters is normally required. There is no known method for the optimal choice of such regularization parameters [14], nevertheless, various strategies offer satisfactory approximations for several imaging problems, e.g., denoising.

The aim of this section is to adapt some of the most popular parameter selection criteria to the case under study. Specifically, we refer to the Morozov's discrepancy principle [14], the L-Curve [14], [15], SURE [16], [17], and GCV [17], [18]. In order to assess these tools, we refer to MARIA [3], [4], to perform focusing of the TomoSAR data. MARIA is an iterative regularization technique that achieves super resolution.

We start by clarifying to which regularization techniques the aforementioned parameter selection methodology is applicable. These regularization methods have a particular structure, to be discussed in the following subsection.

A. Solution Operator

As explained previously in Section II, the nonlinear TomoSAR inverse problem consists in recovering an estimate $\hat{\mathbf{b}}$ of the PSP vector $\mathbf{b}$ in (8). The latter can be expressed through the use of a solution operator $\mathbf{F}$, via

$$\hat{\mathbf{b}} = \left\{\mathbf{F}\mathbf{y}\mathbf{y}^{+}\mathbf{F}^{+}\right\}_{\mathrm{diag}} \tag{24}$$

where operator $\{\cdot\}_{\mathrm{diag}}$ retrieves a vector composed of the elements at the main diagonal of the embraced matrix. Pursuing better accuracy in presence of multiplicative noise and in order to handle the multiple nondeterministic sources, the solver in (24) may rather employ the sample covariance matrix in (4), such that $\hat{\mathbf{b}} = \{\mathbf{F}\mathbf{Y}\mathbf{F}^{+}\}_{\mathrm{diag}}$.

By instance, the popular descriptive regularization CLS method [3], [26], solves the linear equation of observation in (1) in the following manner:

$$\hat{\mathbf{s}}_{\mathrm{CLS}} = \mathbf{F}_{\mathrm{CLS}}\,\mathbf{y} = \left(\mathbf{A}^{+}\mathbf{A} + \alpha\mathbf{I}\right)^{-1}\mathbf{A}^{+}\mathbf{y} \tag{25}$$

with $\mathbf{F}_{\mathrm{CLS}}$ as a solution operator and with α as a regularization parameter. The regularization term $\alpha\mathbf{I}$ in $\mathbf{F}_{\mathrm{CLS}}$ in (25) is necessary to make the solution well-conditioned. Hence, solving (1) requires the choice of a suitable value $\alpha > 0$, which in the limiting case when $\alpha \to 0$, the problem is reduced to its unconstrained ill-posed least squares version. How to select a proper value of α, referred from now on as $\xi = \alpha$, is argued later on.

Note that the expression given in (24) admits the usage of $\mathbf{F}_{\mathrm{CLS}}$ to obtain an estimate $\hat{\mathbf{b}}_{\mathrm{CLS}}$. Thus, having a solution operator $\mathbf{F}$ permits solving the linear equation of observation in (1) and the nonlinear TomoSAR problem in (24).

As observed further below, in principle, the addressed strategies for choosing regularization parameters can be applied to all those regularization techniques that make use of a solution operator $\mathbf{F}$. A list of several regularization techniques that meet this requirement can be found in [3, Table I]; among these methods is MARIA, presented next.

B. Overview of MARIA

MARIA is an iterative statistical regularization method that offers similar advantages as the DOA-inspired parametric techniques do, which are: resolution enhancement, suppression of artifacts and reduction of the ambiguity levels [3], [4]. MARIA provides an approximate solution to the ML optimization problem in (15) via [4]

$$\hat{\mathbf{b}}^{[i+1]}_{\mathrm{MARIA}} = \mathbf{T}^{[i]}\left[\hat{\mathbf{b}}^{[i]}_{\mathrm{BMR}} - \mathbf{w}^{[i]}\right];\ i = 0, 1, \ldots, I; \tag{26}$$

where vector $\hat{\mathbf{b}}^{[i]}_{\mathrm{BMR}} = \{\mathbf{F}^{[i]}_{\mathrm{BMR}}\mathbf{Y}\mathbf{F}^{[i]+}_{\mathrm{BMR}}\}_{\mathrm{diag}}$ is recognized as the BMR estimate of the PSP. The subtraction of the bias vector $\mathbf{w}^{[i]} = \{\mathbf{F}^{[i]}_{\mathrm{BMR}}\mathbf{R}_{\mathbf{n}}\mathbf{F}^{[i]+}_{\mathrm{BMR}}\}_{\mathrm{diag}}$ from $\hat{\mathbf{b}}^{[i]}_{\mathrm{BMR}}$ corrects the shift due to the noise in the observed data, whereas the diagonal matrix $\mathbf{T}^{[i]} = \mathbf{D}(\{\mathbf{A}^{+}\mathbf{F}^{[i]+}_{\mathrm{BMR}}\mathbf{F}^{[i]}_{\mathrm{BMR}}\mathbf{A}\}_{\mathrm{diag}})$ is an adaptive window operator that provides smoothing to the already rectified BMR estimate. Matrix $\mathbf{F}^{[i]}_{\mathrm{BMR}}$ stands for the so-called solution operator, defined as

$$\mathbf{F}^{[i]}_{\mathrm{BMR}} = \mathbf{D}\left(\hat{\mathbf{b}}^{[i]}\right)\mathbf{A}^{+}\mathbf{R}^{-1}_{\mathbf{y}}. \tag{27}$$

Notice that (27) requires of a first estimate of the PSP $\hat{\mathbf{b}}^{[0]}$ in order to construct the matrices $\mathbf{D}(\hat{\mathbf{b}}^{[0]})$ and $\mathbf{R}_{\mathbf{y}} = \mathbf{A}\mathbf{D}(\hat{\mathbf{b}}^{[0]})\mathbf{A}^{+} + N_0\mathbf{I}$.

The dependence on a first estimate $\hat{\mathbf{b}}^{[0]}$ implies that no unique regularization method to recover $\hat{\mathbf{b}}_{\mathrm{MARIA}}$ in (26) exists [3]. Different solutions are retrieved for different initial estimates $\hat{\mathbf{b}}^{[0]}$, when the discrepancies between them are highly significant. Yet, the adaptive iterative implementation of MARIA alleviates the issue in certain extent.

MARIA refines the estimated PSP $\hat{\mathbf{b}}^{[i]}$ after each iteration, until achieving convergence $\hat{\mathbf{b}}^{[i=I]}_{\mathrm{MARIA}}$. In practice, the iterative process is finished either by reaching a user tolerance control level or a maximum number of iterations. We recommend using Capon beamforming [1], [20], to retrieve the initial input to (26). Having a satisfactory first approximation $\hat{\mathbf{b}}^{[0]}$ reduces the processing time, since less iterations are required to converge.

Factor N_0 in $\mathbf{R}_{\mathbf{y}}$ in (27) assumes the role of a diagonal-loading regularization parameter, which assures matrix $\mathbf{R}_{\mathbf{y}}$ to be invertible. The choice of this factor is extremely relevant, since a wrong assigned value could cause under/over regularization of the retrieved solutions. Consequently, the regularization parameter, referred from now on as $\xi = N_0$, must guarantee retrieving a good-fitted (well-regularized) reconstruction.

In the following, we explain how to pick ξ by employing the considered parameter selection criteria. These strategies are normally divided into two categories.

1) One assuming previous knowledge of the noise variance; including methods as the Morozov's discrepancy principle and SURE.
2) Another one extracting all necessary information from the measurements themselves, with techniques as the L-Curve and GCV.

C. Morozov

The Morozov's discrepancy principle selects the regularization parameter ξ in (25) and (27), based on the noise level in the data. A solution is acceptable if it produces a measurement with the same magnitude as the estimated error in the data [14]. Accordingly, the idea of the Morozov's discrepancy principle is to choose a $\xi > 0$, such that

$$\left\|\mathbf{A}\hat{\mathbf{s}}\left(\xi\right)-\mathbf{y}\right\|^2=\delta^2, \tag{28}$$

where $\delta = \mathrm{E}(\|\mathbf{n}\|) = \sqrt{L\sigma^2}$ and with $\hat{\mathbf{s}}(\xi) = \mathbf{F}(\xi)\mathbf{y}$. The term σ^2 stands for the Gaussian noise variance. Factor N_0 in $\mathbf{R}_{\mathbf{y}}$ has a different sense for MARIA, it rather refers to the regularization parameter ξ to be chosen. For such a case, we do not assume that $\sigma^2 = N_0$.

Let $\mathbf{y}' = \mathbf{U}^{+}\,\mathbf{y}$, with $\mathbf{U}$ obtained from the TSVD of the steering matrix $\mathbf{A}$. The numerical implementation of the Morozov's method, introduced in [14, Ch. 5], consists in finding a proper ξ as the unique zero of the function

$$f\left(\xi_n\right)=\sum_{l=1}^{K}\left(\frac{\xi_n}{\mu_l^2+\xi_n}\right)^2\left(y'_l\right)^2 + \sum_{l=K+1}^{L}\left(y'_l\right)^2-\delta^2;\; n=1,\ldots,N; \tag{29}$$

in which $\{\mu_l\}_{l=1}^{K}$ are the singular values of matrix $\mathbf{A}$, which are not set to zero. In order to define which singular values $\{\mu_l\}_{l=K+1}^{L}$ set to zero, a PCA of matrix $\mathbf{A}$ is performed. As explained in [27, Ch. 7], the total variance of the data in $\mathbf{A}$ is the sum of all their singular values $\sum_{l=1}^{L}\mu_l$. The first eigenvector $\mathbf{u}_1$ points in the most significant direction of the data and explains a fraction of the total variance. The next eigenvectors $\mathbf{u}_2,\ldots,\mathbf{u}_L$ accounts for decreasing fractions. Hence, we only preserve those singular values which explain most of the data. When μ_1 is much larger than μ_K in (29) it can lead to instabilities due to the amplification of the truncated error. Preventing this, in the simulations presented in Section V, we keep 80% of the singular values of matrix $\mathbf{A}$ to compute (29), the remaining 20% are set to zero.

The main advantage of this implementation is that it does not require the computation of several estimates $\{\hat{\mathbf{s}}(\xi_n)\}_{n=1}^{N}$ to select the most suitable ξ, making it faster in contrast to the other related techniques.

D. L-Curve

The L-Curve method seeks a balance between the norm of a penalty term and the norm of the residual [14], [15]. It basically consists in forming a smooth curve by plotting the points

$$L_C\left(\xi_n\right)=\left[\ln\left\|\left\{\mathbf{A}\hat{\mathbf{s}}\left(\xi_n\right)-\mathbf{y}\right\}\right\|,\ln\left\{\left\|\hat{\mathbf{s}}\left(\xi_n\right)\right\|\right\}\right], \tag{30}$$

for a collection of candidates $\{\xi_n\}_{n=1}^{N}$ with $\{\hat{\mathbf{s}}(\xi_n)=\mathbf{F}(\xi_n)\mathbf{y}\}_{n=1}^{N}$. The resultant curve has the shape of a letter L, with a smooth corner; the proper value for ξ is found as near as possible to this corner.

The L-Curve consists of two parts, a flat (horizontal) fragment, where the errors due to regularization dominate, and a steep (horizontal) part, where the perturbation errors dominate. The presence of perturbations (i.e., $\mathbf{y}+\triangle\,\mathbf{y}$, $\mathbf{Y}+\triangle\,\mathbf{Y}$,) when solving practical problems is unavoidable, since the data (i.e., $\mathbf{y}$, $\mathbf{Y}$) is obtained from measurements, containing errors.

The horizontal part of the L-Curve corresponds to those solutions where so much regularization is introduced; the solution stays very smooth and $\|\hat{\mathbf{s}}(\xi_n)\|$ changes a little with the regularization parameter. In contrast, the vertical part of the L-Curve corresponds to solutions where $\|\hat{\mathbf{s}}(\xi_n)\|$ varies dramatically with the regularization parameter while, at the same time, the residual norm does not change much. When the horizontal and vertical parts of the L-Curve start approaching each other, both terms, $\|\hat{\mathbf{s}}(\xi_n)\|$ and $\|\mathbf{A}\hat{\mathbf{s}}(\xi_n)-\mathbf{y}\|$, change more significantly with the regularization parameter and at a different rate. Therefore, the reason to use the $\ln-\ln$ scale in (30) is emphasizing the corner separating the vertical and the horizontal parts [28]. However, depending on the problem, the corner in the L-Curve may not be clearly defined.

Finding the corner of an L-Curve is not a trivial task; for such goal, this article recommends using the algorithm presented in [29]. The suggested algorithm is based on an estimation of the local curvature of the L-Curve from three sampled points, as defined by Menger [30], and a sampling update rule based on the golden section search [31].

