SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2018, Vol. 61 042302:1–042302:9
doi: 10.1007/s11432-016-9068-y

# Baseline distribution optimization and missing data completion in wavelet-based CS-TomoSAR

Hui BI[1,2*], Jianguo LIU[3], Bingchen ZHANG[1] & Wen HONG[1]

[1]*Science and Technology on Microwave Imaging Laboratory, Institute of Electronics, Chinese Academy of Sciences, Beijing* 100190*, China;*
[2]*University of Chinese Academy of Sciences, Beijing* 100049*, China;*
[3]*Department of Earth Science and Engineering, Imperial College London, London* SW7 2AZ*, U.K.*

Received 4 November 2016/Revised 6 February 2017/Accepted 23 March 2017/Published online 25 August 2017

**Abstract** In this paper, we propose a coherence of measurement matrix-based baseline distribution optimization criterion, together with an $L_1$ regularization missing data completion method for unobserved baselines (not belonging to the actual baseline distribution), to facilitate wavelet-based compressive sensing-tomographic synthetic aperture radar imaging (CS-TomoSAR) in forested areas. Using $M$ actual baselines, we first estimate the optimal baseline distribution with $N$ baselines ($N > M$), including $N - M$ unobserved baselines, via the proposed coherence criterion. We then use the geometric relationship between the actual and unobserved baseline distributions to reconstruct the transformation matrix by solving an $L_1$ regularization problem, and calculate the unobserved baseline data using the measurements of actual baselines and the estimated transformation matrix. Finally, we exploit the wavelet-based CS technique to reconstruct the elevation via the completed data of $N$ baselines. Compared to results obtained using only the data of actual baselines, the recovered image based on the dataset obtained by our proposed method shows higher elevation recovery accuracy and better super-resolution ability. Experimental results based on simulated and real data validated the effectiveness of the proposed method.

**Keywords** tomographic synthetic aperture radar imaging (TomoSAR), compressive sensing (CS), baseline distribution optimization, coherence of measurement matrix

**Citation** Bi H, Liu J G, Zhang B C, et al. Baseline distribution optimization and missing data completion in wavelet-based CS-TomoSAR. Sci China Inf Sci, 2018, 61(4): 042302, doi: 10.1007/s11432-016-9068-y

## 1 Introduction

Tomographic synthetic aperture radar imaging (TomoSAR) is a radar imaging modality which exploits the acquisitions of multi-baseline with slightly different incidence angles to synthesize an elevation aperture, and hence obtain the three-dimensional (3-D) synthetic aperture radar (SAR) image by reconstructing the reflectivity function along the elevation direction using spectral analysis or compressive sensing (CS) methods [1–4]. In TomoSAR, elevation recovery quality is mainly decided by the length of the elevation aperture, the number of baselines, and the baseline distribution. Because the acquisition of a large number of baselines is time-consuming and often constrained by costs and feasibility, it is essential to obtain many more baselines with larger apertures so as to achieve a high-resolution elevation image [5,6].

CS, which was proposed by Donoho [7] and Candès et al. [8] in 2006, is an important development in sparse signal processing which allows sparse signals to be recovered from significantly fewer samples, as

* Corresponding author (email: bihui1991@163.com)

 

required by the well-known Shannon-Nyquist sampling theorem [9, 10]. In 2009, CS was introduced to TomoSAR, and has been used to reconstruct sparse elevation distributions successfully [3,11,12]. However, in forested areas, the sparse elevation distribution is generally unsatisfied, and thus CS cannot be used for recovery directly. To tackle this issue, Aguilera et al. [13] introduced a wavelet basis to represent the elevation reflectivity function sparsely, and obtained images with higher elevation resolution by solving an $L_1$ regularization problem. In CS-based reconstruction, measurement matrices must satisfy certain conditions, including the restricted isometry property (RIP) [14], the null space property (NSP) [14], the exact reconstruction criteria (ERC) [15], and the coherence [14]. In general, it is NP-hard to estimate the RIP, NSP, and ERC for a given matrix, while coherence is relatively easy to compute and can efficiently evaluate the properties of measurement matrices. Therefore, we explore the potential of coherence as the criterion to judge the performance of measurement matrices constructed based on baseline distributions, so as to achieve baseline distribution optimization in wavelet-based CS-TomoSAR.

