Deriving Digital Surface Models from Geocoded SAR Images and Back-Projection Tomography

Elías Méndez Domínguez, *Member, IEEE*, David Small, *Senior Member, IEEE*, and Daniel Henke

***Abstract*—Digital surface models (DSMs) are sets of elevation data of the Earth's surface, useful for applications such as urban studies and height estimation of buildings. They can be derived from a set of synthetic aperture radar (SAR) images acquired in an interferometric or tomographic configuration. Each image acquisition is usually focused in radar geometry. In this work, we present steps required to derive DSMs from SAR single-look complex (SLC) products focused in map geometry (geocoded). We modified existing tomographic reconstruction techniques to be able to operate with geocoded SLCs and extended methods to operate with 3-D geocoded SLCs. The performance analysis showed that methods using 3-D geocoded SLC products yielded DSMs with fewer outliers and retained more information of the illuminated area, with a cost of higher computational complexity. Compressive sensing methods using 2-D geocoded SLCs can be a good alternative due to their comparatively moderate computational complexity.**

***Index Terms*—Digital surface model (DSM), synthetic aperture radar (SAR), tomography, urban.**

I. Introduction

REMOTE sensing systems enable mapping the Earth's surface as a digital set of elevation data, or digital surface models (DSMs). These models are useful for urban studies and building detection [1]–[6]. DSMs can be derived from optical images, LiDAR or synthetic aperture radar (SAR) data [7]. The operation of LiDAR and cameras depends on the availability of daylight and on weather conditions. In contrast to an airborne laser scanning (ALS) or terrestrial laser scanner (TLS), the side-looking view of SAR offers simultaneously information on the facades and walls of buildings as well as the rooftops and the heights of the objects [8]. In addition, InSAR can provide a global digital elevation model (DEM) of the earth's surface at a lower cost [9], [10]. SAR sensors can provide 3-D imaging by extending the synthetic aperture also in elevation. This configuration is known as interferometric SAR (InSAR) [11] or tomographic SAR (TomoSAR) [12] depending on the aperture length in elevation and number of baselines.

SAR tomography enables topographic mapping after processing a set of images taken at slightly different viewing angles or different times. During the course of the data collection, the illuminated area is assumed to be unchanged. The images are typically 2-D and focused in radar geometry [12] after applying azimuth compression in the frequency domain. This alleviates the computational complexity of the subsequent processing. However, azimuth compression in the frequency domain of data acquired with 1) nonlinear flight paths, 2) strongly varying flight attitude angles (roll, pitch, and heading), or 3) topographic variations have to be performed in a block-wise manner [13]. Block-wise pulse compression involves a degraded azimuth resolution, and thus, the resulting DSM has fewer details. This dilemma can be resolved by aligning the azimuth signals with a DEM of the area and performing pulse compression in time domain. Images formed or projected onto a DEM are referred to as geocoded [14], typically in a map geometry. SAR image formation in time domain requires a comparatively high computational complexity and offers few benefits for data recorded in stable illumination conditions, like spaceborne SAR. The small angular diversity of a tomographic spaceborne SAR data acquisition does not require back-projection for 3-D reconstruction [15], [16]. These reasons explain the scarce use of map geometry for 3-D image formation purposes. The use of map geometry was promoted in [17], where the authors presented a method to convert interferometric products derived in slant range geometry into map geometry. The work in [18] describes some differences in terms of sensitivity to errors in baseline length and angle for interferometric processing with SAR images in radar or map geometry.

Exceptionally, tomographic reconstruction in [19] and [20] has been performed by applying diverse spectral estimation methods to a set of 3-D geocoded SLC products focused by means of a time domain back-projection algorithm (TDBP). A similar approach has been used in [21] and [22] to analyze glaciers and forest structure. A combination of TDBP and the maximum likelihood spectral estimation method was applied in [23] for change detection purposes. Tomographic reconstruction with circular SAR data [24] is often performed in map geometry with TDBP or fast factorized back-projection [25]. The aforementioned literature exemplifies the link between airborne SAR data and geocoded SLC products formed by means of the TDBP algorithm. Aside from accommodating adverse illumination conditions, the use of TDBP and geocoded SLCs is of special interest as the resulting DSMs are formed in a 3-D grid of the real world. The subsequent postprocessing, such as noncoherent tomographic processing and change detection, simplifies as 1) DSMs formed in the same 3-D grid coregister automatically, 2) it enables operations at a voxel level, and 3) if the 3-D grid is built above the digital terrain model (DTM)

Manuscript received January 4, 2021; revised March 3, 2021 and April 8, 2021; accepted April 9, 2021. Date of publication April 15, 2021; date of current version May 6, 2021. This work was supported in part by armasuisse W+T.
(Corresponding author: Elías Méndez Domínguez.)

The authors are with the Remote Sensing Laboratories, Department of Geography, University of Zurich, Zurich CH-8057, Switzerland (e-mail: elias.mendez@geo.uzh.ch; david.small@geo.uzh.ch; Daniel.henke@geo.uzh.ch).

Digital Object Identifier 10.1109/JSTARS.2021.3073508

of the area, then the SAR-based DSM does not contain voxels below the ground level. In addition, if the use of a DEM is recommended for azimuth compression, then 3-D images formed with a 3-D grid could be more suitable for compression in elevation, as targets are focused at their real 3-D position. As a consequence, tomographic reconstruction methods based on 3-D geocoded imagery might offer some performance improvements in comparison to that obtained with 2-D images.

In this work, we study diverse tomographic reconstruction techniques suitable for deriving DSMs. We first describe adaptions of the existing tomographic reconstruction methods when using 2-D geocoded SLC products. For this purpose, we describe the procedure to derive the normal dimension (perpendicular to the slant-range and azimuth plane) in map geometry with TDBP. This permits extraction of the signals in the normal dimension and to perform pulse compression in elevation. We extend the methods published in [19] and [23], by deriving DSMs when applying techniques capable of operating with 3-D geocoded SLCs. The performance of the methods is then evaluated and compared in terms of their respective ability to reproduce the corresponding airborne laser scanning (ALS) DSM. On the one hand, we provide a global comparison of the DSMs obtained with the different tomographic reconstruction methods, and on the other hand, we provide insight into a particular method when it uses 2-D or 3-D images.

The rest of this article is organized as follows. Section II summarizes the different tomographic reconstruction techniques. Section III presents a four-stage procedure to derive DSMs from geocoded SLC products. We describe each stage, and emphasize differences to methods applied to SAR images in radar geometry. Section IV illustrates the results with high-resolution airborne SAR data acquired in medium density urban scenarios. Section V wraps up with a discussion of the results and conclusions.

II. Review of 3-D Reconstruction Methods

A DSM can be derived from K acquisitions of a TomoSAR dataset by a) applying pulse compression to the signals acquired in the three imaging dimensions, often referred to as range, azimuth, and elevation [12], and b) deriving the 3-D coordinates of the scatterers with the power of the resulting compressed signals. Fig. 1 shows a schematic tomographic acquisition with K flight passes. The azimuth dimension is parallel to the flight path, and the normal or elevation dimension is perpendicular to both range and azimuth. We focus here on pulse compression in elevation. The techniques to process the signals in elevation can be divided into two categories [26], [27]: spectral estimation, and compressive-sensing.

Spectral estimation methods can be nonparametric or parametric. The former apply a set of filters to the K-length received signal. The most popular method is the matched filter, or beamforming (BF) [12]. BF is computationally efficient but delivers images with poor resolution and prominent sidelobes in elevation [28]. The method in [29] exploits the singular value decomposition of the steering vector matrix to achieve better resolution in elevation. Adaptive BF [19], [30], [31], or Capon beamforming (CBF), employ the covariance matrix of

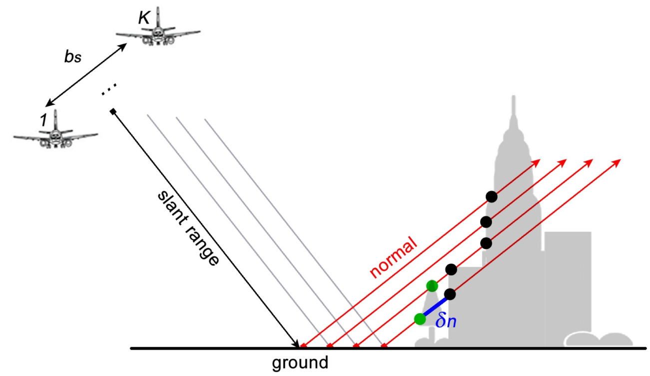


Fig. 1. TomoSAR acquisition geometry. In TomoSAR, each normal line can contain a set of $0 \leq n_s \leq (K-1)$ scatterers of the 3-D world, while for an InSAR configuration $0 \leq n_s \leq 1$. The aperture length in elevation is b_s determines the spatial resolution in that dimension δ_n. Some normal lines contain scatterers of the ground (red squares), of the buildings (black dots), and the tree (green dots).

the received signals to reduce sidelobes in elevation. CBF is sensitive to errors in the steering vectors [19]. This can be solved by applying a diagonal loading factor to the covariance matrices as in robust CBF [32], or the variants described in [33] and [34].

Parametric methods model the received K-length signal as a sum of sinusoids based on their statistical properties. Multiple signal classification (MUSIC) is one of the most utilized super-resolving 3-D reconstruction methods in this domain. It is based on the eigen-decomposition of the covariance matrix [20], [30], [35] and involves a 1-D parameter search, i.e., MUSIC searches for one scatterer at a time. If a ground cell contains multiple correlated scatterers, the covariance matrix becomes singular and the performance of MUSIC degrades.

Multidimensional MUSIC (MD-MUSIC) [36], [37] is the natural extension of MUSIC when performing a multidimensional search. Two additional variants of MUSIC were proposed in [36], and applied to SAR tomography in [38]. These methods, referred to as weighted signal subspace fitting (SSF) and weighted noise subspace fitting, can cope with data showing complex statistical properties and can be applied in the presence of highly correlated targets [38]. Alternative methods such as nonleast squares (NLS) [15], [39], [40] and the maximum likelihood estimator (MLE) [41]–[43] perform a multidimensional search to derive the elevation and amplitude of n_s layovered scatterers within each resolution ground cell. These methods can improve the performance but their computation complexity make them unpractical for high numbers of images [44].