Three regularization parameters $\xi_j<\xi_k<\xi_l$ identify three points $L_C(\xi_j)<L_C(\xi_k)<L_C(\xi_l)$ along the curve, from which the local curvature is obtained, via

$$C\left(L_C\left(\xi_j\right),\,L_C\left(\xi_k\right),\,L_C\left(\xi_l\right)\right)=\frac{4T}{\left\|L_C\left(\xi_j\right)-L_C\left(\xi_k\right)\right\|\left\|L_C\left(\xi_k\right)-L_C\left(\xi_l\right)\right\|\left\|L_C\left(\xi_l\right)-L_C\left(\xi_j\right)\right\|} \tag{31}$$

where T is the area of a triangle created with the three given points. Having this in mind, ξ is selected from the point with the largest change direction within an equal distance, which, in terms of the previous definition, is the point of the curve with the largest positive curvature.

In order to find the L-Curve's corner, a golden section search is performed [31]. First, the limits of the search $[\xi_1,\,\xi_4]$ are defined, such that $\xi_i=10^{x_i}$. Later on, the remaining points are calculated through

$$x_2=\frac{x_4+\beta x_1}{1-\beta};\; x_3=x_1-\left(x_4-x_2\right); \tag{32}$$

with $\beta=(1+\sqrt{5})/2$ as the so-called golden ratio.

Four points, $L_C(\xi_1)$, $L_C(\xi_2)$, $L_C(\xi_3)$, and $L_C(\xi_4)$, are taken from the L-Curve in order to compute two curvatures per iteration, $C_1=C(L_C(\xi_1),\,L_C(\xi_2),\,L_C(\xi_3))$ and $C_2=C(L_C(\xi_2),L_C(\xi_3),\;L_C(\xi_4))$. These curvatures are compared under two options.

1) If $C_1>C_2$ then the next reassignments are made: $\xi_4\leftarrow\xi_3,\xi_3\leftarrow\xi_2,L_C(\xi_4)\leftarrow L_C(\xi_3)$ and $L_C(\xi_3)\leftarrow L_C(\xi_2)$. The recalculation of ξ_2 and $L_C(\xi_2)$ is made.
2) If $C_1<C_2$ then the next reassignments are made: $\xi_1\leftarrow\xi_2,\xi_2\leftarrow\xi_3,L_C(\xi_1)\leftarrow L_C(\xi_2)$ and $L_C(\xi_2)\leftarrow L_C(\xi_3)$. The recalculation of ξ_3 and $L_C(\xi_3)$ is made.

The loop stops when the distance between $\xi_1$and ξ_4 becomes smaller than a specified threshold. To secure a positive curvature of C_2 some checks are performed, so that, while $C_2 < 0$, the limit value is shifted $\xi_4 \leftarrow \xi_3$ and the respective computations are made to advance to the next iteration.

E. Stein's Unbiased Risk Estimate

In 1978, with the publication of the renowned article "Smoothing noisy data with spline functions" [18], Craven and Wahba introduced the basis for SURE and GCV. They propose selecting an appropriate regularization parameter ξ, as the one that minimizes an estimate of the MSE between the actual and the recovered signal, calculated without knowing the true signal. Then, ξ is chosen so that it satisfies

$$\frac{1}{L}\sum_{l=1}^{L}\left(y_l - \hat{y}_l\right)^2 = \sigma^2 \tag{33}$$

with $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{s}}(\xi)$, $\hat{\mathbf{s}}(\xi) = \mathbf{F}(\xi)\mathbf{y}$ and σ^2 as the Gaussian noise variance.

The MSE is estimated via the unbiased approximation [18, (1.8)]

$$\hat{\eta}\left(\xi_n\right) = \frac{1}{L}\left\|\left(\mathbf{I} - \mathbf{AF}\left(\xi_n\right)\right)\mathbf{y}\right\|^2 - \frac{\sigma^2}{L}\mathrm{tr}\left\{\left(\mathbf{I} - \mathbf{AF}\left(\xi_n\right)\right)^2\right\} + \frac{\sigma^2}{L}\mathrm{tr}\left\{\left(\mathbf{AF}\left(\xi_n\right)\right)^2\right\};\ n = 1, \ldots, N. \tag{34}$$

Subsequently, using the property [2, (A.27)],

$$\mathrm{tr}\left\{\mathbf{O} + \mathbf{P}\right\} = \mathrm{tr}\left\{\mathbf{O}\right\} + \mathrm{tr}\left\{\mathbf{P}\right\}, \tag{35}$$

in which $\mathbf{O}$ and $\mathbf{P}$ are arbitrary square matrices, taking into account that $\mathrm{tr}\{\mathbf{I}\} = L$ and performing a sequence of evident manipulations, (34) is simplified, yielding the SURE strategy

$$\hat{\eta}_{\mathrm{SURE}}\left(\xi_n\right) = \frac{1}{L}\left\|\left(\mathbf{I} - \mathbf{AF}\left(\xi_n\right)\right)\mathbf{y}\right\|^2 - \sigma^2 + \frac{2\sigma^2}{L}\mathrm{tr}\left\{\mathbf{AF}\left(\xi_n\right)\right\};\ n = 1, \ldots, N. \tag{36}$$

Equation (36) is equivalent to the SURE expression given in [16, (6)], however, the equation in [16] is specific to the denoising problem, which assumes that $M = L$. Thus, the main reason to derive SURE as in (36) is to provide an expression suitable to the TomoSAR case.

Within the TomoSAR context, the SURE technique consists then in finding a value for ξ, which produces the minimum MSE between the true complex reflectivity vector $\mathbf{s}$ in (1) and the recovered one $\hat{\mathbf{s}}$, approximated using (36).

F. Generalized Cross Validation

The previously addressed parameter selection criterion, along with the Morozov's discrepancy principle in (29), implies the *a priori* knowledge of the Gaussian noise variance σ^2; however, this value is not always known. Under such circumstances, it is better to make use of a different strategy. We recommend using the GCV method instead, defined by [18, (1.9)]

$$\hat{\eta}_{\mathrm{GCV}}\left(\xi_n\right) = \frac{\frac{1}{L}\left\|\left(\mathbf{I} - \mathbf{AF}\left(\xi_n\right)\right)\mathbf{y}\right\|^2}{\left[\frac{1}{L}\mathrm{tr}\left\{\mathbf{I} - \mathbf{AF}\left(\xi_n\right)\right\}\right]^2};\ n = 1, \ldots, N. \tag{37}$$

Similar to SURE in (36), GCV computes an estimate of the MSE between the actual complex reflectivity vector $\mathbf{s}$ in (1) and the retrieved signal $\hat{\mathbf{s}}$, but without requiring knowledge of σ^2. Thus, ξ is chosen as the minimizer of (37).

Contrary to SURE in (36), the GCV technique is more dependent on having a large number of passes L in order to retrieve more accurate estimations of the MSE [18], such scenario is typically not the case for the TomoSAR problem. Subsequently, although both SURE and GCV have the same principle, they may retrieve different MSE estimates, and as a result a different ξ may be chosen. Also, although having *a priori* knowledge on the noise variance is a noticeable advantage, the way that Morozov in (29) and SURE in (36) handle this information, may not be convenient for the TomoSAR case. Being TomoSAR a highly ill-conditioned inverse problem, most of the perturbations in the observed data are due to the formation of the received signal, rather than due to the additive noise. Morozov and SURE give more weight to the perturbations due to the additive noise: Factor σ^2 in (28) and (33) accounts for the noise variance only, it does not gather significant information about the perturbations due to the signal formation. Therefore, the respective retrieved approximations may result misleading for the TomoSAR problem. Consequently, for practical scenarios, we expect a more reliable behavior from those techniques which do not require of σ^2 as *a priori* information (i.e., the L-Curve and GCV), since they extract all necessary information from the measurements themselves, giving a better balance to all kind of perturbations.

V. Simulation Results

This section analyses the capabilities of the treated parameter selection criteria through simulations. The performance of the addressed MOS rule is studied utilizing MUSIC (refer to Table I), whereas the addressed strategies for choosing regularization parameters make use of MARIA in (26).

As mentioned previously, parametric techniques as MUSIC assume that the PSP is composed by a finite number of point-type like backscattering sources. Therefore, due to the characteristics of MUSIC, we refer to an urban like scenario, composed of several point-type like targets. Keep in mind, however, that MARIA is in principle also appropriate for scenarios composed by distributed targets [3], [4].

The simulated tomographic acquisition geometry spans a PLOS synthetic aperture of 70 m with $L = 15$ passes evenly distributed. The considered L-band SAR sensor (0.23 m wavelength) is assumed to be at a nominal altitude of 3000 m. We consider a slant-range distance of about 4000 m between the master track and the targets, meaning a Fourier resolution [8] of about 6 m.

The MUSIC and MARIA focusing techniques are applied on simulated sample covariance matrices $\mathbf{Y}$, constructed as in (4) for $J = 250$ independent looks (see Fig. 1). The data covariance matrices $\mathbf{Y}$ gather the echoes coming from the scatterers displaced along the PLOS height direction. We consider four

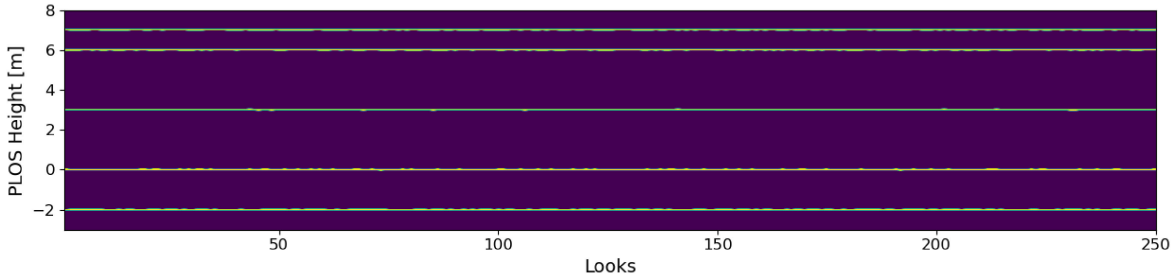

Fig. 1. Distribution among $J = 250$ looks of 5 point-type like targets displaced along the PLOS height direction. Each target is composed of 100 scatterers following a Gaussian distribution, with phase center (mean) located at $\dot{z}_1 = -2$ m, $\dot{z}_2 = 0$ m, $\dot{z}_3 = 3$ m, $\dot{z}_4 = 6$ m, and $\dot{z}_5 = 7$ m, respectively, and spread (standard deviation) of 0.01 m.

TABLE II
CASES OF STUDY

Case	Phase-Centers	
	PLOS Location [m]	Quantity
1	$\dot{z}_1 = -2$ and $\dot{z}_2 = 0$	2
2	$\dot{z}_1 = -2, \dot{z}_2 = 0$ and $\dot{z}_3 = 3$	3
3	$\dot{z}_1 = -2, \dot{z}_2 = 0, \dot{z}_3 = 3$ and $\dot{z}_4 = 6$	4
4	$\dot{z}_1 = -2, \dot{z}_2 = 0, \dot{z}_3 = 3, \dot{z}_4 = 6$ and $\dot{z}_5 = 7$	5

cases of study, each one with different number of targets, as specified in Table II. For the aim of this work, we denote as phase centers to the mean heights of the multiple backscattering sources. Accordingly, each target in Table II is composed of 100 scatterers with equal reflectivity, following Gaussian distributions with phase-centers (means) placed as indicated in Table II. For all targets, a narrow spread (standard deviation) of 0.01 m is settled. The reason of using Gaussian distributions to simulate the backscattering responses is twofold.

1) The location of the phase-centers matches the mean values.
2) The scatterers composing each Gaussian distribution fluctuate randomly for each independent look (see Fig. 1), adding statistical uncertainty to the measurements. In this way, we do not only depend on the additive noise to introduce decorrelation.

MUSIC usually assumes that the number of point-type like targets, composing the PSP, is much smaller than the number of passes L [1, Ch. 5]. Consequently, we set a limit of five targets for the simulations.

Fig. 2 presents the true (theoretical) PSP for the fourth case of study in Table II (blue) and the recovered PSP after applying matched filtering on the simulated data (red). The expected PSP for the other cases of study can be inferred from the blue plot, removing the corresponding targets. Observe that only two phase-centers are detected via matched filtering. The locations of the point-type like targets (along the PLOS height direction) is chosen with the aim of demonstrating the super resolution capabilities of MUSIC and MARIA, in contrast to matched filtering. The targets are placed close to each other and unevenly spaced, with the objective of stressing these methods out. Particularly, the last two targets are more difficult to discriminate due to their close proximity.

The quality of the retrieved solutions, after focusing, is quantified using three metrics.