In this paper, a coherence of measurement matrix-based baseline distribution optimization criterion is proposed together with an $L_1$ regularization missing data completion technique of unobserved baselines (not belonging to the actual baseline distribution), to facilitate wavelet-based CS-TomoSAR in forested areas. From $M$ actual baselines with distribution $\boldsymbol{B}_M = [b_1, b_2, \ldots, b_M]$, we first estimate the optimal $N$ baseline distribution $\boldsymbol{B}_N = [\boldsymbol{B}_M, \boldsymbol{A}_{N-M}] = [b_1, b_2, \ldots, b_M, a_1, a_2, \ldots, a_{N-M}]$ by using the coherence criterion, where $\boldsymbol{A}_{N-M} = [a_1, a_2, \ldots, a_{N-M}]$ represents the virtual baseline distribution to be estimated. The transformation matrix between $\boldsymbol{B}_M$ and $\boldsymbol{A}_{N-M}$ is then reconstructed by solving an $L_1$ regularization optimization problem. The data of $\boldsymbol{A}_{N-M}$ is calculated based on the measurements of $\boldsymbol{B}_M$ and the estimated transformation matrix. Finally, a wavelet-based CS approach is used to reconstruct the elevation power in the forested area via the completed data of $\boldsymbol{B}_N$. Compared to reconstructed results obtained using only the data of $M$ actual baselines, the recovered image based on the completed dataset acquired by the proposed method shows better performance, e.g., higher elevation recovery accuracy, better resolution. Experimental results based on simulated and real airborne data verified the validity of the proposed method.

## 2 Wavelet-based CS TomoSAR

In TomoSAR, let $\boldsymbol{G}_{\mathrm{K}} = [g_1, \ldots, g_M]^{\mathrm{T}}$ and $\boldsymbol{\gamma} = [\gamma(s_1), \ldots, \gamma(s_L)]^{\mathrm{T}}$ denote the actual baseline measurements $\boldsymbol{B}_M$ and the elevation's discrete complex reflectivity function at a specific azimuth-range resolution cell. The TomoSAR imaging model can be represented as

$$\boldsymbol{G}_{\mathrm{K}} \;=\; \boldsymbol{\Phi}_{\mathrm{K}}\boldsymbol{\gamma}, \tag{1}$$

where $\boldsymbol{\Phi}_{\mathrm{K}} \in \mathbb{C}^{M\times L}$ is the measurement matrix, which is constructed according to the imaging geometry with $\boldsymbol{\Phi}_{\mathrm{K}}(m,l) = \mathrm{e}^{\mathrm{j}\frac{4\pi}{\lambda r}b_m s_l}$. $\lambda$ is the wavelength, $r$ is the slant range, $b_m$ is the elevation aperture of the $m$th acquisition in $\boldsymbol{B}_M$, and $s_l\,(l = 1, 2, \ldots, L)$ is the $l$th position along the elevation. Similar to (1), the TomoSAR imaging model based on the unobserved baseline distribution $\boldsymbol{A}_{N-M} = [a_1, a_2, \ldots, a_{N-M}]$ is

$$\boldsymbol{G}_{\mathrm{U}} \;=\; \boldsymbol{\Phi}_{\mathrm{U}}\boldsymbol{\gamma}, \tag{2}$$

where measurement matrix $\boldsymbol{\Phi}_{\mathrm{U}} \in \mathbb{C}^{(N-M)\times L}$ is constructed based on the imaging geometry and $\boldsymbol{A}_{N-M}$. By combining (1) and (2), we can obtain the imaging model based on the $N$ baselines with distribution $\boldsymbol{B}_N = [\boldsymbol{B}_M, \boldsymbol{A}_{N-M}] = [b_1, b_2, \ldots, b_M, a_1, a_2, \ldots, a_{N-M}]$ as

$$[\boldsymbol{G}]_{N\times L} \;=\; \begin{bmatrix}\boldsymbol{G}_{\mathrm{K}}\\ \boldsymbol{G}_{\mathrm{U}}\end{bmatrix} \;=\; \begin{bmatrix}\boldsymbol{\Phi}_{\mathrm{K}}\\ \boldsymbol{\Phi}_{\mathrm{U}}\end{bmatrix}\boldsymbol{\gamma} \;=\; \boldsymbol{\Phi}_{N\times L}\boldsymbol{\gamma}_{L\times 1}. \tag{3}$$