Methods based on compressive sensing aim to reduce the computation complexity of the multidimensional search while providing the performance similar to NLS or MLE. The main idea is to find an orthogonal basis where the signal in elevation becomes sparse, i.e., the signal has nonzero samples at the elevation positions of the targets. The three-staged method in [44], referred to as SL1MMER (scale-down by L_1-norm minimization, model selection, and estimation reconstruction), performs first a sparsification of the signals in elevation to alleviate the need for a multidimensional search. Subsequently, a refinement of the solution is performed by removing spurious peaks that might result from the minimization process; and finally, the complex-valued reflectivity of each scatterer is derived with least squares. The works in [6] and [45] summarize the performance and characteristics of additional tomographic reconstruction methods, such as M-RELAX [30] or those based on the generalized-likelihood ratio test [46]–[48].

The tomographic reconstruction methods can be categorized as single-look or multilook [26]. The former, e.g., BF, provide a valid solution when operating with K single-look images. These methods preserve the range and azimuth resolution of the K images but are more sensitive to phase noise. Multilook methods [49] exploit information from neighboring samples to increase robustness against phase noise at the cost of slightly degraded spatial resolution. Single-look methods can be translated into their corresponding multilook variant.

III. Deriving DSMs With Geocoded SLC Products

A DSM can be derived from geocoded SLC products by applying the following four-stage procedure:

1) compute the 2-D or 3-D K input SAR images;
2) for each ground resolution cell derive the number of $0 \leq n_s \leq (K-1)$ scatterers in layover and its elevation dimension;
3) perform pulse compression with the signals in elevation; and
4) apply detection of the maxima in elevation, i.e., locate the 3-D coordinates of the n_s strongest peaks of the compressed signal in elevation.

Some pulse compression methods can exploit 3-D geocoded SLC products, such as BF, Capon, and MUSIC [19]. For this reason, the pulse compression scheme utilized in step 3) determines whether the user has to focus 2-D or 3-D K images in step 1). The fourth step is not applied if the pulse compression method requires a multidimensional search. Fig. 2 shows the general procedure to derive DSMs from geocoded 2-D or 3-D images. In the following, we describe each of these steps.

A. Focusing 2-D or 3-D Geocoded SLC Products

From the overviews of TDBP and setups in [20] and [50], the 2-D K images of a TomoSAR acquisition can be automatically focused and geocoded by means of TDBP and the use of an external DEM. If a DEM is not available, we can use a plane located at the lowest ground height of the area. Omitting radiometric calibration factors and signal weightings, the backscatter of an element acquired from channel k is computed at height h_i via

$$\gamma^{k,i}(e,n,h_i) = \left[\sum_{\tau=\tau_1}^{\tau_2} s_M\left(\frac{2R_{\text{sr}}}{c},\tau\right)\right] \cdot e^{-j\cdot(4\pi/\lambda)\cdot R_s} \quad (1)$$

where (e,n,h_i) are the map coordinates of the voxel, $\tau_2 - \tau_1$ is the aperture length in azimuth, s_M is the bandpass upsampled range compressed signals (i.e., after matched filtering, upsampling and signal bandpass conversion), R_{sr} the slant range at slow time τ, and c the speed of light. λ is the wavelength and R_s is the slant range at closest approach.

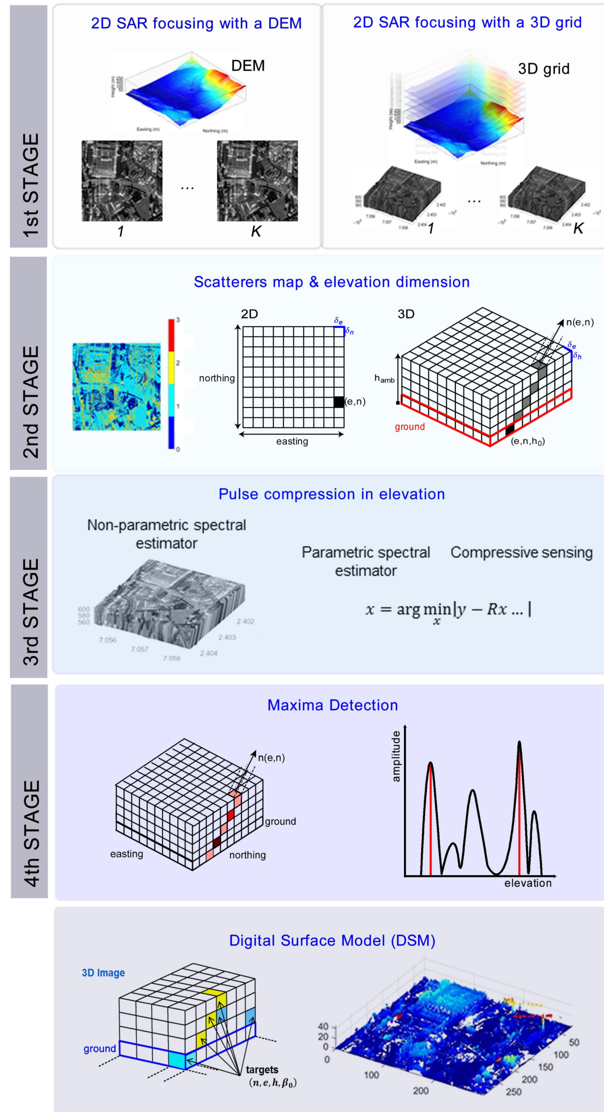


Fig. 2. Processing chain for deriving DSMs with geocoded SAR SLC products. First stage: focusing of the 2-D or 3-D geocoded SLC products with TDBP using an external DEM. If the pulse compression in elevation works with 3-D images, then the 3-D geocoded SLC products can be focused using TDBP and a 3-D grid. The 3-D grid is often based on an external DEM. Second stage: computation of the scatterers map and the elevation dimension. This permits pulse compression in elevation and detection of maxima. Third stage: perform pulse compression in elevation. An SAR-based DSM is obtained after applying detection of the maxima with the pulse compressed signals in elevation. The SAR-based DSM is a 3-D image, where a voxel is left empty when the corresponding sample of the elevation signal is not a maximum. Detection of maxima is not applied when pulse compression in elevation is performed with parametric spectral estimators or techniques based on compressive sensing theory. The 3-D geocoded SAR image provides information on the radar brightness and 3-D location of the targets, while the SAR-based DSM provides information on the 3-D location only.

The user has to focus 3-D geocoded SLCs instead of 2-D SLCs if the pulse compression method in stage 3 (see Fig. 2) works with those products. A 3-D geocoded SLC γ^k can be obtained by using (1) with $h_i = h_0 + i\delta_h$, $i = 0, \ldots, i_{\text{amb}}$ where h_0 is the height of the ground given by the DEM, δ_h is the height sampling spacing and $h_{\text{amb}} = h_0 + i_{\text{amb}}\delta_h$ is the ambiguity height [11]. $\gamma^{k,0}$ is the 2-D geocoded image from channel k focused on the ground. From now on, we assume the use of a horizontally regularly spaced DEM with a sample interval of $\delta_e \times \delta_n$, and that the geocoded SLCs are processed in single look. The baseband conversion term in (1) is required for pulse compression in elevation. In this term, the distance at closest approach has to be computed during TDBP focusing. In contrast to SAR products in radar geometry, the value of the slant range of a geocoded image sample is not inherent to the raster grid being used.

B. Map of Scatterers and Elevation

The number n_s of scatterers in layover of ground resolution cell with coordinates (e, n) can be computed with diverse methods [51]. Here, we use the efficient detection criterion, where $0 \leq n_s(e,n) \leq (K-1)$ is given by

$$\begin{aligned} n_s(e,n) = \ &\arg \min_{m\in[0,K-1]} [(N_{\text{looks}} - k)\cdot m \\ &\cdot \ln\left(\frac{\sqrt[K-m]{\prod_{i=m+1}^{K}\lambda_i(e,n)}}{\frac{1}{(K-m)\sum_{i=m+1}^{K}\lambda_i(e,n)}}\right) \\ &+ m\cdot(2K-m)\cdot C_{N_{\text{looks}}}] \end{aligned} \tag{2}$$

where N_{looks} is the number of looks used to compute the $K \times K$ sample covariance matrix of the resolution cells, and $\lambda_{i=1,\ldots,K}(e,n)$ are the corresponding eigenvalues in descending order. In (2), a more accurate map of the number of scatterers was obtained when $C_{N_{\text{looks}}} = \sqrt{N_{\text{looks}} \cdot \log(N_{\text{looks}})}$. Other techniques are the Bayesian information criterion or the minimum description length [52]. The map of scatterers computed with (2) is required for detection of maxima and the parametric pulse compression methods.

For geocoded SLC products, the elevation or normal line of a ground cell cannot be derived analytically but can be calculated in postprocessing, e.g., with the method described in [23]. Assuming a stripmap mode with a linear trajectory, the vector with the coordinates of a normal line (vector normal to the slant range-azimuth plane) $\boldsymbol{n}$ of γ_0^k with origin at the ground cell (e, n) can be approximated by

$$\boldsymbol{n}(e,n) = (e + \Delta e_i, n + \Delta n_i, h_i)_{i=0,\ldots,i_{\text{amb}}} \tag{3}$$

where Δe_i and Δn_i are the offsets in samples of the 2-D displacement field, computed by cross correlation between images $\gamma^{k,i}$ and $\gamma^{k,0}$ geocoded at heights $h_i = h_0 + i\Delta h$ and h_0, respectively, with $\Delta e_0 = \Delta n_0 = 0$. The expression in (3) is required to 1) perform pulse compression in elevation, and 2) perform maxima detection. We simplify notation by assuming that the offsets Δe_i and Δn_i are identical for all samples in $\gamma^{k,i}$. This assumption is a valid approximation for data acquired with linear trajectories and stripmap mode, since in that case we can assume that in the image domain (range, azimuth, and elevation) are perpendicular to each other, and a common elevation dimension exists [16]. The tracks are assumed to be parallel to each other in case of using datasets recorded in multiple passes. For data acquired with nonlinear flight paths or in a spotlight mode, the values of Δe_i and Δn_i are sample dependent and the three imaging dimensions (radial distance, azimuthal angle, and elevation) are not perpendicular to each other. Fig. 2 shows in the second stage a voxelized 3-D image highlighting the normal line of the ground cell (e, n) of its corresponding 2-D geocoded image.