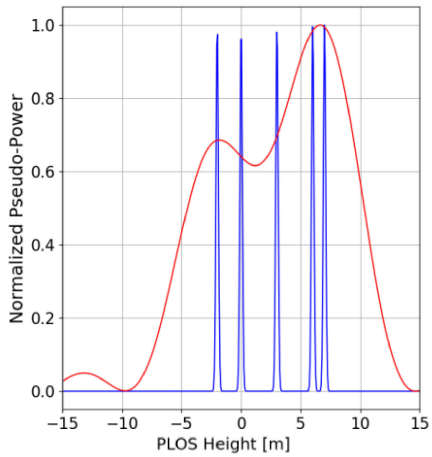

Fig. 2. Blue: Expected PSP for the fourth case of study in Table II. Red: Retrieved PSP after applying matched filtering on the simulated data.

1) *Detection Rate*: It refers to the number of times that the amount of phase centers is correctly recovered, with respect to the expected quantity defined in Table II. For such a purpose, we look for the local maxima above a threshold value of 0.05 within the retrieved (0 to 1) normalized pseudo-power. The reason to use a threshold value is avoiding false detections.
2) *RMSE:* When the number of phase-centers is correctly identified, the RMSE between the true locations $\dot{\mathbf{z}}$ (as described in Table II) and the ones found $\hat{\dot{\mathbf{z}}}$ is calculated, via

$$RMSE\left(\dot{\mathbf{z}}, \hat{\dot{\mathbf{z}}}\right) = \sqrt{\sum_{i=1}^{I} \frac{\left(\dot{z}_i - \hat{\dot{z}}_i\right)^2}{I}}. \quad (38)$$

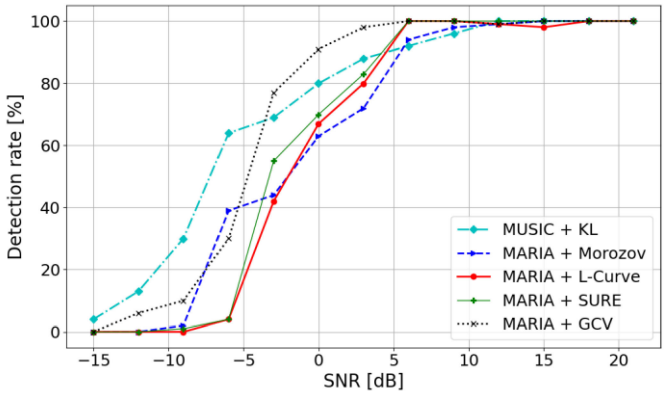

Fig. 3. Detection rate against SNR for the fourth case of study in Table II. Each point in the plot is the average of 250 Monte Carlo simulations.

3) *Fréchet Distance*: When the number of phase centers is correctly identified, we make use of the FD [32], [33], to measure the similarity between the expected (theoretical) PSP $\mathbf{b}$ and the recovered PSP $\hat{\mathbf{b}}$. The FD ranges between zero and one; it equals zero when both signals are the exact same and goes up to one, depending on how different the signals are. We refer to the algorithm in [33] to compute the FD: Let P and Q be discrete polynomial curves; the respective discrete sequence is expressed as $P = (u_1, \ldots, u_M)$ and $Q = (v_1, \ldots, v_M)$. The coupling Ψ between P and Q is a sequence $(u_1, v_1), (u_2, v_2), \ldots, (u_M, v_M)$, which respects the order of the points in P and Q. The length of the coupling

$$\ell(\Psi) = \max_{i=1,\ldots,M} \ell(u_i, v_i) \tag{39}$$

is the length of the longest link in Ψ. Finally, the FD is defined by

$$FD(P, Q) = \min\{\ell(\Psi)\}. \tag{40}$$

Fig. 3 depicts the detection rate against SNR for the fourth case of study in Table II. Each point in the plot is the average of 250 Monte Carlo trials. Observe how the addressed techniques converge to a higher detection rate as the SNR increases. All methods have improved performance above 5 dB, consequently, a SNR of 7 dB is considered in the simulations reported next. The considered white noise power spectral density is assumed to be known *a priori*.

In the following, two hundred and fifty Monte Carlo simulations are conducted for each case of study in Table II, and for each parameter selection criterion. In the case of MUSIC, we take into account all possible model orders $\{n\}_{n=1}^{N=L-1}$, resulting in a total of $250 \times (L-1) \times 4 = 14\,000$ executed simulations. The idea is to compare the performance of MUSIC for all model orders and to observe which model orders are the most suitable ones. In order to reconstruct the PSP along the PLOS height direction in detail, we make use of $M = 150$ samples within the PLOS height range from -5 to 10 m.

Figs. 4 and 5, and Table III present the compendium of the results gotten from all simulations. The colored matrices in Fig. 4 correspond to MUSIC, they are read row by row, one case of study at a time, comparing the measurements obtained for all $(L-1)$ model orders. The color bars at right hand indicate the attained values with respect to the specified metric: 1) Detection rate, 2) average RMSE, and 3) average FD. Note that we use different colors in every graphic, with the aim of making the transitions more notorious. In Fig. 3(b) and (c), we set to one those model orders that have a detection rate of 0% for a given case of study. Note that this circumstance makes the computation of the RMSE, as defined in (38), unfeasible.

The curves superimposed onto each row of the colored matrices in Fig. 4, show the percentage of times that the treated MOS rule [i.e., KL in (22)] chose a particular model order for the given case of study. The summation of all the percentages along each row equals 100%. By instance, KL chose the sixth model order about 10% of the times, for the fourth case of study.

Fig. 5 depicts the superimposed vertical profiles of the retrieved PSP, organized per case of study and per parameter selection strategy. For completeness, Table III presents the detection rate, average processing time, average RMSE and average FD, attained by each treated parameter selection criterion. The corresponding measurements are gotten after normalizing (from 0 to 1) both curves, the reconstructed PSP and the expected PSP, since MUSIC and MARIA, as most super resolution techniques, do not preserve radiometric accuracy.

A. Results Obtained by MUSIC

Observe in Fig. 4 that, excluding the two largest model orders, those model orders above five achieve satisfactory measurements. In contrast, the model orders lower than the expected number of targets attain poor results. Note also that the best choice corresponds to the expected number of targets.

As discussed previously in Section III, KL tends to select models with a high order, since it does not have a penalty term to prevent this. However, the model orders which are relatively higher than the most proper one (i.e., the expected number of targets) also tend to retrieve good-fitted solutions.

Generally, the model orders below the expected number of targets cause underfitting. On the other hand, excluding the fourth case of study, the model order immediately above the expected number of targets normally cause overfitting. This means that the probability of error increases within this range of values, since the only admissible choice is the expected number of targets.

Consequently, seeking a better performance for KL, we recommend restricting the lower limit of the range of possible model orders. For the given simulations, we consider only those model orders above five as possible options for KL. This setting helps achieving better overall measurements, especially in regard to the detection rate, improved by about 10 % in our case. We set the lower limit to five, since it is the maximum expected number of targets among all cases of study and most of the model orders above retrieve satisfactory measurements. However, this setting has the drawback of removing plausible options for the cases of study with fewer targets (less than five targets). The latter may translate into attaining worse detection rates for such cases of study, as confirmed in Table III.

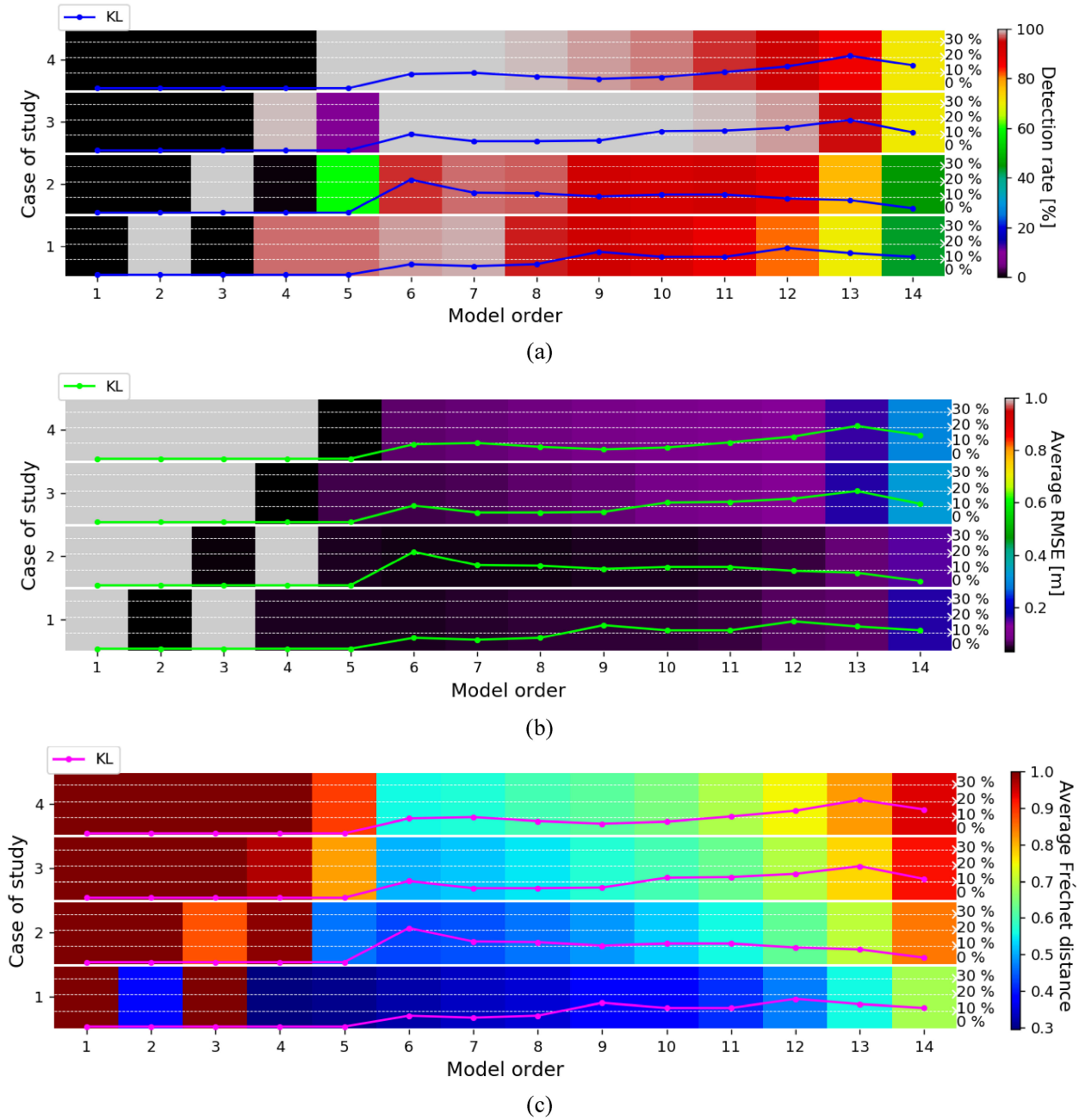

Fig. 4. Colored matrices are read row by row, one case of study at a time, comparing the measurements obtained for all $(L-1)$ model orders. The color bars at the right hand side indicate the attained values with respect to the specified metric. (a) Detection rate. (b) Average RMSE. (c) Average FD. The curve superimposed onto each row of the colored matrices, show the percentage of times that the MOS rule (i.e., KL) chose a particular model order for the given case of study. The summation of all the percentages along each row equals 100%.

Removing the two largest model orders also helps accomplishing a better performance, since they retrieve poorer measurements. Nevertheless, restricting the lower limit is more important. The reported results in Table III show that, even when the largest model orders are included, the treated MOS rule performs successfully. MUSIC, in Table I, attains a detection rate above 85% and an average RMSE below 0.15 m in all cases of study, which demonstrates the high accuracy achieved when KL in (22) is incorporated to MUSIC and the lower limit is restricted. The FD augments in the cases of study with more targets; this is expected, since the more elaborated the PSP is the more challenging it is to estimate. The reported average processing time (of about 1 s) is gotten after processing all possible $(L-1)$ model orders, necessary to construct the colored matrices in Fig. 4. This value is reduced by shortening the model orders' range.