Then, the covariance matrix $\boldsymbol{C} \in \mathbb{C}^{N\times N}$ of measurements $\boldsymbol{G}$ can be calculated as

$$\boldsymbol{C} \;=\; \mathrm{E}\left(\boldsymbol{G}\boldsymbol{G}^{\mathrm{H}}\right) \;=\; \boldsymbol{\Phi}\mathrm{diag}\left(\boldsymbol{p}\right)\boldsymbol{\Phi}^{\mathrm{H}}, \tag{4}$$

where symbols $\mathrm{E}(\cdot)$ and $(\cdot)^{\mathrm{H}}$ are the expectation and conjugate transpose operators, respectively, and $\mathrm{diag}(\boldsymbol{p}) \in \mathbb{R}^{L\times L}$ is a diagonal matrix with a major diagonal corresponding to the non-negative elevation power distribution $\boldsymbol{p} = [|\gamma(s_1)|^2, \ldots, |\gamma(s_L)|^2]^{\mathrm{T}}$ and the zero elements off major diagonal [13]. A precondition of CS-TomoSAR imaging is a sparse elevation distribution. However, the reflectivity function $\boldsymbol{\gamma}$ is always non-sparse in a forested area; therefore, an orthonormal basis, $\boldsymbol{\Psi} \in \mathbb{C}^{L\times L}$, is needed to project $\boldsymbol{\gamma}$ sparsely as $\boldsymbol{\alpha} \in \mathbb{C}^{L\times 1}$ with $\boldsymbol{\alpha} = \boldsymbol{\Psi}^{\mathrm{H}} \cdot \boldsymbol{\gamma}$, where $\boldsymbol{\alpha}$ is a sparse vector with $K$ non-zero elements and $K \ll L$. We can then recover $\boldsymbol{p}$ by solving the following optimization problem

$$\min_{\hat{\boldsymbol{p}}} \left\|\boldsymbol{\Psi}^{\mathrm{H}}\hat{\boldsymbol{p}}\right\|_1 \quad \text{s.t.} \quad \left\|\boldsymbol{C} - \boldsymbol{\Phi}\mathrm{diag}(\hat{\boldsymbol{p}})\boldsymbol{\Phi}^{\mathrm{H}}\right\|_F \leqslant \varepsilon, \tag{5}$$

where $\hat{\boldsymbol{p}}$ is the reconstructed elevation power, $\varepsilon$ is a constant decided by the noise level, and operator $\|\cdot\|_F$ is the Frobenius norm of a matrix.

## 3 Proposed method

### 3.1 Coherence of measurement matrix-based baseline distribution optimization

In CS-TomoSAR, elevation reconstruction quality is dependent on the performance of measurement matrix $\boldsymbol{\Phi}$. Meanwhile, $\boldsymbol{\Phi}$ is constructed based on the baseline distribution. Therefore, an optimal baseline distribution is very helpful in the improvement of CS-TomoSAR imaging quality. In this paper, a coherence criterion is introduced to evaluate the properties of measurement matrices so as to obtain the optimal distribution of a fixed number of baselines in wavelet-based CS-TomoSAR.

In the proposed method, we first estimate the optimal baseline distribution $\boldsymbol{B}_N$ with $N$ baselines using the coherence criterion [16]

$$\min_{\boldsymbol{\Phi}} \left\{ \frac{1}{L(L-1)} \sum_{l\neq l'} \mu_{ll'}(\boldsymbol{\Phi}) \right\} \quad \text{s.t.} \quad d_p(\boldsymbol{\Phi})/L \leqslant \beta_{p\%}, \tag{6}$$

where coherent coefficient $\mu_{ll'}(\boldsymbol{\Phi})$ between elements $u_l$ and $u_{l'}$ in the $l$ and $l'$ columns of $\boldsymbol{\Phi}$ is

$$\mu_{ll'}(\boldsymbol{\Phi}) = \frac{\left|u_{l'}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}u_l\right|}{\sqrt{u_l^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}u_l u_{l'}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}u_{l'}}}, \tag{7}$$

and the support length $d_{p\%}(\boldsymbol{\Phi})$ of the coherent coefficient is

$$d_{p\%}(\boldsymbol{\Phi}) = \max_{(l,l')\in\Lambda_{p\%}} |l-l'|, \tag{8}$$

where $\Lambda_{p\%}$ is the index set that denotes the maximum $p\%$ portion of the coherent coefficient energy, and

$$\Lambda_{p\%} = \left\{(l,l') \,|\, l > l', \mu_{ll'} \geqslant \xi_{p\%}\right\}, \quad \text{where} \quad \sum_{\Lambda_{p\%}} \mu_{ll'}^2(\boldsymbol{\Phi}) = p\% \cdot \sum_{l>l'} \mu_{ll'}^2(\boldsymbol{\Phi}), \tag{9}$$