To derive $\boldsymbol{n}(e, n)$, we need to focus one 3-D geocoded SLC $\boldsymbol{\gamma}^k$ product. This can be performed using a patch of $\boldsymbol{\gamma}^k$ as input to reduce computation time. However, one must recognise that the size of the patch needs to be large enough to enable the computation of $\Delta e_{i_{\text{amb}}}$ and $\Delta n_{i_{\text{amb}}}$. To obtain a more reliable displacement field, it is desired that the patch contains objects with strong edges or point targets, discarding uniform areas such as grasslands or water bodies using, for example, the local coefficient of variation [53].

C. Pulse Compression in Elevation With 2-D geocoded SLC products

Tomographic reconstruction methods based on 2-D images use those focused at a height $h_i = h_0$, i.e., use the set $\{\gamma^{1,0}, \gamma^{2,0}, \ldots, \gamma^{K,0}\}$. The backscatter at the ground resolution cell (e, n, h_0) of the kth acquisition is computed with (1) and denoted as $\gamma^{k,0}(e, n, h_0)$. From [40], the received signal $\boldsymbol{y}$ at ground cell (e, n) can be expressed as the sum of the elevation profile weighted by a linear phase term as follows:

$$\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{\epsilon} \tag{4}$$

where $\boldsymbol{y} = (\gamma^{1,0}(e,n,h_0), \ldots, \gamma^{K,0}(e,n,h_0))^T$ is of size $K \times 1$, $\boldsymbol{R}$ is the $K \times L_e$ matrix with the steering vectors in its columns, i.e., $R(k,l) = e^{-j2k_c\Delta Rs(k,l)}$, being k_c is the wavenumber, l is the discrete elevation position, and L_e is the length of the discretized elevation dimension; $\boldsymbol{x}$ is the unknown reflectivity vector in elevation of size $L_e \times 1$, and

$$\Delta Rs(k,l) = \begin{cases} 1 & , k = 1 \\ \Delta Rs(k,l) = R_s(k,l) - R_s(1,l) & , k > 1 \end{cases} \tag{5}$$

with R_s being the slant range. In map geometry, the slant range of a ground cell is not derived in postprocessing but simultaneously during pulse compression in azimuth. This increases memory requirements, as we need to store the distance at closest approach corresponding to each voxel (e, n, h_i) and image k. The elevation position l refers to the voxel at location $(e + \Delta e_l, n + \Delta n_l, h_l)$. The matrix $\boldsymbol{R}$ is evaluated in the normal dimension of the ground cell (e, n) with (3). For CBF and MUSIC, the power of the compressed signals $\boldsymbol{x}$ is given by

$$|\boldsymbol{x}_{\text{CBF}}(l)|^2 = \frac{1}{\boldsymbol{r}^H(l)\hat{\boldsymbol{C}}^{-1}\boldsymbol{r}(l)} \tag{6}$$

$$|\boldsymbol{x}_{\text{MUSIC}}(l)|^2 = \frac{1}{\boldsymbol{r}^H(l)\hat{\boldsymbol{E}}_n\hat{\boldsymbol{E}}_n^H\boldsymbol{r}(l)} \tag{7}$$

where $(.)^H$ indicates the Hermitian operator, $\boldsymbol{r}(l)$ is the lth column (steering vector at elevation position l) of the matrix $\boldsymbol{R}$, $\hat{\boldsymbol{C}} = \frac{1}{L_k}\sum_{n=1}^{L_k}(\boldsymbol{y}_n\boldsymbol{y}_n^H)$ is the covariance matrix for the signals at ground cell (e, n), $\boldsymbol{y}_n$ is the measurement vector of the nth look, and L_k is the number of looks. The noise space

$\hat{\boldsymbol{E}}_n$ is determined by the $(K-n_s)$ eigenvectors corresponding to the $(K-n_s)$ smallest eigenvalues. The calculation of the covariance matrices $\hat{\boldsymbol{C}}$ varies depending on the distribution of the phase centers (uniform or not) [51]. The covariance matrices can be computed by an adaptive ensemble averaging inside a sliding window [54] or with nonlocal means [55], [56].

If the pulse compression method performs a multidimensional search, then the DSM is computed by maximizing an objective function. Let us rewrite (4) by assuming the presence of n_s scatterers with elevations $\boldsymbol{l} = [l_1, \ldots, l_{n_s}]$ as

$$\boldsymbol{y} = \boldsymbol{R}(\boldsymbol{l})\boldsymbol{x}(\boldsymbol{l}) + \boldsymbol{\epsilon} \tag{8}$$

where $\boldsymbol{R}(\boldsymbol{l})$ is of size $K \times n_s$ depending on the unknown elevations of the scatterers derived by means of a K-dimensional search of $\boldsymbol{l}$, and l_i is the voxel at location $(e+\Delta e_i, n+\Delta n_i, h_i)$. The objective function of the methods studied in this work is

$$\boldsymbol{l}_{\text{MD-MUSIC}} = \arg\max_{\boldsymbol{R}(\boldsymbol{l})} \mathrm{Tr}(\boldsymbol{P_A}\hat{\boldsymbol{E}}_s(\boldsymbol{l})\hat{\boldsymbol{E}}_s^H(\boldsymbol{l})) \tag{9}$$

$$\boldsymbol{l}_{\text{SSF}} = \arg\min_{\boldsymbol{R}(\boldsymbol{l})} \mathrm{Tr}[(\boldsymbol{I} - \boldsymbol{P_A}(\boldsymbol{l}))\hat{\boldsymbol{E}}_s(\boldsymbol{l})\boldsymbol{W}_{\text{SSF}}(\boldsymbol{l})\hat{\boldsymbol{E}}_s^H(\boldsymbol{l})] \tag{10}$$

$$\boldsymbol{l}_{\text{NLS}} = \arg\max_{\boldsymbol{R}(\boldsymbol{l})} [\boldsymbol{y}^H\boldsymbol{R}(l)(\boldsymbol{R}(l)^H\boldsymbol{R}(l))^{-1}\boldsymbol{R}^H(l)\boldsymbol{y}] \tag{11}$$

$$\boldsymbol{l}_{\text{MLE}} = \arg\max_{\boldsymbol{z}(\boldsymbol{l})} [e^{-\boldsymbol{z}^H(\boldsymbol{l})\hat{\boldsymbol{\Gamma}}^{-1}\boldsymbol{z}(\boldsymbol{l})}] \tag{12}$$

where $\boldsymbol{P_A}(\boldsymbol{l}) = \boldsymbol{R}(\boldsymbol{l})[\boldsymbol{R}(\boldsymbol{l})^H\boldsymbol{R}(\boldsymbol{l})]^{-1}\boldsymbol{R}(\boldsymbol{l})^H$. The matrix $\boldsymbol{R}(\boldsymbol{l})$ contains the steering vectors of the corresponding n_s scatterers, and $\hat{\boldsymbol{E}}_s$ is the signal space determined by the n_s eigenvectors corresponding to the n_s largest eigenvalues and

$$\boldsymbol{W}_{\text{SSF}}(\boldsymbol{l}) = [\boldsymbol{D_s}(\boldsymbol{l}) - \hat{\sigma}_n^2\boldsymbol{I}]^2\boldsymbol{D_s}^{-1}(\boldsymbol{l}). \tag{13}$$

In (13), $\boldsymbol{D_s}(\boldsymbol{l}) = \mathrm{diag}([\lambda_1, \ldots, \lambda_{n_s}])$, $\boldsymbol{I}$ is the identity matrix of size $n_s \times n_s$, and $\hat{\sigma}_n^2$ is the noise standard deviation. In (12), $\boldsymbol{z}(\boldsymbol{l}) = \sum_{n=1}^{n_s}\boldsymbol{R}(l_n)$, $\boldsymbol{R}(l_n)$ is the steering vector of the nth target at elevation position l_n, and $\hat{\boldsymbol{\Gamma}}$ is the coherence matrix.

The DSM is derived in a similar fashion using methods based on compressive sensing theory. In the absence of noise, and based on the model in [44], the reflectivity at elevation $\boldsymbol{x}$ can be estimated with

$$\boldsymbol{x} = \arg\min_{\boldsymbol{x}}\{||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2^2 + \lambda_K||\boldsymbol{x}||_1\} \tag{14}$$

or

$$\boldsymbol{x} = \min_{\boldsymbol{x}} ||\boldsymbol{x}||_1 \;\; \text{s.t.} \;\; ||\boldsymbol{y} - \boldsymbol{R}\boldsymbol{x}||_2 < \sigma_\epsilon \tag{15}$$

being λ_K, a Lagrange multiplier that depends on the number of images K and the noise level σ_ϵ. The expressions in (14) and (15) are solved here with the least angle regression algorithm (LARS) and the stage-wise orthogonal matching pursuit algorithm (StOMP) [57], respectively. After LARS (or StOMP), only a few elevation positions remain plausible for each location of the n_s scatterers. We finally minimize (14) or (15) based on the value of n_s.

D. Pulse Compression in Elevation With 3-D Geocoded SLC Products

Tomographic reconstruction methods based on 3-D geocoded SLCs use $\{\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2, \ldots, \boldsymbol{\gamma}^K\}$. The number of applicable reconstruction methods is reduced by the required computational complexity or because the method does not admit 3-D inputs. For CBF, the power at location (e, n, h_i) is given by

$$|\boldsymbol{x}_{\text{CBF}}(e, n, h_i)|^2 = \frac{1}{\boldsymbol{r}_i^H\hat{\boldsymbol{C}}_i^{-1}\boldsymbol{r}_i^H} \tag{16}$$

where $\boldsymbol{r}_i$ is the ith steering vector at position (e, n, h_i) of the matrix $\boldsymbol{R}$ and

$$\hat{\boldsymbol{C}}_i = \frac{1}{L_k}\sum_{n=1}^{L_k}(\boldsymbol{y}_n(e, n, h_i)\boldsymbol{y}_n^H(e, n, h_i)). \tag{17}$$

The covariance matrix is computed from the recorded signals at height h_i. This operation increases significantly the computation time. However, the signal to noise ratio of a scatterer improves when focused at its actual 3-D position [19]. The main rationale behind this approach is that for a target located at (e, n, h_i), the signal in $\boldsymbol{y}_n(e, n, h_i)$ is expected to be more accurate than $\boldsymbol{y}_n(e, n, h_0)$, and by extension, the corresponding covariance matrix. This modification can be applied to MUSIC in (7).