B. *Results Obtained by MARIA*

As observed in Fig. 5, MARIA clearly achieves finer resolution in comparison to MUSIC, but requiring, in most cases, more processing time, as verified in Table III. All addressed

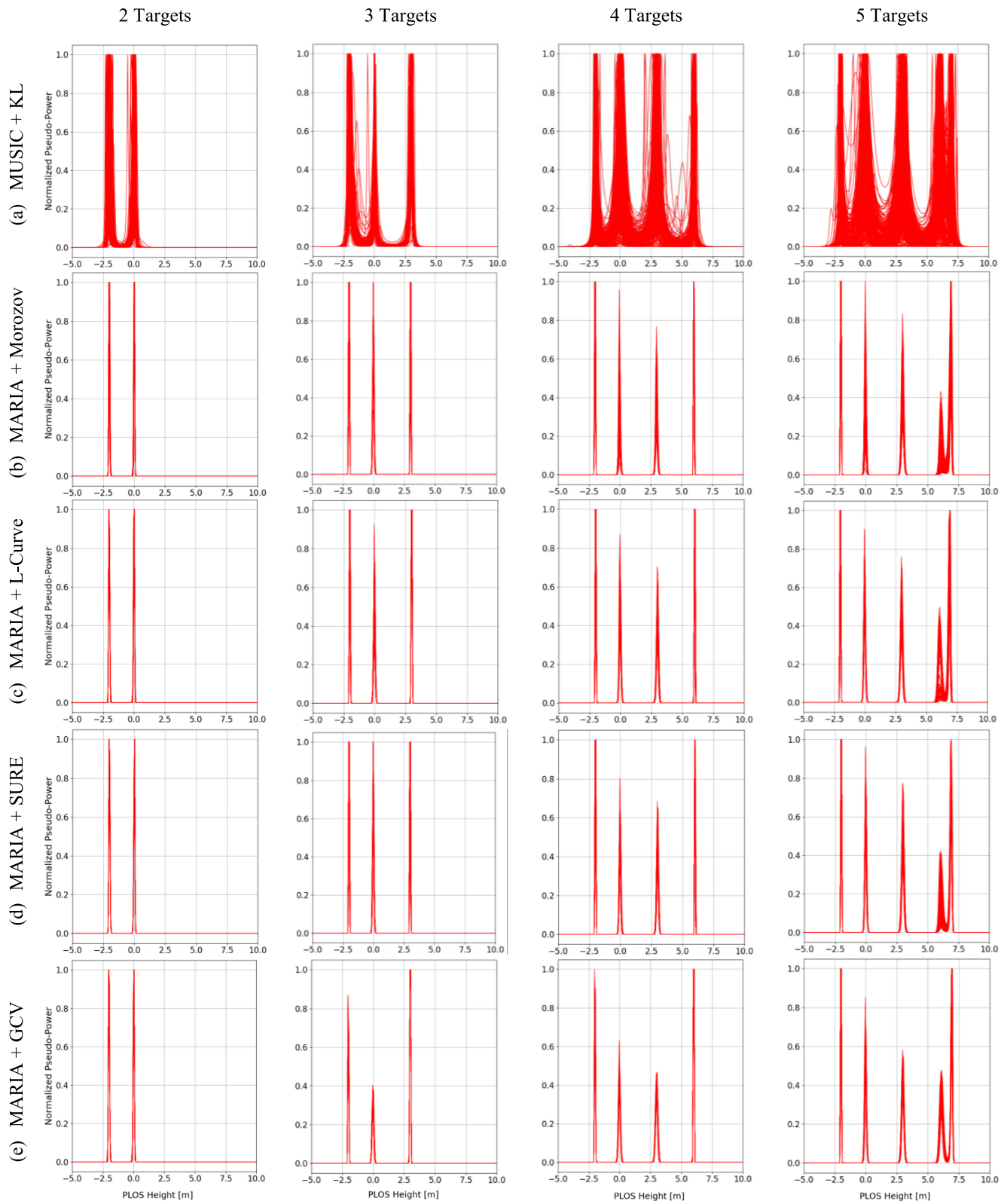

Fig. 5. Superimposed vertical profiles of the retrieved PSP. From left to right: First, second, third, and fourth case of study, as in Table II. We apply MUSIC to perform focusing after incorporating the treated MOS rule (a) KL, whereas, we make use of MARIA for the same purpose, but incorporating the addressed strategies for choosing regularization parameters: (b) Morozov, (c) L-Curve, (d) SURE, and (e) GCV.

TABLE III
SIMULATION RESULTS[1, 2]

			2 Targets	3 Targets	4 Targets	5 Targets
MUSIC	KL	Average Processing Time [s]	0.8841	0.8935	0.9518	0.8723
		Detection Rate [%]	87	89	96	92
		Average RMSE [m]	0.0678	0.0469	0.1327	0.1396
		Average FD	0.5823	0.5134	0.7306	0.7720
MARIA	Morozov	Average Processing Time [s]	0.8741	0.9224	0.9202	0.9557
		Detection Rate [%]	100	100	100	93
		Average RMSE [m]	0.0352	0.0335	0.0290	0.0939
		Average FD	0.3913	0.3828	0.6180	0.8966
	L-Curve	Average Processing Time [s]	2.0344	2.0419	2.0124	2.0288
		Detection Rate [%]	100	100	100	100
		Average RMSE [m]	0.0356	0.0356	0.0303	0.0659
		Average FD	0.3835	0.4467	0.5107	0.7054
	SURE	Average Processing Time [s]	4.2965	4.2708	4.2716	4.3039
		Detection Rate [%]	100	100	100	100
		Average RMSE [m]	0.0357	0.0341	0.0292	0.0622
		Average FD	0.3851	0.3860	0.5204	0.8092
	GCV	Average Processing Time [s]	4.2380	4.3019	4.3343	4.2901
		Detection Rate [%]	100	100	100	100
		Average RMSE [m]	0.0357	0.0371	0.0325	0.0634
		Average FD	0.3910	0.6648	0.6078	0.7390

[1]From left to right: first, second, third and fourth case of study. [2]The average processing time refers to the retrieval of a single vertical profile; the computations are performed in an Intel Xeon Gold 6154 CPU at 3.70 GHz, using a single thread.

parameter selection criteria (i.e., Morozov, L-Curve, SURE, and GCV) are *a posteriori*; they require of a collection of test values $\{\xi_n\}_{n=1}^{N}$ in order to pick an adequate regularization parameter. This explains the higher processing time, besides that MARIA is an iterative method. In the reported simulations, we set to 10 the maximum number of iterations that MARIA can reach. In the case of SURE and GCV, we define a search range from 10^{-1} to 10^{-9}, whereas for the L-Curve from 10^{-1} to 10^{-8}, and for Morozov from 10^{1} to 10^{-8}. The different search ranges are defined seeking to attain the best overall performance.

The L-Curve method chooses ξ as the regularization parameter found as near to the corner as possible along the curve constructed from (30). With the aim of finding this position, we utilize the search algorithm proposed in [29]. To better understand how the algorithm works, we present an example in Fig. 6, gotten from a single realization of the fourth case of study in Table II. Observe that the location of the L-Curve's corner in Fig. 4(a) corresponds to the largest positive curvature in Fig. 4(b), calculated via (31). Such search algorithm reduces considerably the amount of test values $\{\xi_n\}_{n=1}^{N}$, attaining less computation time in contrast to SURE and GCV.

In order to select a proper value for ξ automatically, not in a graphical manner, the use of a search algorithm is indispensable for the L-Curve method. This is not the case for Morozov, SURE, and GCV, which only involve finding a minimum value. Nevertheless, the results gotten for the L-Curve confirm the advantages of utilizing a search algorithm. Consequently, we

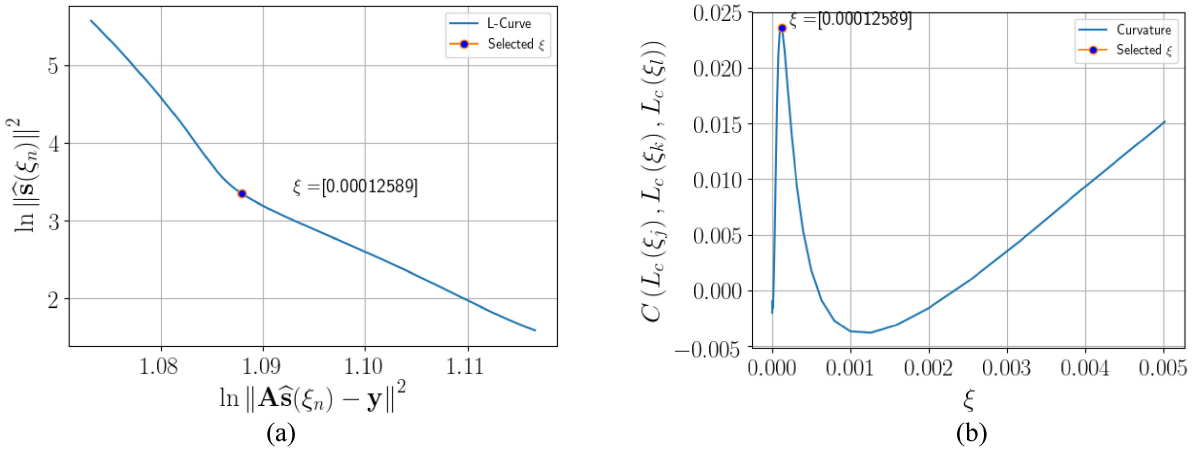

Fig. 6. (a) L-Curve's corner location using the golden section search algorithm described in Section IV. (b) Curvature value associated to the retrieved position.

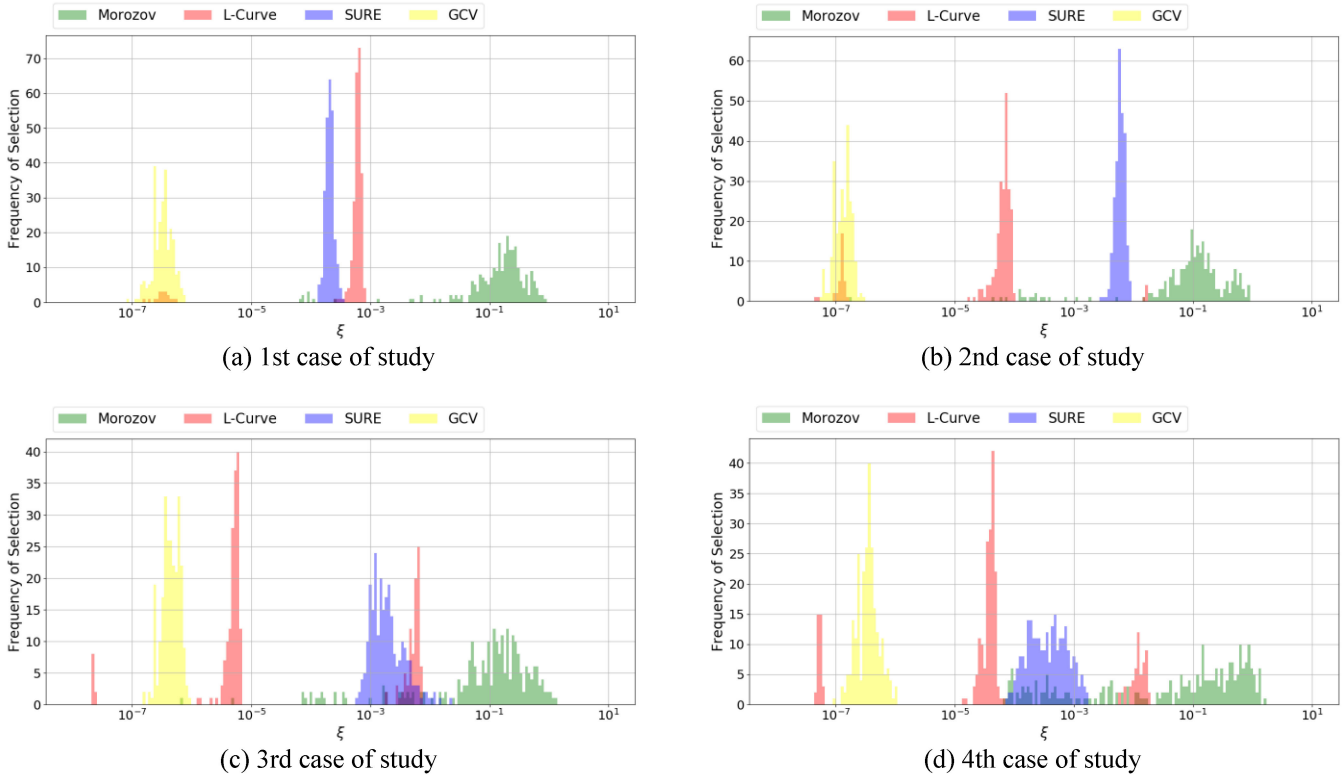

Fig. 7. Frequency of selection for ξ among the different possible assigned values and for each case of study in Table II: (a) First, (b) second, (c) third, and (d) fourth case of study.

employ the golden search algorithm [31] in the other three competing strategies. Different search algorithms (e.g., bisection search [31], line search [34], polynomial-based techniques [31], etc.) could be specifically adapted to Morozov, SURE, and GCV in order to accomplish a better performance, e.g., less computation time. However, this subject is out of the scope of the article, since it requires of dedicated work. The main goal of the article is providing a methodology explicitly adapted to tackle with the TomoSAR problem. More efficient implementations of this methodology are to be addressed in the future.