where $\xi_{p\%}$ is a threshold, and $0 < \beta_{p\%} < 1$ is the upper bound of the support length of the $p\%$ coherent coefficient. The meaning of the proposed optimal criterion in (6) is that the smaller the average coherent coefficient, the better performance of sparse recovery in the elevation when the support length satisfies a proper error scope. If $\boldsymbol{\Phi}$ is an orthogonal matrix, we have $\forall l \neq l'$, $\mu_{ll'}(\boldsymbol{\Phi}) = 0$, and $d_p(\boldsymbol{\Phi}) = 0$. The simulated Annealing (SA) algorithm [17] is used to solve the problem in (6). Then, the optimal baseline distribution $\boldsymbol{B}_N$ including actual baselines $\boldsymbol{B}_M$ and unobserved baselines $\boldsymbol{A}_{N-M}$ can be derived.

It should be noted that in the procedure for baseline distribution optimization, the location of $\boldsymbol{B}_M$ is changeless. Meanwhile, we traverse many kinds of possible distributions of $\boldsymbol{A}_{N-M}$ so as to obtain the optimal baseline distribution $\boldsymbol{B}_N$ based on the coherence criterion shown in (6). In addition, the

elevation aperture length will not be changed in this processing.

### 3.2 Missing data completion via $L_1$ regularization

After acquiring the optimal baseline distribution $\boldsymbol{B}_N$ based on the coherence criterion, an $L_1$ regularization based technique is used to estimate the data of $\boldsymbol{A}_{N-M}$ by means of the measurements of $\boldsymbol{B}_M$ and the geometric relationship between $\boldsymbol{B}_M$ and $\boldsymbol{A}_{N-M}$. Let $\boldsymbol{V} \in \mathbb{C}^{(N-M)\times M}$ denote the transformation matrix between $\boldsymbol{G}_{\mathrm{K}}$ and $\boldsymbol{G}_{\mathrm{U}}$; then, we have

$$\boldsymbol{G}_{\mathrm{U}} \;=\; \boldsymbol{V}\boldsymbol{G}_{\mathrm{K}} \;=\; \boldsymbol{V}\boldsymbol{\Phi}_{\mathrm{K}}\boldsymbol{\gamma}. \tag{10}$$

Therefore, the precondition for recovering $\boldsymbol{G}_{\mathrm{U}}$ is the accurate estimation of $\boldsymbol{V}$ by

$$\boldsymbol{\Phi}_{\mathrm{U}} \;=\; \boldsymbol{V}\boldsymbol{\Phi}_{\mathrm{K}}, \tag{11}$$

where $\boldsymbol{\Phi}_{\mathrm{U}}$ and $\boldsymbol{\Phi}_{\mathrm{K}}$ are known based on the imaging geometric relationship between the sensors and the chosen azimuth-range resolution cell. The transpose operation $(\cdot)^{\mathrm{T}}$ for (11) can be rewritten as

$$\boldsymbol{\Phi}_{\mathrm{U}}^{\mathrm{T}} \;=\; \boldsymbol{\Phi}_{\mathrm{K}}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}, \tag{12}$$

which can be solved by the following optimization problem:

$$\hat{\boldsymbol{V}} \;=\; \left[\min_{\boldsymbol{V}^{\mathrm{T}}}\left\{\left\|\boldsymbol{\Phi}_{\mathrm{U}}^{\mathrm{T}} - \boldsymbol{\Phi}_{\mathrm{K}}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{T}}\right\|_F^2 \;+\; \alpha\left\|\boldsymbol{V}^{\mathrm{T}}\right\|_1\right\}\right]^{\mathrm{T}}, \tag{13}$$

where $\hat{\boldsymbol{V}}$ is the estimated transformation matrix based on $L_1$ regularization. The details of the iterative recovery procedure for solving the optimization problem (13) are shown in Algorithm 1.