Multidimensional search-based parametric spectral estimators become impractical due to prohibitive computational complexity. For this reason, we evaluate here the performance of a single-dimensional search-based MLE, operating in a fashion similar to MUSIC. First, we compute the power given by (12) with

$$\boldsymbol{x}_{\text{MLE}}(e, n, h_i) = e^{-\boldsymbol{z}^H(h_i)\hat{\boldsymbol{\Gamma}}_i^{-1}\boldsymbol{z}(h_i)} \tag{18}$$

where $\boldsymbol{z}(h_i) = \boldsymbol{R}(h_i)$ is the steering vector of the voxel (e, n, h_i), and $\hat{\boldsymbol{\Gamma}}_i$ is the sample coherence matrix at height h_i. The argument of the exponential function in (18) resembles the CBF in (6) with the difference being that MLE utilizes the coherence matrix instead of the covariance matrix [41].

E. Maxima Detection

Maxima detection is applied to the output power resulting from (6), (7), or (18). For each ground cell (e, n), we retain the 3-D coordinates (e, n, h) of the n_s largest peaks in $\boldsymbol{x}(\boldsymbol{n}(e, n))$. This operation consists of an examination of the normal lines of each ground cell with the expression in (3). After applying maxima detection or maximizing the resulting DSM, referred to as I_{DSM}, we have a 3-D image with a voxel size given by $(\delta_e \times L_e) \cdot (\delta_n \times L_n) \cdot \delta_h$, with $L_k = L_e \times L_n$ being the multilook factor in easting and northing, and $\delta_e \times \delta_n$ is the ground sample distance (GSD). The occupancy matrix O is a 3-D binary matrix whose elements take the value 0 or 1 indicating absence or presence of a scatterer, respectively, in each voxel v of I_{DSM}. $O(v) = 1$ if and only if the backscatter of the voxel $v \in \boldsymbol{n}(e, n)$ is one of the $n_s(e, n)$ strongest. Note that $\sum_{i=0}^{i_{\text{amb}}} O((e+\Delta e_i, n+\Delta n_i), h_i) = n_s(e, n)$.

F. Experimental Settings

The methods in this work were evaluated using multilooking. In map geometry, an SAR image focused on a DEM with a sample interval of $(L_e\delta_e \times L_n\delta_n)$ is not equivalent to the $(L_e \times L_n)$-multilooked version of the SAR image focused on a DEM with a GSD of $(\delta_e \times \delta_n)$. In (1), the height sampling spacing δ_h tradesoff the precision of the geolocation of the scatterers and the computation time of the entire 3-D image focusing chain. To avoid aliasing, we ensure $\delta_h \leq \delta_n \cdot \cos(\theta_{\text{inc}})$, where δ_n is the Rayleigh resolution in elevation [12] given by $\delta_n = \frac{\lambda R_s}{2b_s}$, θ_{inc} is the incident angle, and b_s is the aperture length in elevation. In practice, the achieved resolution in elevation is better than the Rayleigh resolution δ_n, and thus, a finer δ_h should be used. In this work, we use $\delta_h = 25$ cm for datasets with a Rayleigh resolution of approximately 30 m in the best cases. The sample intervals in northing and easting were set to 10 cm for the images acquired in the first test site and to 25 cm for those of the second test site. These values were chosen based on the ground resolution cell size, the size of DEM used for image focusing, and the computation time required by TDBP.

The scatterers map was obtained after denoising with block matching 3-D [58], the eigenvalues of the covariance matrices. For illustration purposes only, the SAR-based DSMs shown here were filtered using the entropy [59]. In this case, we removed all points where entropy was lower than 0.3. We did not apply any additional filtering or denoising to the signals in the processing chain nor to the resulting DSMs.

G. Performance Analysis

The performance of the methods was evaluated based on an ALS-based DSM of the area of interest. Based on this ALS-based DSM, we built a 3-D voxelized image with height information with a voxel size equal to that of the DEM used for SAR focusing. This enables the derivation of quality indicators in a fashion similar to image classification or change detection, i.e., a voxel to voxel comparison. We used the κ coefficient [60] as a tradeoff between false alarms and correct detections to evaluate the capability of each method to reproduce the ALS-based DSM. False alarms were assumed to be caused by outliers, related to sidelobes or phase noise. A false alarm occurs when $I^{\text{ALS}}_{\text{DSM}}(v) \neq I^{\text{SAR}}_{\text{DSM}}(v)$, where $I^{\text{ALS}}_{\text{DSM}}$ and $I^{\text{SAR}}_{\text{DSM}}$ are the ALS-based DSM and the SAR-base DSM under evaluation. A correct detection occurs in the case of equality. To better emphasize the differences between the performances of the different methods, we list the ratios between the κ achieved by the best approach in comparison to the others. This indicator is referred to here as κ_r, and has the value of unity for the SAR-based DSM resembling the ALS-based DSM most.

As global quality indicators, we computed the mean and standard deviation of the absolute value of the height error as $\mu_{\Delta I}$ and $\sigma_{\Delta I}$. These indicators, expressed in meters, were obtained from the difference image ΔI given by

$$\Delta I(e,n) = \sum_{h=h_0}^{h_{\text{amb}}} \left(\mid I^{\text{SAR}}_{\text{DSM}}(e,n,h) - I^{\text{ALS}}_{\text{DSM}}(e,n,h) \mid \right). \quad (19)$$

If the SAR-based DSM is identical to the ALS-based DSM, then $\mu_{\Delta I}$ and $\sigma_{\Delta I}$ are zero. The standard deviation provides a measure of the number of outliers in the SAR-based DSM under evaluation. The quality indicators $\mu_{\Delta I}$ and $\sigma_{\Delta I}$ are computed with the entire area of interest and can be used to compare the performance of the methods; however, the values of those indicators do not provide a representative estimate of the height accuracy achieved by a particular method because 1) many objects, such as rooftops, trees, shadows, water bodies, and asphalt-covered roads, might not have a significant backscatter at a particular frequency band, and thus, the SAR-based DSM does not contain information about them, 2) the SAR data and the LiDAR-based DSM were not acquired simultaneously, and thus, the presence of cars or other small objects differ, 3) the SAR data are acquired from a downward side-looking antenna in a single flight, while the ALS-based DSM is acquired from a downward off-nadir-looking laser, and 4) the SAR-based DSM contains errors caused by presence of moving objects and artifacts. To provide a more significant estimate of the height accuracy achieved by a given method, we derived the mean $\mu_{H_{\text{ALS}}}$ and standard deviation $\sigma_{H_{\text{ALS}}}$ of the height of some objects in the ALS-based DSM. We performed the same process with the objects in the SAR-based DSMs to obtain $\mu_{H_{\text{SAR}}}$ and $\sigma_{H_{\text{SAR}}}$. If the SAR- and ALS-based DSMs are identical, then $\mu_{H_{\text{ALS}}} = \mu_{H_{\text{SAR}}}$ and $\sigma_{H_{\text{ALS}}} = \sigma_{H_{\text{SAR}}}$. The differences in the values of μ_H provide an estimate of the respective height bias, while the differences between the values of σ_H provide a measure of presence of outliers. To ease interpretation of the results, we report the difference in the mean values of the ALS- and SAR-based DSMs given by $\Delta\mu_H = |\mu_{H_{\text{ALS}}} - \mu_{H_{\text{SAR}}}|$.

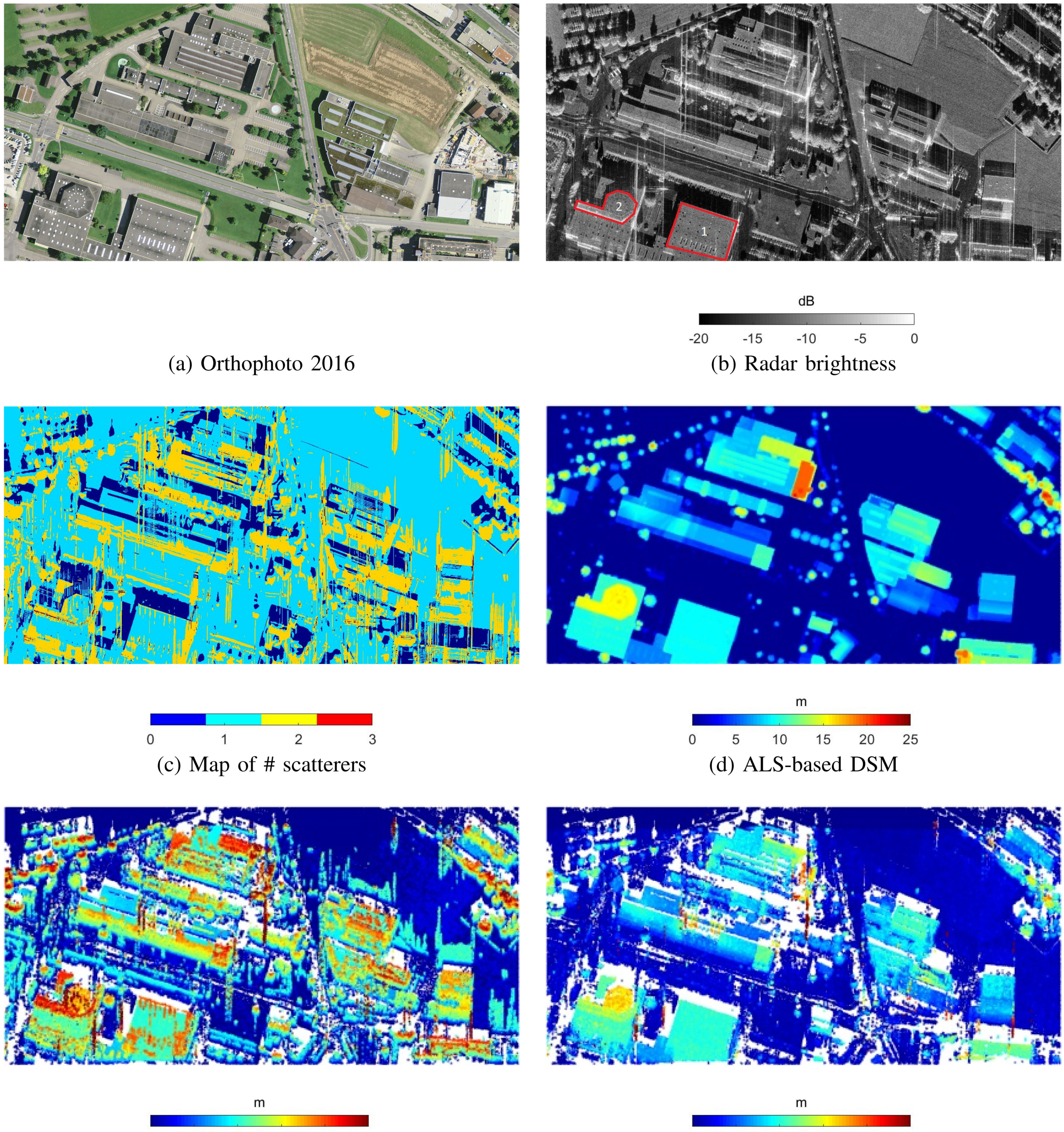
Finally, we also provide the ratio between the computation times of the fastest method with respect to the other. This ratio, referred to as t_r, provides the user an indicator of the computational complexity required by each approach. For the fastest method, $t_r = 1$. The method's computational footprint is determined by the time required to perform pulse compression in range, azimuth, and elevation.