Morozov in (29) selects a proper value of ξ without computing the PSP [necessary to construct the solution operator $\mathbf{F}_{\mathrm{BMR}}$ in (27)], being, due to this reason, faster than the other competing techniques. As observed in Table III, all parameter selection criteria perform satisfactory for the first, second, and third cases of study in Table II. The main differences between strategies take place in the fourth case of study. Morozov appears as the less reliable technique, whereas the L-Curve, SURE, and GCV are just as reliable. Take into account, nonetheless, that using different search ranges, may enhance the performance of Morozov.

As observed in Fig. 7, all treated parameter selection strategies tend to choose different values of ξ among simulations. This means that MARIA is sufficiently robust to correctly solve

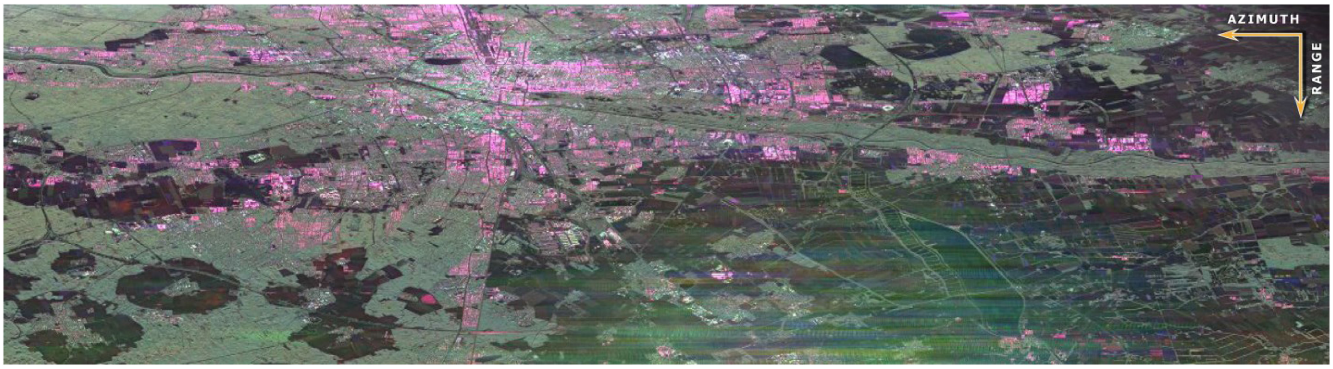

Fig. 8. SLC SAR image of the test site in Munich, Germany, 2015 (near range on top). The colors correspond to the channels HH (red), VV (blue), and HV (green).

the TomoSAR problem using different values of ξ. Yet, the regularization parameter must not be arbitrarily chosen. Not all values of ξ are adequate for MARIA, as we can confirm in the fourth case of study in Table III.

The TomoSAR problem is defined within the DOA estimation theory. Therefore, the principal goal of MARIA, together with the incorporated parameter selection criteria, is to accomplish satisfactory detection rates and RMSE measurements. As verified in Table III, these values are consistent among methods. However, different regularization parameters make MARIA retrieve different values of pseudo-power. This behavior is more notorious for GCV, in the second and third cases of study, as observed in Fig. 5 and through the reported FD in Table III. This means that, although MARIA is able to correctly solve the TomoSAR problem using different regularization parameters, they lead to different physical interpretations, i.e., a different attained pseudo-power.

As seen in Fig. 5, after convergence (or by reaching the maximum number of iterations $i = I$) MARIA recovers sparse vertical profiles. A solution operator $\mathbf{F}_{\mathrm{BMR}}^{[i=I]}$ in (27), which attains this kind of results, may cause under/overfitted solutions when applied to the addressed parameter selection tools. The PSP is so refined that the approximations performed by the parameter selection strategies may result highly inaccurate. To tackle with this issue, it is advisable to rather employ a solution operator $\mathbf{F}_{\mathrm{BMR}}$ constructed after the first iteration $i = 1$. In the reported simulations, we apply Capon beamforming [1], [20], to provide MARIA with the initial input to (26). Next, we make use of the solution operator $\mathbf{F}_{\mathrm{BMR}}^{[i=1]}$, gotten after the first iteration, to compute the different parameter selection strategies. Later on, after choosing ξ, MARIA is applied normally.

Finally, based on the explanations provided previously and based on the reported simulations, we conclude the following.

1) The provided explicit implementations of the Morozov's discrepancy principle in (29), the L-Curve in (30), SURE in (36), and GCV in (37) are suitable for the TomoSAR problem.
2) MARIA tends to attain finer resolution with the L-Curve, SURE, and GCV, in contrast to Morozov. Morozov retrieves a worse detection rate in Table III for the fourth case of study in Table II, where the last two targets are more difficult to discriminate due to their close proximity.
3) Morozov and SURE are only applicable when the noise variance is known *a priori*, since it is required for their computations. Conversely, L-Curve and GCV can be applied in a more general basis, as they do not require of such information.
4) Although L-Curve and GCV retrieve similar metrics (i.e., detection rate, RMSE, and FD) in Table III, L-Curve requires less computation time.
5) Taking the latter into consideration, we suggest using L-Curve as the common standard for MARIA. However, when the noise variance is known *a priori*, we recommend using Morozov, which trades off a minor decrease in the detection rate for much less processing time.

We make use of L-Curve for the experimental results reported in Section VI, since we have no *a priori* knowledge on the noise variance.

VI. Experimental Results

In 2015, the UAVSAR system of the Jet Propulsion Laboratory (JPL) and the National Aeronautics and Space Administration (NASA) was flown in order to acquire L-band fully polarimetric TomoSAR data collections from Munich, the third largest city in Germany [5], [35]. The utilized Gulfstream G-III aircraft flew at a nominal altitude of 12.5 km with a swath of 22 km and length of 60 km. The incidence angles range from 25° to 65°. For the specified microwave frequency band, with 0.24 m wavelength and 80 MHz chirp bandwidth, the resultant SLC imagery has a resolution of 1.66 m in range and 0.8 m in azimuth. Fig. 8 shows one SLC image out of the stack; the presence of radio frequency interference is due to the several external sources, by instance, those coming from the Munich's airport.

The TomoSAR acquisition geometry consists of seven passes at different altitudes, as specified in Table IV. These were completed on a heading of 193°. The expected vertical Fourier resolution is of about 2.8 m in the near range and of about 6 m in the far range.

For demonstration purposes, we define the area where the Maximilianeum edifice is located as our ROI. Fig. 9 shows the corresponding intensity images, with respect to the master track, for all polarizations (HH, HV, and VV). The azimuth

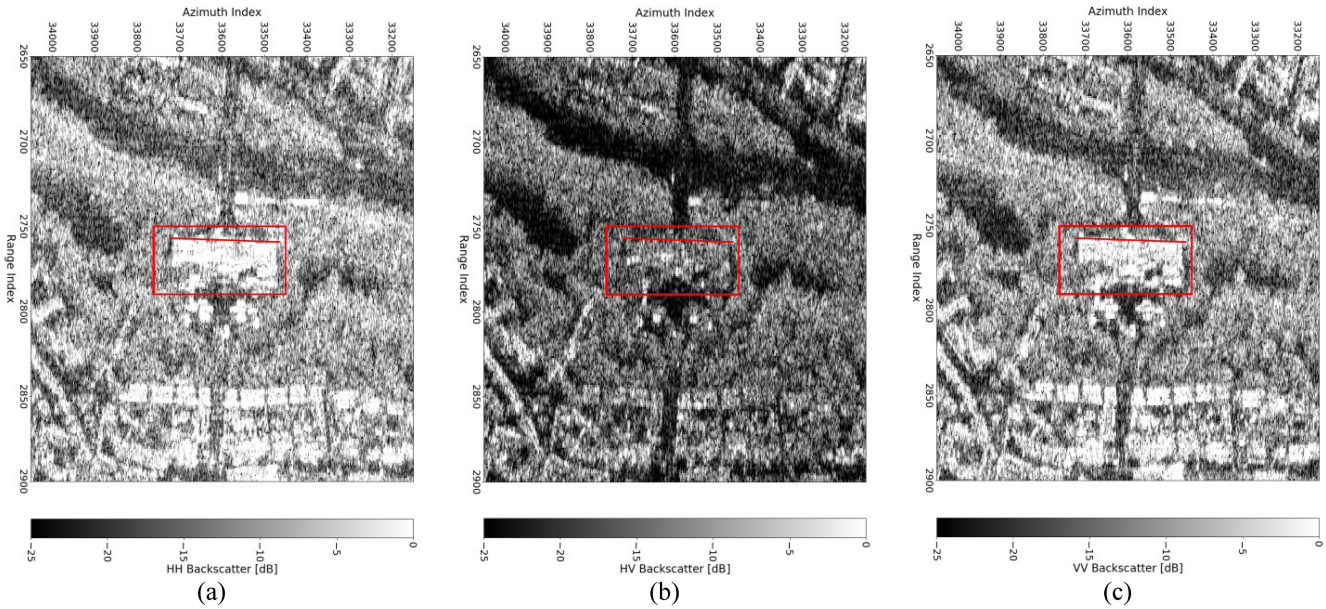

Fig. 9. Quick look intensity images of the ROI for all polarizations. (a) HH. (b) HV. (c) VV. The ROI is specified through a red rectangle, whereas the tomograms presented afterward correspond to the red line crossing the ROI.

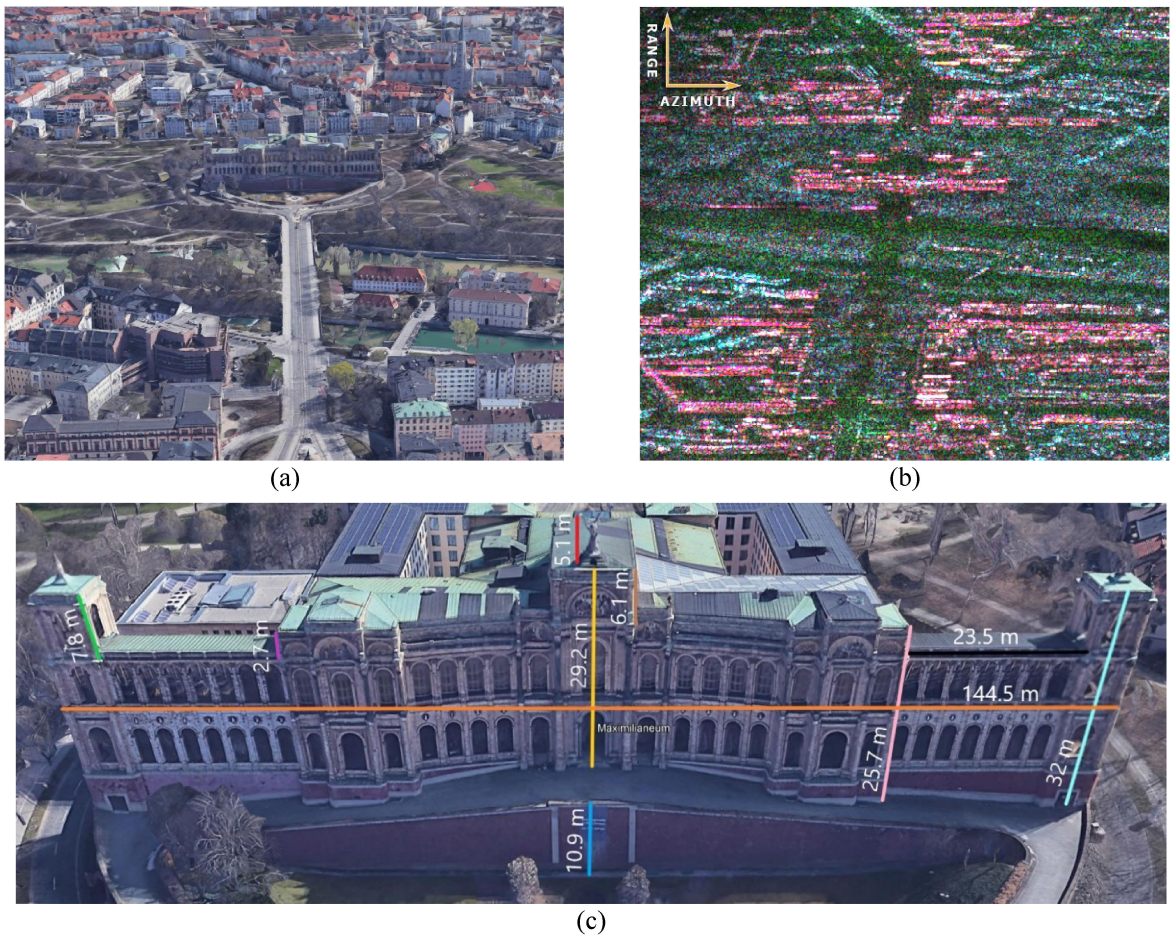

Fig. 10. (a) Google Earth image of the test region, where the Maximilianeum building is located. (b) Polarimetric SLC SAR image of the test area [the colors correspond to the channels HH (red), VV (blue), and HV (green)]. (c) Front view of the Maximilianeum edifice (Google Earth), specifying the height of the structures that constitute it.