**Algorithm 1** Iterative algorithm for $L_1$ regularization reconstruction

**Input:** Mapping matrix $\boldsymbol{\Phi}_{\mathrm{K}}$ and $\boldsymbol{\Phi}_{\mathrm{U}}$; Iterative parameter $\mu$; Error parameter $\sigma$
**Initial:** $\left(\boldsymbol{V}^{\mathrm{T}}\right)^{(0)} = \boldsymbol{0}$
**Iteration:** **While** $i \leqslant I$ and Residual $>\sigma$
**Step 1:** $\boldsymbol{W}^{(i)} = \boldsymbol{\Phi}_{\mathrm{U}}^{\mathrm{T}} - \boldsymbol{\Phi}_{\mathrm{K}}^{\mathrm{T}}\left(\boldsymbol{V}^{\mathrm{T}}\right)^{(i)}$
**Step 2:** $\left(\Delta\boldsymbol{V}^{\mathrm{T}}\right)^{(i)} = \left(\left(\boldsymbol{\Phi}_{\mathrm{K}}^{\mathrm{T}}\right)^{\mathrm{H}}\boldsymbol{\Phi}_{\mathrm{K}}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{\Phi}_{\mathrm{K}}^{\mathrm{T}}\right)^{\mathrm{H}}\boldsymbol{W}^{(i)}$
**Step 3:** $\left(\boldsymbol{V}^{\mathrm{T}}\right)^{(i+1)} = \mathrm{sgn}\left(\left(\boldsymbol{V}^{\mathrm{T}}\right)^{(i)} + \mu\left(\Delta\boldsymbol{V}^{\mathrm{T}}\right)^{(i)}\right)\cdot\max\left(\left(\left(\boldsymbol{V}^{\mathrm{T}}\right)^{(i)} + \mu\left(\Delta\boldsymbol{V}^{\mathrm{T}}\right)^{(i)}\right) - \mu\alpha, 0\right)$
**Step 4:** $\mathrm{Residual} = \left\|\left(\boldsymbol{V}^{\mathrm{T}}\right)^{(i+1)} - \left(\boldsymbol{V}^{\mathrm{T}}\right)^{(i)}\right\|_F^2$
**Step 5:** $i = i + 1$
**end**
**Output:** $\hat{\boldsymbol{V}} = \left(\left(\boldsymbol{V}^{\mathrm{T}}\right)^{(i+1)}\right)^{\mathrm{T}}$

In the iterative recovery algorithm shown in Algorithm 1, parameters $\mu$ and $\alpha$ should be set appropriately. We will briefly discuss their selections in the following.

(1) $\mu$ controls the convergence speed of the iterative algorithm. It should satisfy

$$0 \;<\; \mu^{-1} \;<\; 1\Big/\left\|\boldsymbol{\Phi}^{\mathrm{T}}\right\|_F^2. \tag{14}$$

When $\mu^{-1}$ moves from 0 to $1/\|\boldsymbol{\Phi}^{\mathrm{T}}\|_F^2$, the convergence speed of the proposed iterative algorithm increases, while the precision of the recovered solution decreases. Therefore, the value of $\mu$ should be selected as a compromise between convergence speed and precision. In this paper, we set $\mu = 1/2\|\boldsymbol{\Phi}^{\mathrm{T}}\|_F^2$ to estimation the transformation matrix $\boldsymbol{V}$.

(2) Regularization parameter $\alpha$ controls the reconstructed precision, which is set adaptively according

to the iterative step $i = 1, 2, \ldots, I$ as

$$\alpha^{(i)} = \left|\left(\boldsymbol{V}^{\mathrm{T}}\right)^{(i)} + \mu\left(\Delta\boldsymbol{V}^{\mathrm{T}}\right)^{(i)}\right|_{K+1} \Big/ \mu, \tag{15}$$

where $|(\boldsymbol{V}^{\mathrm{T}})^{(i)} + \mu(\Delta\boldsymbol{V}^{\mathrm{T}})^{(i)}|_{K+1}$ is the $(K+1)$st largest amplitude element of $|(\boldsymbol{V}^{\mathrm{T}})^{(i)} + \mu(\Delta\boldsymbol{V}^{\mathrm{T}})^{(i)}|$ with $K$ being the number of non-zero elements we set in the transformation matrix.

After the above recovery of the transformation matrix, the data $\boldsymbol{G}_{\mathrm{U}}$ from the unobserved baseline distribution $\boldsymbol{A}_{N-M}$ can be estimated based on (10). We can then reconstruct the elevation power by solving the optimization problem shown in (5) based on the completed data $\boldsymbol{G}$ of $\boldsymbol{B}_N$.

## 4 Experiment and discussion

In this section, we present experimental results obtained using simulated and real airborne data along with a performance analysis to illustrate the validity of the proposed method in wavelet-based CS-TomoSAR tailored to the forested area.