In this work, we do not analyze the super-resolution capabilities provided by each method, as ALS-based DSM products do not contain information on the walls or facades of the buildings. As a consequence, SAR-based DSMs from methods providing super-resolution could yield a lower κ coefficient in comparison to those without super-resolution capabilities. The usage of a combined TLS- and ALS-based DSM could be a better solution in order to account for this property. The super-resolution capability of a method is independent of whether or not it is applied to images in radar or map geometry. We refer the reader to e.g., [61], [62] for more details.

IV. Experimental Results

A. Data and Test Site

Here, we introduce SAR datasets that were acquired with Fraunhofer FHR's *Ka*-band MEMPHIS sensor [63] over two test sites. The first test site was Hinwil (Switzerland); an orthophoto is shown in Fig. 3(a); the second test site was Memmingen (Germany)—an orthophoto is shown in Fig. 5(a). The sensor



(a) Orthophoto 2016
(b) Radar brightness
(c) Map of # scatterers
(d) ALS-based DSM
(e) SAR-based DSM (StOMP)
(f) SAR-based DSM (MLE, 3D images)

Fig. 3. Orthophoto, SAR image, scatterers map, and DSMs of the test site in Hinwil (Switzerland). In the SAR image, red rectangles indicate two objects where the height accuracy was computed locally.

was equipped with four receiving antennas, enabling single-pass multibaseline cross-track interferometry (3 baselines). Table I lists the main system parameters. The 2-D SAR image focusing was performed with a graphic processor unit-based TDBP processor [64]. The geolocation error of the images provided by MEMPHIS was found to be a few centimeters in an analysis of the signatures of corner reflectors deployed in the area of interest [65].

B. Graphical and Numerical Results

The four-stage procedure described in Section III was applied to the MEMPHIS data recorded over the test sites.

The first test site is located in the industrial quarter of Hinwil (Switzerland). Fig. 3(a) and (b) shows an orthophoto of the test site and the radar brightness of one channel after applying multilooking with a 5×5 sliding window. The red polygons

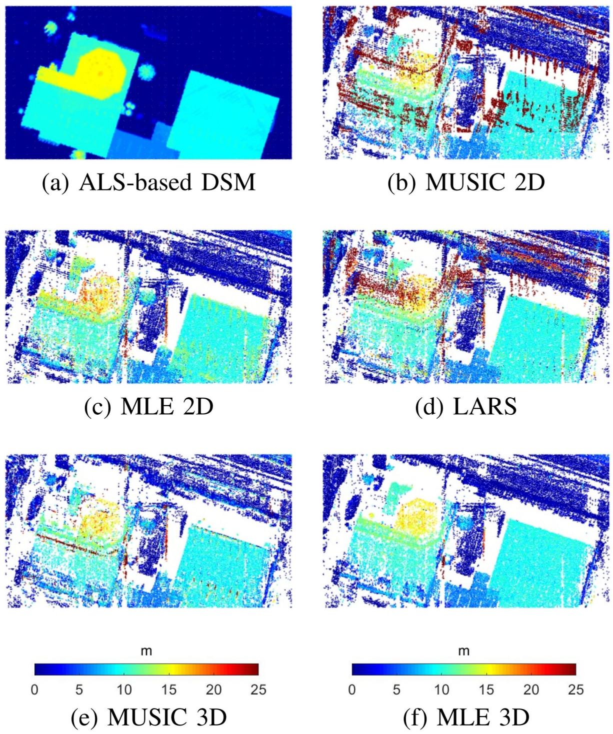


Fig. 4. Top view of six DSMs of the building used for numerical evaluation in Hinwil. The colormap encodes the height above the ground in meters.

TABLE I
MEMPHIS System Parameters

Carrier frequency	35 GHz (Ka-band)
Range Bandwidth	900 MHz
PRF	1500 Hz
Average airplane velocity	77 m/s
Airplane altitude a.g.l.	1300-1400 m
Antenna tilt angle	20°-35°
Slant range resolution	0.167 m
Max. azimuth sample interval	0.082 m
Min. Rayleigh elevation resolution	30 m
Nominal baseline lengths	0.0055 m, 0.165 m, 0.275 m
Slant range at image center	1547 m
Nominal depression angle	30°

indicate areas where the average ($\mu_{H_{\mathrm{SAR}}}$) and standard deviation ($\sigma_{H_{\mathrm{SAR}}}$) of the height was computed. Fig. 3(c) and (d) illustrates maps of the local number of scatterers and the top view of the ALS-based DSM used for the performance analysis. The map of scatterers shows an absence of targets (dark blue) in the shadows cast by buildings or trees. The ground cells usually contain a single scatterer (light blue). Fig. 3(e) and (f) shows two SAR-based DSMs obtained with StOMP and MLE after filtering points with an entropy higher than 0.3. The SAR-based DSMs did not have scatterers on some asphalt-covered roads due to their inherent high entropy value. Visual inspection of the images in Fig. 3(e) and (f) shows that the DSMs derived using MLE based on 3-D geocoded SLCs resembles more the ALS reference DSM than does the StOMP-based DSM.

Table II lists quality indicators obtained after applying the different pulse compression methods in elevation. Methods based on 3-D geocoded SLCs as input provided better global quality indicators (κ_r, $\mu_{\Delta I}$, and $\sigma_{\Delta I}$), than those based on 2-D images. The local indicators $\Delta\mu_H$ and $\sigma_{H_{\mathrm{SAR}}}$ were computed with two objects, each containing a section of the same building. The first object, labeled "1" in Fig. 3(a), is a flat rooftop. The second object, labeled "2," is a flat octagonal-like shaped rooftop module. The mean and standard deviation of the height above ground of the first rooftop are 9 m and 0.17 m based on the DSM derived with ALS data. The octogonal rooftop component has an average height of 15.5 m and a standard deviation of 0.24 m. Fig. 3 shows that the radar brightness of the two rooftops of the building is significant at *Ka*-band, and thus, a robust SAR-based reconstruction of the DSM is possible. For the two objects, based on $\Delta\mu_H$ in Table II, MLE, LARS, and StOMP provided the best results, while the methods based on 3-D images performed in last position. However, MLE yielded the best values for $\sigma_{H_{\mathrm{SAR}}}$ when operating with 3-D images. The parametric spectral estimators and the compressive sensing approaches are a tradeoff solution based on the values of $\Delta\mu_H$ and $\sigma_{H_{\mathrm{SAR}}}$. The multidimensional search-based variants of MUSIC, such as MD-MUSIC and SSF, outperformed the single-dimensional search-based MUSIC based on all quality indicators listed in Table II. The nonparametric spectral estimator MUSIC was found to be the fastest approach when operating with 2-D images, followed by StOMP. The spectral parametric approaches SSF and MD-MUSIC demanded the most computation time due to the presence of many ground cells with more than two scatterers in layover.

Some graphical results from Hinwil are shown in Fig. 4. The ALS-based DSM of the two objects used for evaluating locally the height errors is depicted in Fig. 4(a). Fig. 4(b)–(f) illustrates some SAR-based DSMs—one will observe that the DSMs derived with 2-D images have more outliers (dark red points) than those with 3-D images. This reflects the values of $\sigma_{H_{\mathrm{SAR}}}$ in Table II. Comparison of the DSMs generated with (e) the compressive sensing approach LARS and (f) the parametric-spectral estimation method MLE shows that the former generated more errors above the rooftops of the buildings.