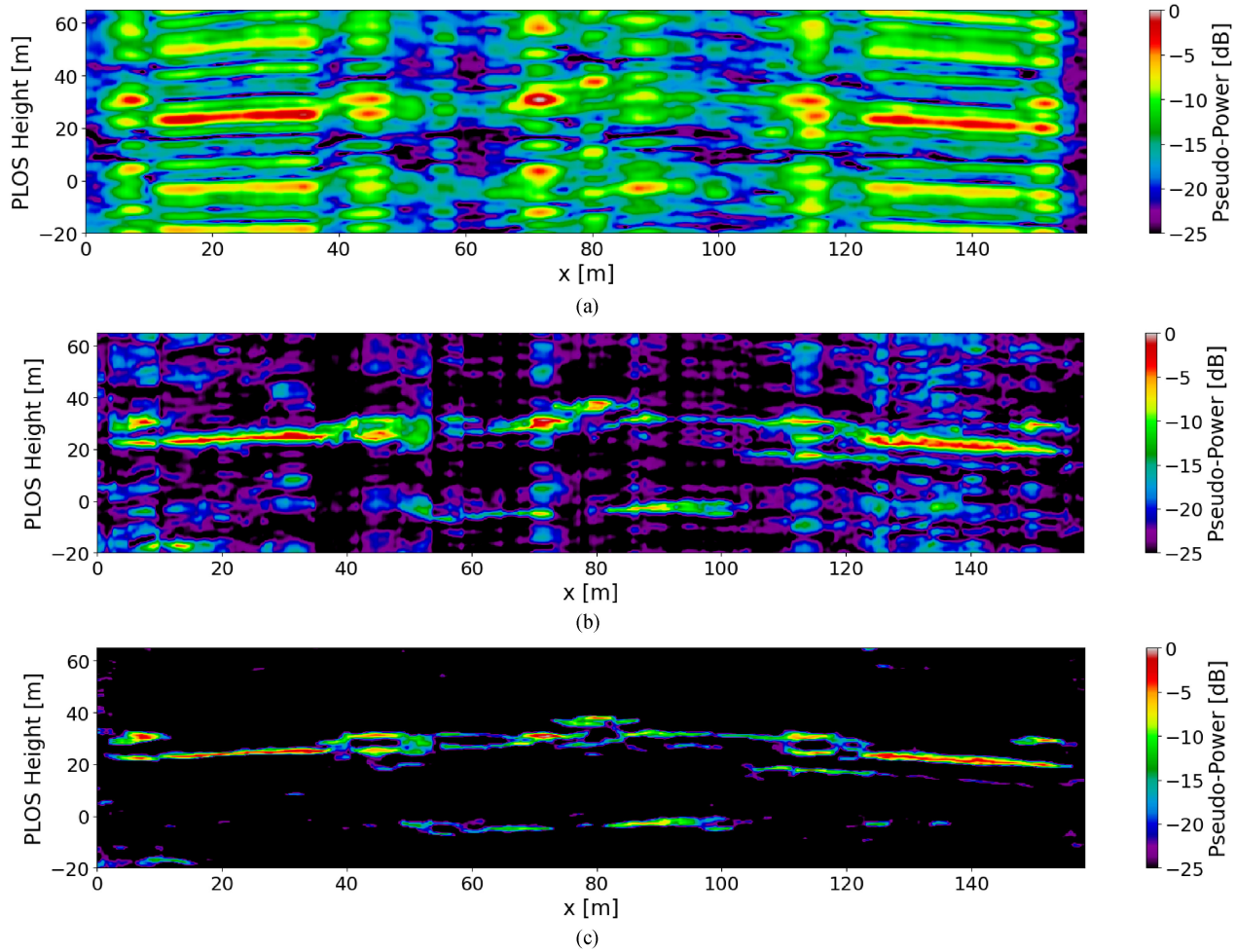

Fig. 11. HH tomograms retrieved from the area depicted by the red line crossing the ROI specified in Fig. 9. We perform focusing using (a) matched filtering, (b) MUSIC + KL, and (c) MARIA + L-Curve.

TABLE IV
TOMOSAR ACQUISITION GEOMETRY

Track	Flight altitude [m]
1	12500
2	12500 + 30
3	12500 + 90
4	12500 + 160
5	12500 + 240
6	12500 + 400
7	12500 + 600

and range indices are displayed, acting as a guide to identify the azimuth and range bounds of the ROI, which is specified through a red rectangle. Note that the Maximilianeum is oriented practically parallel to the flight direction. The tomograms presented afterward refer to the red line crossing the ROI, spanning about 160 m along azimuth. Fig. 10(a) shows the Google Earth image of the test region, whereas Fig. 10(b) shows the respective polarimetric SLC SAR image [the colors correspond to the channels HH (red), VV (blue), and HV (green)]. Fig. 10(c) specifies the height of the different structures constituting the Maximilianeum. The tomographic slices presented afterward correspond to this edifice.

Multilooking is performed on the set of data covariance matrices through Boxcar filtering, using a 5×10 (range/azimuth) pixel window. As a reference, we first apply matched filtering to focus the multilooked TomoSAR data, i.e., $\hat{\mathbf{b}} = \{\mathbf{A}^{+}\mathbf{Y}\mathbf{A}\}_{\text{diag}}$ [4]. The resultant tomograms are presented in Figs. 11(a), 12(a), and 13(a), for polarizations HH, HV, and VV, correspondingly. The several structures constituting the edifice in Fig. 10(c) are distinguished, by instance: the two towers at the extremes, both wings (one at each side) and the central building. Conversely, observe the presence of ambiguities above and below the building; these are more notorious in Fig. 11(a), between the range from 40 to 65 m and between the range from 0 to -20 m, along the PLOS height axis. The low resolution, along with the presence of ambiguity, hampers the interpretation of the results.

Subsequently, we retrieve the tomographic slices from the exact same region as in Fig. 9, but using MUSIC + KL and MARIA + L-Curve instead. These results verify their feature enhancing capabilities, namely, suppression of artifacts, ambiguity levels

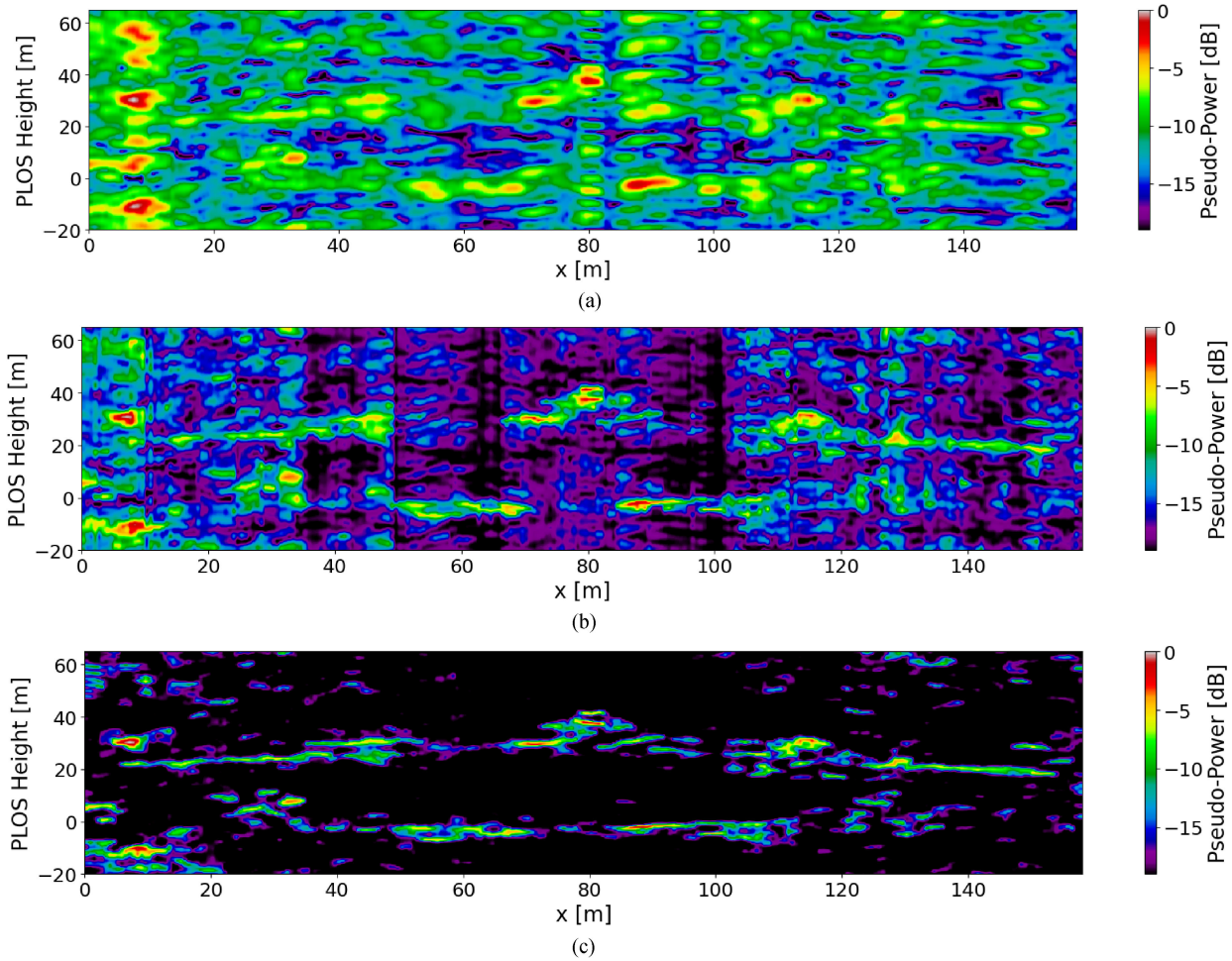

Fig. 12. HV tomograms retrieved from the area depicted by the red line crossing the ROI specified in Fig. 9. We perform focusing using (a) matched filtering, (b) MUSIC + KL, and (c) MARIA + L-Curve.

reduction, and increased resolution. In order to better appreciate such capabilities, and since most super resolution techniques (e.g., MUSIC and MARIA) do not preserve radiometric accuracy, all tomograms shown hereafter are normalized with respect to the (pseudo) power recovered using matched filtering, which is known to be more accurate in this aspect. The tomographic slices are presented in a dB scale, where 0 dB refers to the peak obtained with matched filtering.

Figs. 11(b), 12(b), and 13(b) show the tomograms recovered using MUSIC for polarizations HH, HV, and VV, respectively. We make use of KL in (22) to select the model orders. As discussed in Section V, we remove the first (two) possible model orders, such that KL mainly selects the larger model orders. Fig. 11(c), 12(c), and 13(c) depict the tomograms obtained with MARIA for polarizations HH, HV, and VV, correspondingly. The L-Curve method is employed to select the involved regularization parameters. We use Capon beamforming [1], [20], to retrieve the initial input to MARIA. In order to apply the L-Curve technique, we refer to the solution operator $\mathbf{F}_{\mathrm{BMR}}^{[i=1]}$ in (27) obtained after the first iteration. Later on, we refer to the chosen ξ to compute MARIA normally.

Next, for an easy assessment, Fig. 14 presents the superimposed vertical profiles for each azimuth position within the displayed tomograms; the (pseudo) power is presented in a linear scale. Through the comparison between the recovered tomograms and the corresponding superimposed vertical profiles, we can observe that, as expected, MARIA attains finer resolution than MUSIC; nonetheless, MARIA requires of more computation time, as seen in Table V.

For comparison purposes, Table V includes the processing time attained by MARIA for all related parameter selection tools. The noise variance was set to an arbitrary value for those techniques that require it. The reported measurements refer to the time needed to process the entire ROI defined in Fig. 9. Note that the computation time of MUSIC decreased drastically with respect to the reported simulations, since only five possible model orders are considered.