### 4.1 Experiment based on simulated data

We set the wavelength and reference slant range of the simulated data as 0.86 m and 4000 m, respectively. For a specific azimuth-range resolution cell, we first exploit the measurements of actual baselines $\boldsymbol{B}_M = [0, 15, 28, 44, 60, 75, 91, 100]$ with $M = 8$, in order to estimate the data of optimal virtual baseline distribution $\boldsymbol{B}_N$ with $N = 15$ using the proposed method; we then exploit the wavelet-based CS algorithm to reconstruct the elevation distribution via the completed dataset by means of the 20 largest wavelet coefficients. For the baseline distribution $\boldsymbol{B}_M$, the Rayleigh resolution in the elevation direction can be approximated by [2]

$$\rho_s = \frac{\lambda r}{2\Delta b} = \frac{0.86 \times 4000}{2 \times 100} = 17.20 \text{ m}. \tag{16}$$

#### 4.1.1 *Recovered accuracy of reflectivity*

We set two scattering areas, representing canopy and ground, with different intensity and scope as the simulated elevation distribution in the forested area (see Figure 1). To a perform meaningful comparison, we artificially introduced Gaussian noise with signal-to-noise ratios (SNRs) equal to $\infty, 5$ dB, 0 dB, and $-5$ dB to the simulated echo data, respectively. Figure 1 shows the normalized profiles reconstructed using the data from the observed actual baseline distribution $\boldsymbol{B}_M$ (Blue) and the estimated optimal virtual baseline distribution $\boldsymbol{B}_N$ (Red), in comparison with the original simulated scene (Black). In Figure 1, we can see that when the SNR equals $\infty$ and 5 dB, compared to the simulated elevation distribution, although the images reconstructed using the data from $\boldsymbol{B}_M$ and $\boldsymbol{B}_N$ have some recovery errors in areas with weaker reflectivity (see the right column in Figure 1), they all can find the scattering center accurately. Meanwhile the recovered profiles based on the completed dataset of $\boldsymbol{B}_N$ have lower sidelobes. When the SNR decreases to 0 dB, compared to results based on data from the actual baseline distribution, the reconstructed profiles based on the completed dataset of the estimated optimal baseline distribution $\boldsymbol{B}_N$ show better estimated accuracy in terms of reflectivity and are much closer to the simulated elevation scene, especially in the canopy area. After adding more noise to the echo data with SNR $= -5$ dB, the images recovered using the data from $\boldsymbol{B}_M$ could not find the scattering center accurately, while the results of $\boldsymbol{B}_N$ could identify the scattering center and successfully recover the elevation power.

#### 4.1.2 *Super-resolving*

The super-resolving ability is discussed in the following. We set two scattering areas (representing ground and canopy) with identical amplitudes as the simulated scene along the elevation with the distance between the two scattering centers changing from 30 m to 0 m. Figure 2 presents the estimated elevations