The second test site is located at the Allgäu airport in Memmingen (Germany). Fig. 5(a) and (b) shows an orthophoto of the test site and the amplitude of the radar brightness of one channel after applying multilooking with a 5×5 sliding window. Red rectangles indicate the objects for computing the local quality indicators $\mu_{H_{\mathrm{SAR}}}$ and $\sigma_{H_{\mathrm{SAR}}}$. Fig. 5(c) and (d) illustrates a map of the number of scatterers and the top view of the ALS-based reference DSM used for the performance analysis. The map shows an absence of targets (dark blue) on some rooftops, the apron of the airport, a portion of the airport runway, and the shadows cast by buildings or trees. The ground cells usually contained a single scatterer (light blue). Double scatterers (yellow) are mainly found in trees. Fig. 5(e) and (f) shows two SAR-based DSMs obtained with StOMP, and MLE after filtering points with an entropy greater than 0.3. In contrast to the Hinwil test site, the rooftops of some buildings in Memmingen did not have a significant backscatter and thus, the SAR-based DSMs did not

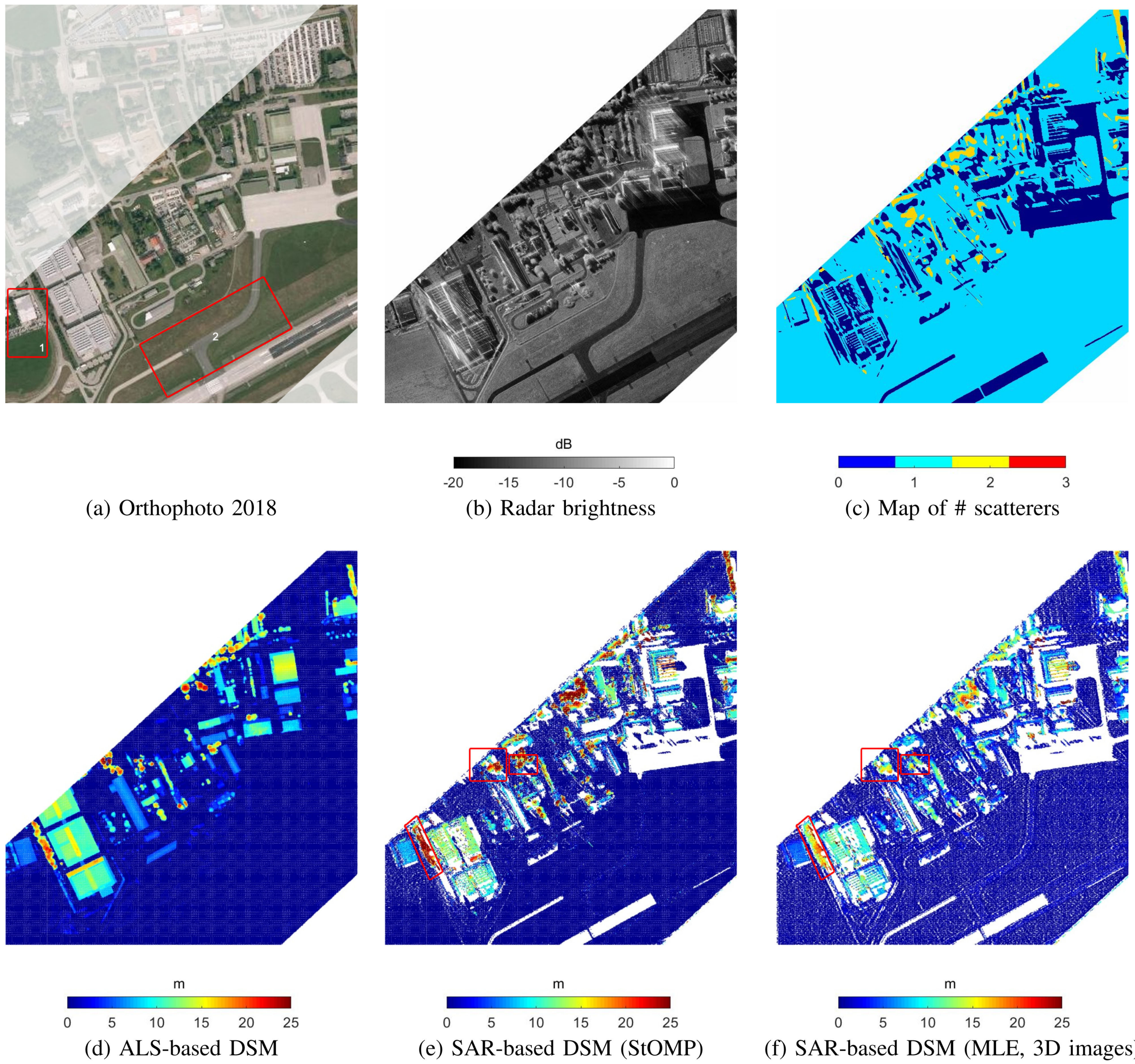


Fig. 5. Orthophoto, SAR image, scatterers map, ALS- and two SAR-based DSMs of the test site in Memmingen (Germany). In the orthophoto the red rectangles indicate the two objects used for computing locally the accuracy in height.

contain them when compared to the ALS-based DSM. The areas indicated by red polygons contained trees; here, one observed the most significant differences between the multiple SAR-based DSMs.

Table III shows the performance analysis after applying the methods described in Section III. Based on the global indicators κ_r, $\mu_{\Delta I}$, and $\sigma_{\Delta I}$ the methods based on 3-D geocoded SLCs performed best. The local indicators were computed with two objects, labeled as 1 and 2 in Fig. 5. The first object is a flat rooftop. The mean and standard deviation of the height above ground of the rooftop were 5.5 m and 0.24 m, respectively. For this object, based on $\Delta\mu_H$ in Table III, the compressive sensing approach StOMP, and the nonparametric spectral estimators using 2-D geocoded SLCs had the smallest errors. The methods based on 3-D images performed worst, but did yield the smallest values of $\sigma_{H_{\text{SAR}}}$. Similar to the Hinwil case, MUSIC operating with 2-D images and StOMP were the fastest methods; however, the spectral parametric approaches required less computation time, as there were fewer ground cells with two or more scatterers in layover. A comparison of Δt obtained with the parametric spectral estimators in Table II with Table III shows the influence of the multidimensional search on the computational complexity. No such influence was observed in the values of t_r provided by the compressive sensing-based approaches.

Fig. 6(a) illustrates the ALS-based DSM of the object used for evaluating locally the height estimation accuracy. The building is located on the left in the DSM. Visual inspection of Fig. 6 shows that the DSM derived with 2-D images and MUSIC had a large number of outliers, reflected in the value of $\sigma_{H_{\text{SAR}}}$ in Table III. One can also observe that for a particular pulse compression method, the derived DSM had fewer errors when using 3-D images. Comparison of the DSMs generated by the compressive sensing approach LARS and the parametric-spectral estimation method MLE showed that the former had more errors above the rooftops of the buildings but fewer than the MUSIC-based DSM when operating with 2-D images.

TABLE II
TEST SITE: HINWIL

		Global				Local, object 1		Local, object 2	
Type	Method	κ_r	$\mu_{\Delta I}$ [m]	$\sigma_{\Delta I}$ [m]	t_r	$\Delta\mu_H$ [m]	$\sigma_{H_{SAR}}$ [m]	$\Delta\mu_H$ [m]	$\sigma_{H_{SAR}}$ [m]
SnP-2D	CBF	2.00	7.27	14.05	1.20	5.01	14.03	1.91	11.43
SnP-2D	MUSIC	1.86	8.17	14.97	1	5.61	14.62	1.87	11.37
SP-2D	MDMUSIC	1.44	2.86	5.60	333.26	1.20	4.87	0.57	7.87
SP-2D	SSF	1.63	2.81	4.88	334.41	1.67	4.12	0.81	7.30
SP-2D	NLS	1.63	3.86	7.86	115.48	1.59	7.42	0.93	8.70
SP-2D	MLE	1.53	2.28	4.32	29.39	0.07	2.41	1.66	4.12
CS-2D	LARS	1.53	3.05	5.81	1.96	0.23	2.33	1.96	9.56
CS-2D	StOMP	1.73	2.39	3.93	1.02	1.08	3.27	0.53	4.87
SnP-3D	CBF	1.37	2.19	3.56	258.80	0.97	1.74	2.67	3.84
SnP-3D	MUSIC	1.63	2.52	3.96	254.54	1.45	2.82	3.7	5.12
SnP-3D	MLE	1	2.13	3.70	254.08	0.36	1.46	2.59	3.67

$\sigma_{H_{\text{ALS}}}$ was 0.17 m and 0.24 m for the first and second object, respectively. SnP-2D refers to spectral nonparametric methods using 2-D images, SP stands for spectral parametric, and CS stands for compressive sensing.

TABLE III
TEST SITE: MEMMINGEN, GERMANY

		Global				Local, object 1		Local, object 2	
Type	Method	κ_r	$\mu_{\Delta I}$ [m]	$\sigma_{\Delta I}$ [m]	t_r	$\Delta\mu_H$ [m]	$\sigma_{H_{SAR}}$ [m]	$\Delta\mu_H$ [m]	$\sigma_{H_{SAR}}$ [m]
SnP-2D	CBF	2.33	1.12	3.39	2.48	0.73	4.11	0.11	0.54
SnP-2D	MUSIC	2.74	1.12	3.39	1	0.83	4.07	0.09	0.44
SP-2D	MDMUSIC	2.74	1.11	3.31	10.07	1.31	2.69	0.10	0.45
SP-2D	SSF	2.74	1.12	3.36	10.13	1.29	2.50	0.09	0.44
SP-2D	NLS	2.74	1.12	3.34	10.77	0.96	3.21	0.13	0.65
SP-2D	MLE	2.55	1.09	3.24	8.24	1.35	1.89	0.11	0.53
CS-2D	LARS	2.74	1.13	3.42	1.58	0.99	3.25	0.14	0.65
CS-2D	StOMP	3.24	1.11	3.23	1.03	0.12	3.04	1.12	1.45
SnP-3D	CBF	1.47	1.23	3.21	177.02	0.97	2.35	0.77	1.04
SnP-3D	MUSIC	1.19	1.09	2.85	167.46	1.31	1.53	0.48	0.57
SnP-3D	MLE	1	1.03	2.78	167.07	1.28	1.66	0.34	0.48

$\sigma_{H_{\text{ALS}}}$ is 0.24 m and 0.07 m for the first and second object, respectively. SnP-2D refers to spectral nonparametric methods using 2-D images, SP stands for spectral parametric, and CS stands for compressive sensing.

The second object (red rectangle in Fig. 5) used for the performance analysis is a large portion of grasslands surrounding the runway and some asphalt-covered roads. The mean and standard deviation of the height above ground of the object were 0.27 m and 0.07 m. Based on the results listed in Table III, the methods using 2-D geocoded SLCs were used to produce DSMs with a more accurate height estimate of the ground surface than those using 3-D images. The best performance was given by MUSIC-based methods, while CBF and StOMP were ranked last.