Fig. 15 depicts the renders of the area where the Maximilianeum edifice [Fig 10(c)] is located, obtained after applying the addressed focusing techniques, namely, matched filtering, MUSIC and MARIA. The 3-D representation gathers 40 consecutive tomograms, spanning a volume of 65 $\times$ 190 $\times$60 m, as specified

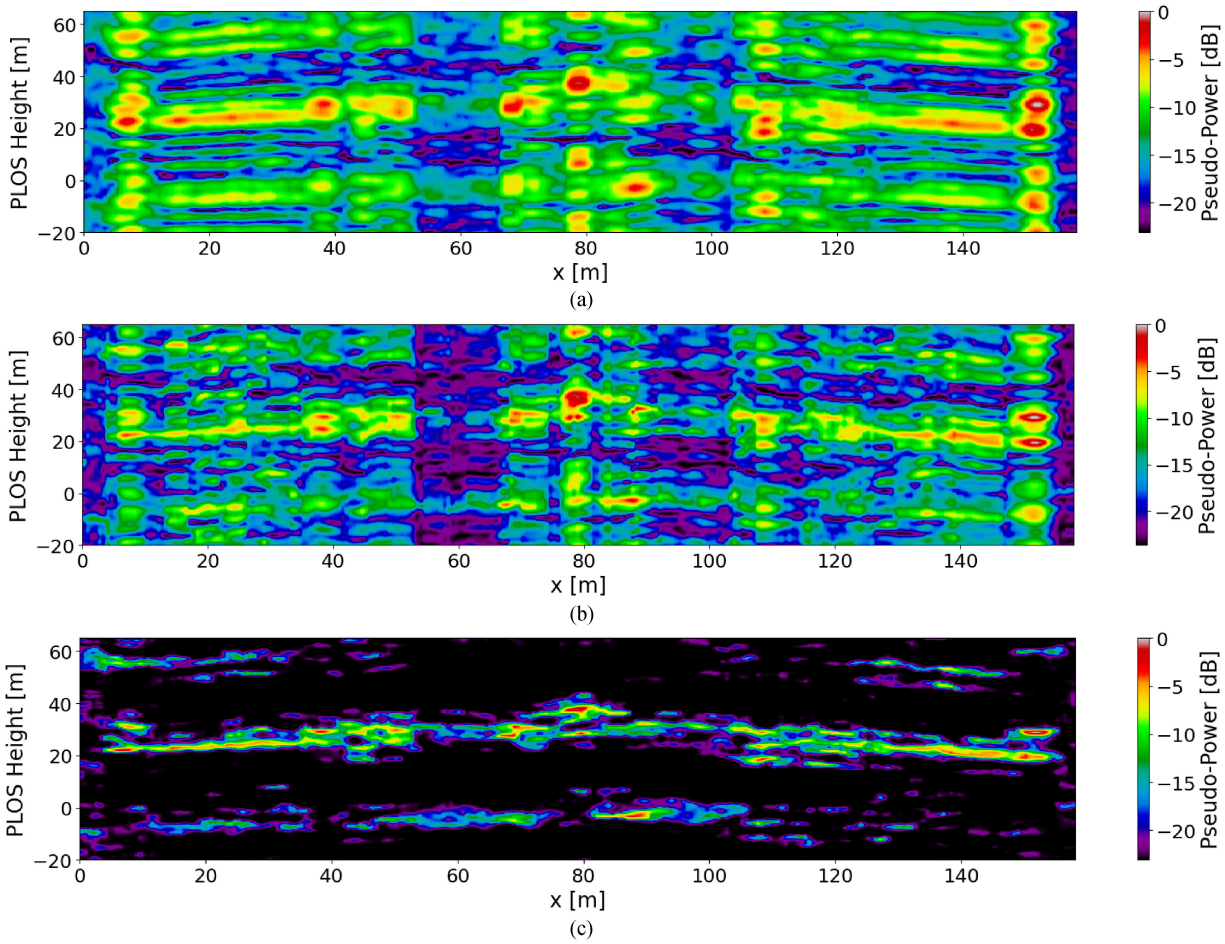

Fig. 13. VV tomograms retrieved from the area depicted by the red line crossing the ROI specified in Fig. 9. We perform focusing using (a) matched filtering, (b) MUSIC + KL, and (c) MARIA + L-Curve.

TABLE V
PROCESSING TIME IN SECONDS[1]

MUSIC	**KL**	160.470
MARIA	**Morozov**	595.039
	L-Curve	3147.028
	SURE	5036.489
	GCV	5114.055

[1]The computations are performed in an Intel Xeon Gold 6154 CPU at 3.70 GHz, using four threads.

in Fig. 13. Proper histogram equalization has been performed, in order to facilitate the identification of the different structures.

Most of the structures constituting the Maximilianeum in Fig. 10(c), are sufficiently spaced to be discriminated via matched filtering. Consider Figs. 11–13; taking matched filtering as a reference, note that all of these structures remain in the results gotten through MUSIC and MARIA; furthermore, no additional (strange) structures have been added. Observe also the resemblance between the tomograms recovered by MUSIC and MARIA, both accomplish similar solutions, with MARIA attaining finer resolution. Consequently, we can infer that both MUSIC and MARIA, retrieve good-fitted reconstructions. The main advantage of using MUSIC and MARIA in this case, has to do with the reduction of the ambiguity levels and the retrieval of refined solutions, which eases identifying the local maxima and finding their locations. MARIA achieves better ambiguity levels reduction than MUSIC; both techniques perform worse for the VV polarization and better for the HH polarization.

The presented experimental results verify then the successful implementation of the treated parameter selection criteria (i.e., KL and L-Curve), specifically adapted to deal with the TomoSAR problem. These parameter selection tools have been respectively incorporated into two focusing techniques (i.e., MUSIC and MARIA), with the aim of demonstrating their correct functioning. Yet, the addressed methodology can be easily adapted to other focusing techniques, which require the correct setting of their parameters.

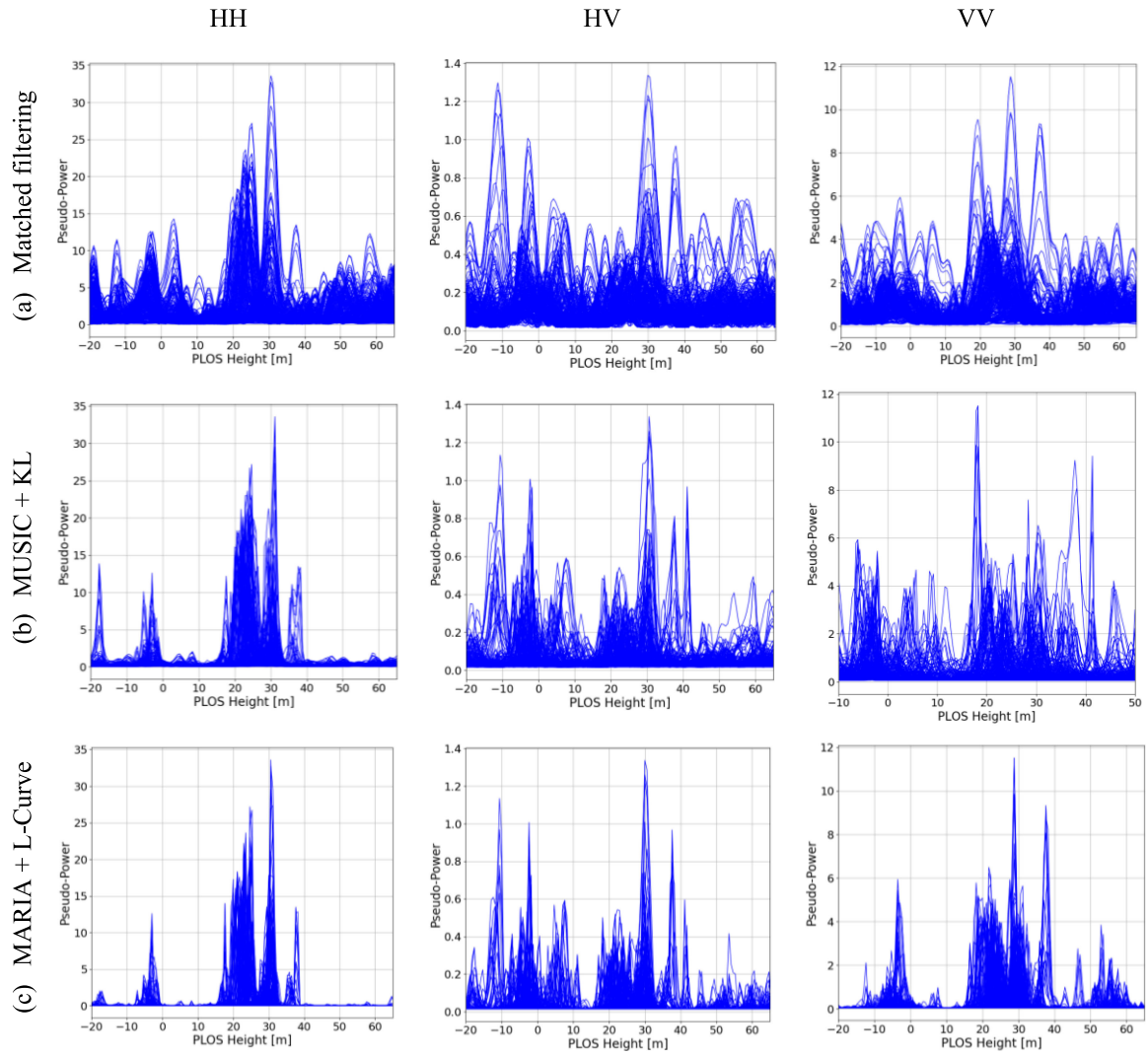

Fig. 14. Superimposed vertical profiles of the tomograms displayed in Figs. 11–13. (a) Matched filtering. (b) MUSIC + KL. (c) MARIA + L-Curve. From left to right: HH, HV, and VV. The (pseudo) power is presented in a linear scale.

VII. Conclusion

Tackling the TomoSAR problem within the context of DOA, allows using super resolved focusing techniques that achieve suppression of artifacts and ambiguity levels reduction. This methodology includes parametric techniques as MUSIC and ESPRIT, and iterative statistical regularization methods as MARIA. Although these techniques offer attractive advantages, they depend on the proper selection of their parameters to guarantee best performance. Otherwise, there is risk of retrieving under/overfitted solutions, distorting our perception of the actual signal.

Several popular parameter selection strategies exist in the literature. However, these criteria must be specifically adapted to the particular problem. The TomoSAR nonlinear inverse problem is especially challenging; the constraints on the number of passes composing the acquisition geometry, along with the large number of samples employed to describe the PSP (in order to take full advantage of the super resolution capabilities of the addressed focusing techniques) make the TomoSAR problem highly ill-conditioned. Accordingly, the main contribution of this article consists in providing the explicit implementations of different parameter selection strategies, specifically adapted to the TomoSAR case. For such a purpose, we refer to a MOS rule, based on ML, to determine the model order of parametric focusing techniques as MUSIC or ESPRIT. On the other hand, we make use of strategies as the Morozov's discrepancy principle, the L-Curve, SURE, and GCV for the proper selection of regularization parameters, necessary for focusing techniques as MARIA.

The ML method of parameter estimation, originally designed to deal with linear problems, has been adapted to the TomoSAR problem. Subsequently, a MOS rule, based on the KL information criterion, is introduced.

The addressed strategies for choosing regularization parameters are also originally designed for linear problems, nonetheless, having a solution operator permits extending them to tackle with the TomoSAR problem. The solution operator allows solving both linear and nonlinear problems.