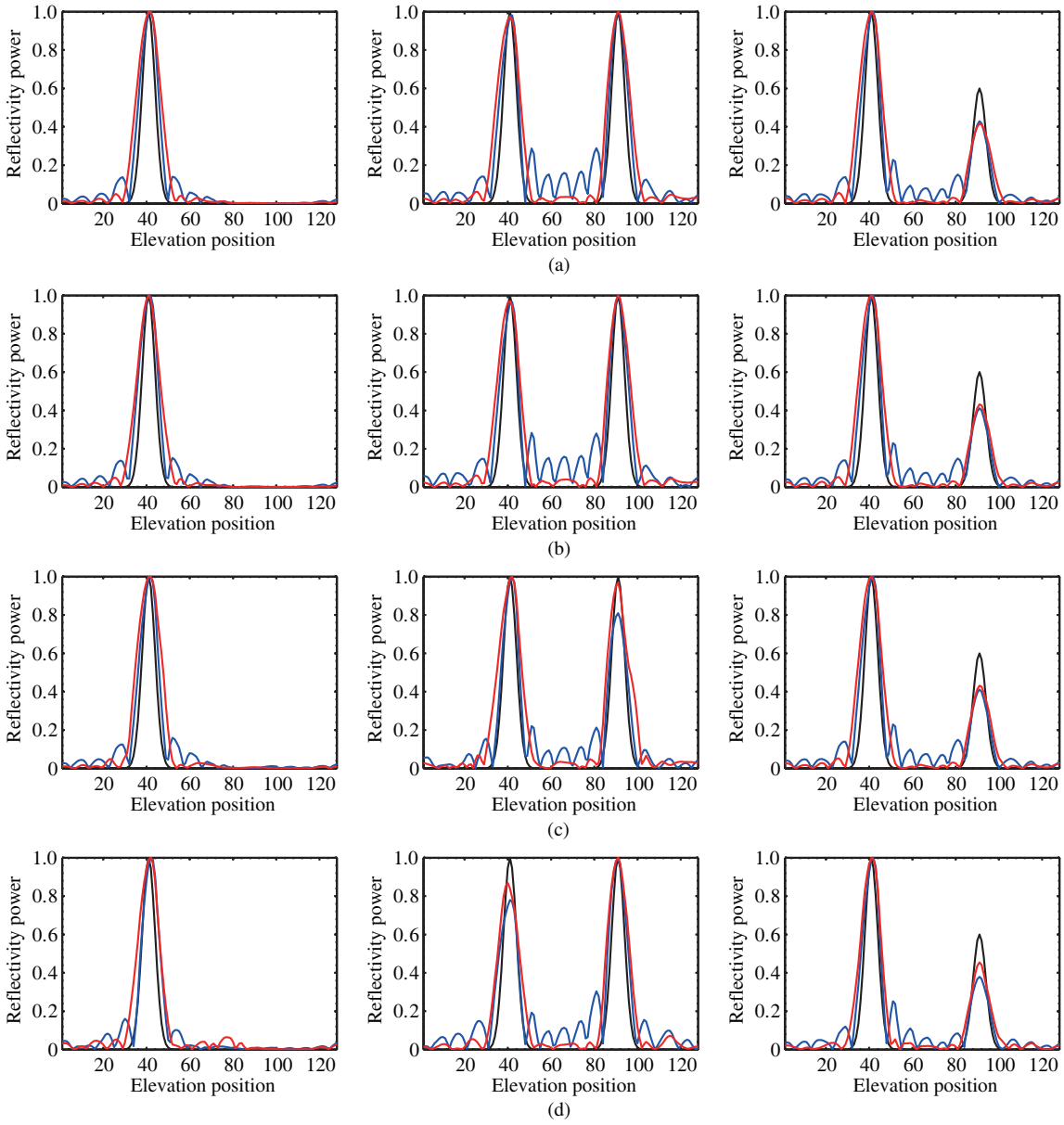


**Figure 1** Recovered normalized elevation profiles based on data from observed actual baseline distribution $\boldsymbol{B}_M = [0, 15, 28, 44, 60, 75, 91, 100]$ with $M = 8$ (Blue) and estimated optimal virtual baseline distribution $\boldsymbol{B}_N$ with $N = 15$ (Red). Two scattering areas with different intensities are used to simulate the elevation distribution in the forested area (Black). The SNRs of the simulated echo data are (a) $\infty$, (b) 5 dB, (c) 0 dB, and (d) $-5$ dB, respectively.

(in meters) of two scattering centers reconstructed by the wavelet-based CS TomoSAR imaging algorithm with data from actual baseline distribution $\boldsymbol{B}_M$ (Upper row) and estimated optimal virtual baseline distribution $\boldsymbol{B}_N$ (Lower row) under different SNRs. In Figure 2, the SNR values are 0 dB, 5 dB, and $\infty$, respectively from the left to the right columns. As calculated in (16), the Rayleigh resolution in the elevation via $\boldsymbol{B}_M$ is approximately 17.2 m. From Figure 2, we can see that no matter which dataset is used in the wavelet-based CS-TomoSAR imaging, the detected distance of two scattering centers in the six plots all less than the Rayleigh resolution. This benefits from the super-resolution ability of the CS-TomoSAR imaging technique. When SNR = 0 dB, the detected distance in the image recovered using the data from $\boldsymbol{B}_M$ is only 10 m, while the image reconstructed using the completed dataset of $\boldsymbol{B}_N$ can also distinguish two centers over a distance as short as 7 m. After the SNR value increases to SNR = 5 dB, the distinguishable distance between the two scattering centers is reduced to 5 m in the