V. DISCUSSION AND CONCLUSION

A. Discussion

In this work, we evaluated multiple tomographic reconstruction methods to derive DSMs based on either 2-D or 3-D geocoded SLC products. The methods were applied to single-pass multibaseline InSAR data at *Ka*-band. Based on the κ_r, the global mean and standard deviation of the height error in Tables II and III, methods using 3-D geocoded SLCs were seen to better reproduce the ALS-based DSM than those using 2-D geocoded SLC products. Visual inspection of the corresponding DSMs revealed that a pulse compression method introduced fewer outliers when using a covariance matrix computed at each height level. This suggests that these methods are less sensitive to phase noise or involve weaker sidelobes in elevation. The local and global standard deviation of height error was utilized as a measure of this property. When the methods were based on 2-D images, the parametric spectral estimators performed best, followed by the techniques based on compressive sensing theory and the nonparametric spectral estimators. As expected, this performance ranking is similar to those reported when images are used in radar geometry [26]. For data acquired with a large number of baselines, methods operating with 2-D geocoded SLCs can be expected to provide DSMs with fewer errors due to sidelobes and phase noise than those shown here.

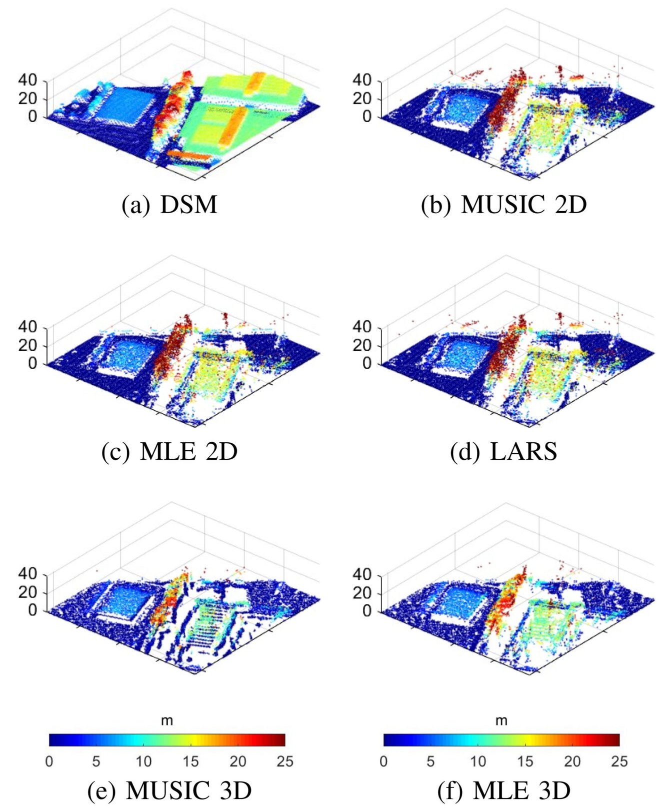


Fig. 6. 3-D reconstruction of buildings in Memmingen using different pulse compression methods in elevation. The colormap encodes the height above ground in meters.

In that case, the methods using 3-D geocoded SLCs and the parametric spectral estimation methods might be impractical due to the required additional computation time. Red rectangles in Figs. 3(b) and 5(a) indicate objects with flat surfaces, and thus, the parameter $\sigma_{H_{\mathrm{SAR}}}$ in the tables relates to the smoothness of the rooftops or ground of the corresponding DSMs. A measure of the planarity [66] of the different DSMs could be an alternative quality indicator for comparing the performances of the methods. In terms of the height accuracy given by $\Delta\mu_H$, the parametric spectral estimators ranked in first place, followed by the compressive sensing-based approaches. The methods using 3-D images performed worst, providing results comparable in quality to the nonparametric spectral estimators operating with 2-D geocoded SLC products.

The DSMs shown in this work can be further improved by applying additional denoising techniques. Denoising of the covariance matrices could be performed with methods such as [54]–[56] to reduce the standard deviation of the height error. The computation of the covariance matrices can include diagonal loading to achieve super-resolution, as in the CBF variants [32]–[34]. However, this process can introduce 1) a loss in the signal to interference noise ratio, 2) an incorrect estimate of the output power if the iterative process involved in the computation of the loading factor does not converge to a valid solution, and 3) an incorrect estimate of the output power without *a priori* knowledge of the error bound of the steering vector. Nonparametric spectral estimation methods can offer a better performance when exploiting volume denoising [67] before detection of maxima. Exploiting the value of the entropy and increasing the multilooking factor can help us to reduce the standard deviation of the height error of a certain DSM as the number of outliers decreases accordingly. Postprocessing techniques, such as clustering, or the use of *a priori* knowledge of the scene could be applied to further improve the DSMs. For single-pass interferometric airborne SAR datasets, single-look tomographic processing is challenging due to the typically limited number of baselines. To preserve details, we recommend using a DEM with a fine sample interval in northing and easting and computing the covariance matrices by adaptively averaging inside a sliding window [54]. However, the sample interval should be large enough to ensure that the covariance matrices do not become singular.

The presence of speckle noise can affect the displacement field and lead to inaccurate offsets when deriving the normal dimension. This can be mitigated by using a 3-D image obtained by incoherent summation of the 3-D K geocoded input images, or by applying image denoising techniques, such as the work in [67]. Raytracing methods are valid alternatives to derive the elevation dimension without a need for focusing one 3-D geocoded image.

In map geometry, a DSM can also be obtained by thresholding the power of the pulse compressed signal along elevation. However, evaluation of those approaches is not trivial, as the computation of the threshold plays a key role in the quality of the resulting DSM. If a DEM of the area of interest is not available, the user can utilize a plane at the lowest ground height or located below the ground. In those cases, the computation time increases since we have to focus more images, and the resulting DSMs might include some points below the ground due to phase noise or sidelobes.

The computational complexity involved in the process of deriving DSMs using SAR images in map geometry is significantly greater than algorithms applied in radar geometry. A possible solution to reduce the computation time is the use of a coarser height sample interval at a cost of degrading the height precision. Interpolation of the signal in elevation could be an alternative solution to be studied.

VII. Conclusion

In this work, we presented steps required to derive a DSM from a set of geocoded SAR images acquired in a tomographic configuration. We described diverse tomographic reconstruction techniques and introduced the necessary operations so that the methods can operate based on 2-D or 3-D images. The performance of the methods was evaluated and compared in terms of their capability to reproduce the corresponding ALS-based reference DSM as well as the height accuracy of the resulting DSM. Both numerical and visual inspections indicated that methods using 3-D geocoded SLCs yielded the best performance, i.e., the resulting DSMs had fewer outliers and retained more information of the illuminated area at a cost in degraded height accuracy. In terms of computational complexity, height accuracy, and number of errors, compressive sensing methods operating with 2-D SAR geocoded SLCs offered high accuracy with comparatively few outliers. Comparisons of DSMs obtained with SAR images in map and radar geometry is foreseen in future work. Adaptations of the parametric spectral estimators and the compressive sensing-based methods to operate with 3-D images will also be studied.

Acknowledgment

The authors would like to thank O. Frey and E. Meier for fruitful discussions and the reviewers for their helpful contributions in improving this article and also would like to thank the Fraunhofer FHR team and the pilots and all institutions involved in the MEMPHIS data collection during the Swiss and German campaigns in 2009 and 2013.

References

[1] T. G. Farr *et al.*, "The shuttle radar topography mission," *Rev. Geophys.*, vol. 45, no. 2, pp. 1–33, 2007.

[2] G. Corsini, M. Diani, F. Lombardini, and G. Pinelli, "Simulated analysis and optimization of a three-antenna airborne InSAR system for topographic mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2518–2529, Sep. 1999.

[3] P. Gamba and B. Houshmand, "Digital surface models and building extraction: A comparison of IFSAR and LIDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 4, pp. 1959–1968, Jul. 2000.

[4] H. Aghababaee, G. Ferraioli, G. Schirinzi, and V. Pascazio, "Regularization of SAR tomography for 3-D height reconstruction in urban areas," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 648–659, Feb. 2019.

[5] M. Mariotti D'Alessandro and S. Tebaldini, "Digital terrain model retrieval in tropical forests through P-band SAR tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6774–6781, Sep. 2019.

[6] C. Rambour, A. Budillon, A. C. Johnsy, L. Denis, F. Tupin, and G. Schirinzi, "From interferometric to tomographic SAR: A review of synthetic aperture radar tomography-processing techniques for scatterer unmixing in urban areas," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 2, pp. 6–29, Jun. 2020.

[7] J. Proulx-Bourque, R. Magagi, N. T. O'Neill, and P. Gravel, "Digital surface model generation using optimal RADARSAT-2 images," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 1198–1201.

[8] X. X. Zhu and M. Shahzad, "Facade reconstruction using multiview spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3541–3552, Jun. 2013.

[9] M. Werner, "Shuttle radar topography mission (SRTM): Experience with the x-band SAR interferometer," in *Proc. CIE Int. Conf. Radar Proc.*, 2001, pp. 634–638.

[10] G. Krieger *et al.*, "TanDEM-X: A satellite formation for high-resolution SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3317–3341, Nov. 2007.

[11] P. Rosen *et al.*, "Synthetic aperture radar interferometry," *Proc. IEEE*, vol. 88, no. 3, pp. 333–382, Mar. 2000.

[12] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline l-band data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.

[13] P. Prats, K. A. Câmara De Macedo, A. Reigber, R. Scheiber, and J. J. Mallorqui, "Comparison of topography-and aperture-dependent motion compensation algorithms for airborne SAR," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 349–353, Jul. 2007.

[14] E. Meier, U. Frei, and D. Nüesch, *Precise Terrain Corrected Geocoded Images*. Karlsruhe, Germany: Wichmann, 1993, pp. 173–186.

[15] X. X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.

[16] Y. Yu, M. Mariotti d'Alessandro, S. Tebaldini, and M. Liao, "Signal processing options for high resolution SAR tomography of natural scenarios," *Remote Sens.*, vol. 12, no. 10, pp. 1–23, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/10/1638

[17] Y. Zheng and H. A. Zebker, "Phase correction of SLC radar images for user-friendly efficient interferogram formation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2694–2701, Jun. 2017.