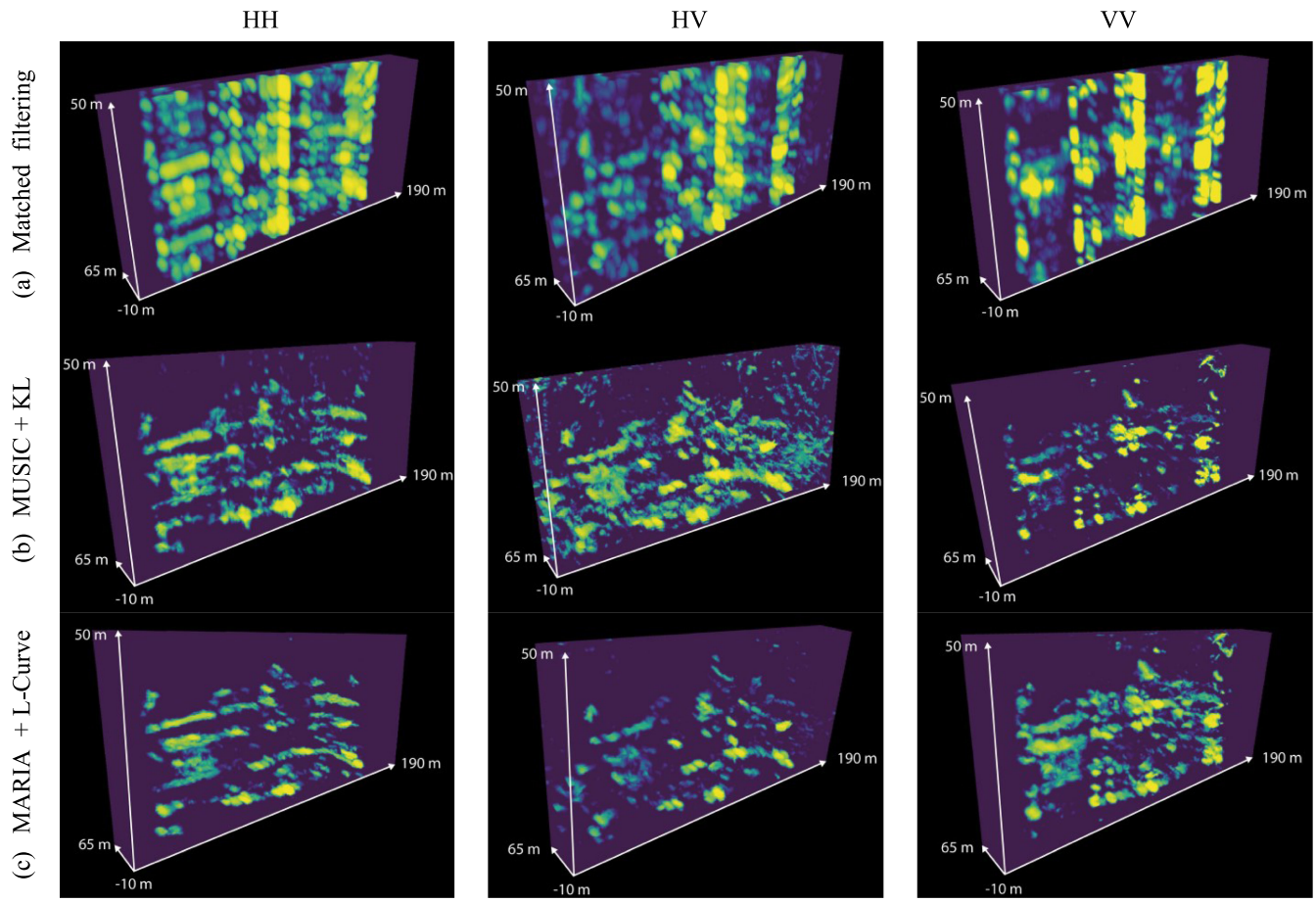

Fig. 15. Renders of the area where the Maximilianeum edifice [Fig 10(c)] is located, obtained after applying: (a) Matched filtering, (b) MUSIC + KL, and (c) MARIA + L-Curve. From left to right: HH, HV, and VV. The 3-D representation gathers 40 consecutive tomograms.

Therefore, the presented criteria for selecting regularization parameters are only applicable to those focusing techniques described through a solution operator, by instance, CLS and MARIA. Two out of the four treated related tools, i.e., Morozov and SURE, require the *a priori* knowledge of the noise variance. The remaining two, i.e., L-Curve and GCV, acquire all necessary information from the measurements themselves. Morozov and SURE give more weight to the perturbations due to the additive noise, rather than to those due to the signal formation. Nevertheless, for the TomoSAR ill-conditioned inverse problem, most of the perturbations in the observed data are due to the formation of the received signal. Consequently, for practical applications, we expect a more consistent behavior from the L-Curve and GCV.

Contrary to the other tools, the L-Curve obliges using a search algorithm to find a proper regularization parameter. For such aim, we make use of a golden section search algorithm, which estimates the local curvature of the L-Curve from three sampled points. The processing time gotten for the L-Curve, in contrast to SURE and GCV, confirm the advantages of utilizing a dedicated search algorithm. Subsequently, we recommend exploring the usage of dedicated search algorithms for the other addressed parameter selection strategies. This subject, however, is to be treated in the future, since it requires of dedicated work.

The successful implementation of the treated parameter selection criteria, is verified trough simulations and experimental results. Relevant pointers and suggestions, consequence of the gained experience, are provided. With the aim of demonstrating a correct functioning, the addressed parameter selection strategies have been respectively incorporated into two focusing techniques (i.e., MUSIC and MARIA). Still, the addressed methodology can be easily adapted to other focusing techniques, which require the correct setting of their parameters. By instance, after minor modifications, KL can be also incorporated into ESPRIT, whereas, Morozov, the L-Curve, SURE, and GCV can be incorporated into those techniques having a solution operator.

The simulations and experimental results show that MARIA achieves finer resolution than MUSIC, at the expense of a higher processing time. The fact that MUSIC has a finite number of possible model orders, defined by the number of passes, makes it faster by nature. Generally, MUSIC retrieves satisfactory results at a lower computational cost. Thus, depending on the application, the user may select less processing time via MUSIC or attaining finer resolution through MARIA.

ACKNOWLEDGEMENT

The authors would like to thank Dr. S. Hensley from the JPL/NASA, for providing the UAVSAR data utilized in the reported experiments.

REFERENCES

[1] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, vol. 1, Upper Saddle River, NJ, USA: Prentice hall, 2005.
[2] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 2004.
[3] G. Martín del Campo, M. Nannini, and A. Reigber, "Statistical regularization for enhanced TomoSAR imaging," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1567–1589, Feb. 2020.
[4] G. Martín del Campo, M. Nannini, and A. Reigber, "Towards feature enhanced SAR tomography: A maximum-likelihood inspired approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1730–1734, Nov. 2018.
[5] G. D. Martín-del-Campo-Becerra, A. Reigber, M. Nannini, and S. Hensley, "Single-Look SAR tomography of urban areas," *Remote Sens.*, vol. 12, Aug. 2020, Art, no. 2555.
[6] G. D. Martín del Campo, Y. V. Shkvarko, A. Reigber, and M. Nannini, "TomoSAR imaging for the study of forested areas: A virtual adaptive beamforming approach," *Remote Sens*, vol. 10, Nov. 2018, Art. no. 1822.
[7] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosc. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.
[8] M. Nannini, R. Scheiber, and A. Moreira, "Estimation of the minimum number of tracks for SAR tomography," *IEEE Trans. Geosc. Remote Sens.*, vol. 47, no. 2, pp. 531–543, Feb. 2009.
[9] M. Nannini, A. Reigber, and R. Scheiber, "A study on irregular baseline constellations in SAR tomography," in *Proc. 8th Eur. Conf. Synthetic Aperture Radar*, 2010, pp. 1150–1153.
[10] M. Schmitt and U. Stilla, "Maximum-likelihood-based approach for single-pass synthetic aperture radar tomography over urban areas," *IET Radar, Sonar Navigation*, vol. 8, no. 9, pp. 1145–1153, Dec. 2014.
[11] M. Nannini, R. Scheiber, R. Horn, and A. Moreira, "First 3-D reconstructions of targets hidden beneath foliage by means of polarimetric SAR tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 60–64, Jan. 2012.
[12] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
[13] R. Giryes, M. Elad, and Y. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407–422, May 2011.
[14] J. L. Mueller and S. Siltanen, *Linear and Nonlinear Inverse Problems With Practical Applications*. vol. 10, Philadelphia, PA, USA: Siam, 2012.
[15] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-Curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.
[16] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sept. 2008.
[17] S. Ramani, Z. Liu, J. Rosen, J. Nielsen, and J. A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3659–3672, Aug. 2012.
[18] P. Craven and G. Wahba, "Smoothing noisy data with spline functions," *Numerische Mathematik*, vol. 31, no. 4, pp. 377–403, Dec. 1978.
[19] H. H. Barrett and K. J. Myers, *Foundations Image Sci.*, New York, NY, USA: Wiley, 2004.
[20] F. Gini, F. Lombardini, and M. Montanari, "Layover solution in multibaseline SAR interferometry," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1344–1356, Oct. 2002.
[21] Y. Huang, L. Ferro-Famil, and A. Reigber, "Under-Foliage object imaging using SAR tomography and polarimetric spectral estimators," *IEEE Trans. Geosc. Remote Sens.*, vol. 50, no. 6, pp. 2213–2225, Jun. 2012.
[22] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging*. Boca Raton, FL, USA: CRC press, 1998.
[23] F. Lombardini and F. Gini, "Model order selection in multi-baseline interferometric radar systems," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 20, pp. 3206–3219, 2005.
[24] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33 no. 2, pp. 387–392, Apr. 1985.
[25] Y. V. Shkvarko, "Unifying regularization and Bayesian estimation methods for enhanced imaging with remotely sensed data—Part I and Part II," *IEEE Trans. Geosc. Remote Sens.*, vol. 42, no. 5, pp. 923–940, May 2004.
[26] Y. V. Shkvarko, "Unifying experiment design and convex regularization techniques for enhanced imaging with uncertain remote sensing data—Part I and Part II," *IEEE Trans. Geosc. Remote Sens.*, vol. 48, no. 1, pp. 82–111, Jan. 2010.
[27] G. Strang. *Introduction to Linear Algebra*, 6th ed. Cambridge, MA, USA: Wellesley-Cambridge, 2016.
[28] P. C. Hansen, "Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems," *Numer. Algorithms*, vol. 6, no no. 1, pp. 1–35, 1994.
[29] A. Cultrera and L. Callegaro, "A simple algorithm to find the L-Curve corner in the regularization of inverse problems," *IOP SciNotes*, vol. 1, no. 2, pp. 1–6, Aug. 2020.
[30] J. C. Leger, "Menger curvature and rectifiability," *Ann. Math.*, vol. 149, pp. 831–869, 1999.
[31] A. Belegundu and T. Chandrupatla, *Optimization Concepts and Applications in Engineering*, Upper Saddle River, NJ, USA: Prentice Hall, 1999.
[32] S. Har-Peled and B. Raichel, "The Fréchet distance revisited and extended," *ACM Trans. Algorithms*, vol. 10, no. 1, Jan. 2014, Art. no. 3.
[33] T. Eiter and H. Mannila, "Computing discrete Fréchet distance," Inf. Syst. Dept., Tech. Univ. Vienna, Vienna, Austria, Tech. Rep. CD-TR 94/64, Apr., 1994.
[34] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 2006.
[35] S. Hensley *et al.*, "UAVSAR tomography of Munich," in *Proc. IGARSS IEEE Int. Geosc. Remote Sens. Symp.*, 2019, pp. 1140–1143.

Gustavo Daniel Martín-del-Campo-Becerra received the engineering degree in electronics and communications engineering from the University of Guadalajara, Guadalajara, Jalisco, Mexico, in 2008, and the M.Sci. and Dr.Sci. (Ph.D. equivalent) degrees in electrical engineering, with specialization in telecommunications, from the Center for Research and Advanced Studies (Cinvestav), National Polytechnic Institute, Guadalajara, Jalisco, Mexico, in 2013 and 2017, respectively.

Since 2017, he has been with the Microwaves and Radar Institute (HR), German Aerospace Center (DLR), Weßling, Germany. His research interests include the applications of signal processing to remote sensing, particularly SAR tomography, inverse problems, random fields estimation, and adaptive spatial analysis.

Sergio Alejandro Serafín-García received the engineering degree in electronics and communications engineering from the University of Guadalajara, Guadalajara, Jalisco, Mexico, in 2017. He is currently working toward the M.Sci. degree in electrical engineering with the Center for Research and Advanced Studies (Cinvestav), National Polytechnic Institute, Guadalajara, Jalisco, Mexico.

His research interests include the applications of signal processing to remote sensing, particularly SAR tomography.

Andreas Reigber (Fellow, IEEE) received the Diploma degree in physics from the University of Konstanz, Konstanz, Germany, in 1997, the Ph.D. degree in engineering from the University of Stuttgart, Stuttgart, Germany, in 2001, and the Habilitation degree from the Berlin University of Technology, Berlin, Germany, in 2008.

He is currently the Head of the SAR Technology Department, Microwaves and Radar Institute, German Aerospace Center, Weßling, Germany, where he is leading the development and operation of state-of-the-art airborne SAR sensors. He is also a Professor of remote sensing and digital image processing with the Berlin University of Technology, Berlin, Germany. His research interests include various aspects of multimodal, multichannel, and high-resolution SAR processing and post-processing.

Dr. Reigber was the recipient of several prize paper awards, among them the IEEE TGRS Transactions Prize Paper Award, in 2001 and 2016 for his works on polarimetric SAR tomography and nonlocal speckle filtering, respectively, and the IEEE TGRS Letters Prize Paper Award, in 2006, for his work on multi-pass SAR processing.

Susana Ortega-Cisneros received the engineering degree in electronics and communications engineering from the University of Guadalajara, Guadalajara, Jalisco, Mexico, in 1990, the M.Sci. degree in solid state electronics from the Center for Research and Advanced Studies (Cinvestav), National Polytechnic Institute, Mexico City, Mexico, in 1995, and the Ph.D. degree in computer science and telecommunications from the Autonomous University of Madrid, Madrid, Spain, in 2005.

Since 2010, she has been with the Center for Research and Advanced Studies (Cinvestav), National Polytechnic Institute, Guadalajara, Jalisco, Mexico. She is involved in the design of digital architectures based on FPGAs, digital signal processors (DSPs), and microprocessors. Her current research interests include digital control, self-timed synchronization, electronic systems applied to biomedicine, embedded microprocessor design, digital electronics, custom DSPs in FPGAs, and, recently, remote sensing applications.