[18] M. I. Duersch and D. G. Long, "Backprojection SAR interferometry," *Int. J. Remote Sens.*, vol. 36, no. 4, pp. 979–999, 2015.

[19] O. Frey and E. Meier, "Combining time-domain back-projection and capon beamforming for tomographic SAR processing," in *Proc. IEEE Int. Geosci. Remote Sensing Symp.*, 2008, vol. 2, pp. 445–448.

[20] O. Frey and E. Meier, "3-D time-domain SAR imaging of a forest using airborne multibaseline data at L-and P-band," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3660–3664, Oct. 2011.

[21] S. Tebaldini, T. Nagler, H. Rott, and A. Heilig, "Imaging the internal structure of an alpine glacier via L-band airborne SAR tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7197–7209, Dec. 2016.

[22] F. Banda, J. Dall, and S. Tebaldini, "Single and multipolarimetric P-band SAR tomography of subsurface ice structure," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2832–2845, May 2016.

[23] E. M. Domínguez, C. Magnard, E. Meier, D. Small, M. E. Schaepman, and D. Henke, "A back-projection tomographic framework for VHR SAR image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4470–4484, Jul. 2019.

[24] O. Ponce, R. Scheiber, P. Prats, I. Hajnsek, and A. Reigber, "Multidimensional airborne holographic SAR tomography reconstruction for glaciers at L-/P-band," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 9–12.

[25] L. M. Ulander, H. Hellsten, and G. Stenstrom, "Synthetic-aperture radar processing using fast factorized back-projection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 3, pp. 760–776, Jul. 2003.

[26] X. Zhu, "Very high resolution tomographic SAR inversion for urban infrastructure monitoring—A sparse and nonlinear tour," Ph.D. thesis, Deutsche Geodätische Kommission, Reihe C, Nr. 666, Verlag der Bayerischen Akademie der Wissenschaften, 2011.

[27] A. Reigber, F. Lombardini, F. Viviani, M. Nannini, and A. Martinez del Hoyo, "Three-dimensional and higher-order imaging with tomographic SAR: Techniques, applications, issues," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 2915–2918.

[28] F. Lombardini, F. Cai, and D. Pasculli, "Spaceborne 3-D SAR tomography for analyzing garbled urban scenarios: Single-look superresolution advances and experiments," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 960–968, Apr. 2013.

[29] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.

[30] 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.

[31] F. Lombardini and A. Reigber, "Adaptive spectral estimation for multibaseline SAR tomography with airborne L-band data," in *Proc. Int. Geosci. Remote Sensing Symp.*, 2003, vol. 3, pp. 2014–2016.

[32] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[33] Z. Wang, J. Li, and P. Stoica, "Comparison between norm constrained and robust Capon beamformers," in *Proc. IEEE Int. Symp. Phased Array Syst. Technol.*, 2003, pp. 170–175.

[34] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2407–2423, Sep. 2004.

[35] 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.

[36] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[37] A. Paulraj, B. Ottersten, R. Roy, A. Swindlehurst, G. Xu, and T. Kailath, "16 subspace methods for directions-of-arrival estimation," *Handbook Statist.*, vol. 10, pp. 693–739, 1993.

[38] Y. Huang, L. Ferro-Famil, and A. Reigber, "Under-foliage object imaging using SAR tomography and polarimetric spectral estimators," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2213–2225, Jun. 2011.

[39] O. Besson and P. Stoica, "Nonlinear least-squares approach to frequency estimation and detection for sinusoidal signals with arbitrary envelope," *Digit. Signal Process.*, vol. 9, pp. 45–56, 1999.

[40] X. X. Zhu and R. Bamler, "Tomographic SAR inversion by L1-norm regularization-the compressive sensing approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, Oct. 2010.

[41] M. Schmitt, "Reconstruction of urban surface models from multi-aspect and multi-baseline interferometric SAR," Ph.D. thesis, Institut für Photogrammetrie und Kartographie, Munich, Germany, Technische Universität München 2014.

[42] M. Schmitt, "Three-dimensional reconstruction of urban areas by multi-aspect TomoSAR data fusion," in *Proc. Joint Urban Remote Sens. Event*, 2015, pp. 1–4.

[43] M. Schmitt, C. Magnard, T. Brehm, and U. Stilla, "Towards airborne single pass decimeter resolution SAR interferometry over urban areas," in *Proc. Photogrammetric Image Anal.*, 2011, vol. 6952, pp. 197–208.

[44] X. X. Zhu and R. Bamler, "Tomographic SAR inversion by L_1-Norm regularization—The compressive sensing approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3839–3846, Oct. 2010.

[45] G. Fornaro, F.Lombardini, A. Pauciullo, D. Reale, and F. Viviani, "Tomographic processing of interferometric SAR data: Developments, applications, and future research perspectives," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 41–50, Jul. 2014.

[46] A. Budillon and G. Schirinzi, "GLRT based on support estimation for multiple scatterers detection in SAR tomography," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 3, pp. 1086–1094, Mar. 2016.

[47] A. Budillon, A. C. Johnsy, and G. Schirinzi, "A fast support detector for superresolution localization of multiple scatterers in SAR tomography," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2768–2779, Jun. 2017.

[48] S. Verde, A. Pauciullo, D. Reale, and G. Fornaro, "Multiresolution detection of persistent scatterers: A performance comparison between multilook GLRT and CAESAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3088–3103, Apr. 2020.

[49] G. Fornaro, A. Pauciullo, D. Reale, and S. Verde, "Multilook SAR tomography for 3-D reconstruction and monitoring of single structures applied to COSMO-SKYMED data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2776–2785, Jul. 2014.

[50] M. Soumekh, *Synthetic Aperture Radar Signal Processing*. New York, NY, USA: Wiley, 1999.

[51] 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.

[52] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[53] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2963–2972, Dec. 2005.

[54] M. Schmitt and U. Stilla, "Adaptive multilooking of airborne single-pass multi-baseline InSAR stacks," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 305–312, Jan. 2014.

[55] O. D'Hondt, C. López-Martínez, S. Guillaso, and O. Hellwich, "Nonlocal filtering applied to 3-D reconstruction of tomographic SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 272–285, Jan. 2018.

[56] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, "NL-SAR: A unified nonlocal framework for resolution-preserving (Pol)(In) SAR denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2021–2038, Apr. 2015.

[57] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[58] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[59] R. Z. Schneider and D. Fernandes, "Entropy concept for change detection in multitemporal SAR images," in Proc. *Eur. Conf. Synthetic Aperture Radar*, 2002, pp. 221–224.

[60] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 98–109, Feb. 2014.

[61] X. X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 247–258, Jan. 2012.

[62] G. D. Martín-del-Campo-Becerra, S. A. Serafín-García, A. Reigber, and S. Ortega-Cisneros, "Parameter selection criteria for Tomo-SAR focusing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1580–1602, 2021.

[63] H. Schimpf, H. Essen, S. Boehmsdorff, and T. Brehm, "MEMPHIS—A fully polarimetric experimental radar," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2002, vol. 3, pp. 1714–1716.

[64] D. Henke, E. Mendez Dominguez, D. Small, M. E. Schaepman, and E. Meier, "Moving target tracking in single- and multichannel SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3146–3159, Jun. 2015.

[65] C. Magnard, M. Frioud, D. Small, T. Brehm, H. Essen, and E. Meier, "Processing of MEMPHIS ka-band multibaseline interferometric SAR data: From raw data to digital surface models," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2927–2941, Jul. 2014.

[66] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS J. Photogrammetry Remote Sens.*, vol. 105, pp. 286–304, 2015.

[67] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013.

Elías Méndez Domínguez (Member, IEEE) received the M.Sc. degree in telecommunications engineering from the University of Vigo, Vigo, Spain, in 2007, and the Ph.D. degree with focus on SAR image change detection from the University of Zürich, Zürich, Switzerland, in 2018.

In 2008, he joined the traineeship program with the Sensors and Radar Technologies and Cybersecurity Unit, I.P.S.C. Institute (Joint Research Centre), Ispra, Italy. In 2009, he joined the Surface Radar Unit, THALES Nederland B.V., Delft, The Netherlands, under the FP7 Marie Curie Early Stage Research Training Program in collaboration with the International Research Centre for Telecommunications and Radar, Delft University of Technology, Delft, The Netherlands. In 2010, he joined the Remote Sensing Laboratories, Department of Geography, University of Zürich. His research interests include synthetic aperture radar (SAR) image processing, and change detection.

David Small (Senior Member, IEEE) received the B.A.Sc. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1988, the M.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, in 1991, and the Ph.D. degree with focus on the development of a processing chain for generating digital elevation models from repeat-pass ERS-1/2 data, from the Remote Sensing Laboratories, Department of Geography, University of Zürich, Zürich, Switzerland, in 1998.

Since 2013, he has lead the Environmental SAR group, SARLab, Remote Sensing Laboratories, University of Zürich. In addition to geometric and radiometric calibration of synthetic aperture radar (SAR) imagery, he works on standardized approaches for terrain flattening to enable analysis ready data products. These are being demonstrated for a wide variety of applications, from SAR-based wet snow mapping in mountainous terrain, as well as applications in forestry and wide-area cryosphere mapping.

Daniel Henke received the Dipl.-Inf. degree in computer science from Albert-Ludwigs University, Freiburg, Germany, in 2007, and the Ph.D. degree with focus on the development of methods to track objects in remotely sensed data, from the University of Zürich, Zürich, Switzerland, in 2014.

From 2014 to 2017, he was a Postdoctoral Researcher under a University Research Grant with the SARLab, University of Zürich, and under the SNF Postdoctoral Mobility Grant with the Department of Earth System Science, School of Physics, University of California, Irvine, CA, USA, and the Smithsonian Tropical Research Institute, Panama. Since 2017, he has been leading the Computational SAR Group, SARLab, Remote Sensing Laboratories, University of Zürich. His research interests include synthetic aperture radar processing, moving target indication and change detection techniques, data fusion, signal processing, probabilistic modeling, and earth system science applications using remote sensing data.