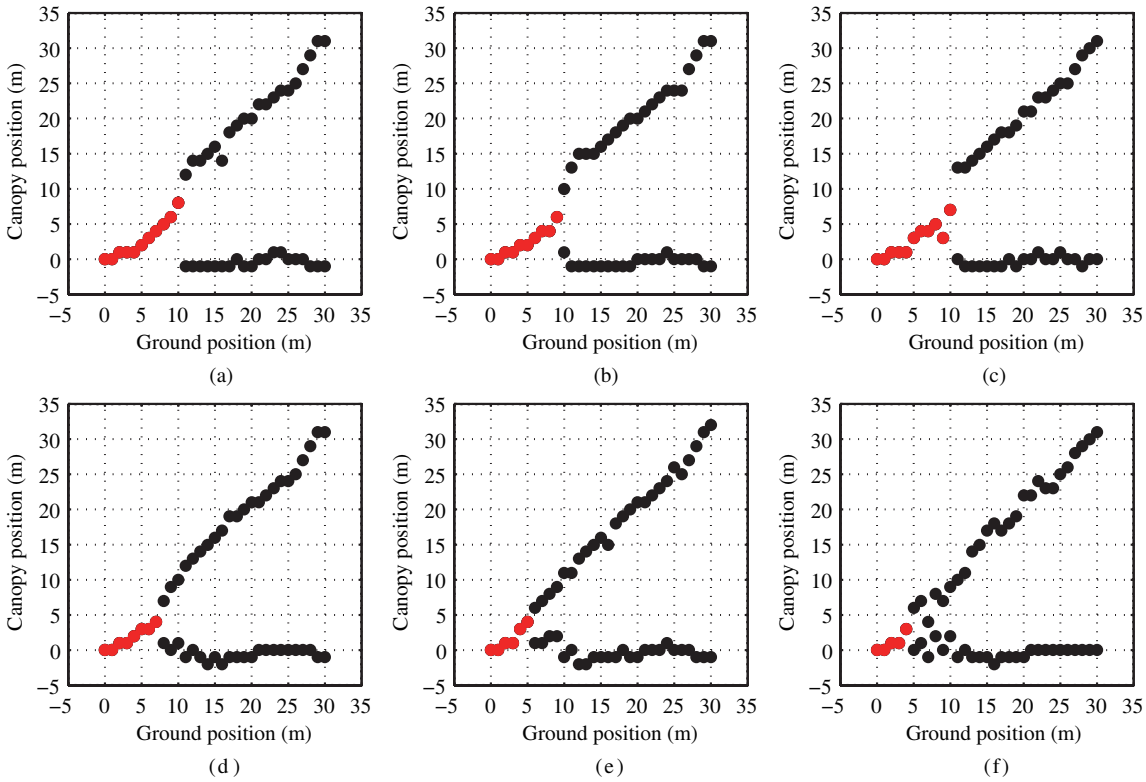


**Figure 2** Estimated elevations (in meters) of two scattering centers with identical distribution of scattering areas (representing ground and canopy) under increasing distances between two centers. The true positions are represented by a horizontal line referring to the ground and a diagonal line referring to the scattering center of the canopy at varying elevations. Estimated elevations based on the data of (Upper) observed actual baseline distribution $\boldsymbol{B}_M$, (Lower) estimated optimal virtual baseline distribution $\boldsymbol{B}_N$ with $N = 15$. From left to right column in each row: the SNRs equal 0 dB, 5 dB, and $\infty$, respectively. (Black dots: two centers detected. Red dots: one center detected.)

image reconstructed using the completed dataset of $\boldsymbol{B}_N$, which is further decreased to 4 m when the imaging procedure is not disturbed by noise. In addition, it can be seen that in both noiseless and noisy environments, after completing the dataset based on our proposed method, the super-resolving ability in the wavelet-based CS-TomoSAR algorithm's recovered images will be improved efficiently.

### 4.2 Experiment based on airborne data

The BioSAR 2008 dataset is used to further demonstrate the validity of the proposed method. The data was acquired by the E-SAR sensor of German Aerospace Centre at a boreal forested area of Northern Sweden in 2008, and includes six L-band focused two-dimensional (2D) complex image data with a slant range resolution of 1.5 m and an azimuth resolution of 1.6 m. One slice along the range direction is selected to present the experiments intuitively. Figure 3(b) shows a span of tomogram span obtained using the wavelet-based CS method (as a function of range and elevation) using a 21-by-21 window based on six actual L-band 2D focused SAR complex imagery data, and Figure 3(c) shows the completed data of ten optimal virtual baseline distribution estimated by our proposed method. For comparison, we also depict the elevation reconstructed result of conventional beamforming algorithm based on six actual L-band 2D focused SAR complex imagery data in Figure 3(a). From Figure 3, we find that compared to the beamforming algorithm's recovered image, the wavelet-based CS method can suppress elevation ambiguity and sidelobes efficiently no matter which dataset is used, and hence improve the image performance effectively. Compared with the result recovered using only the six actual SAR complex imagery data (shown in Figure 3(b)), it can be seen that the completed data can successfully recover the scattering intensity, especially in weak intensity areas, e.g., Area 1 (4400–4500 m) and Area 2 (4700–4800 m). In Area 1, we can see that the image based on the six observed imagery data can no longer distinguish the

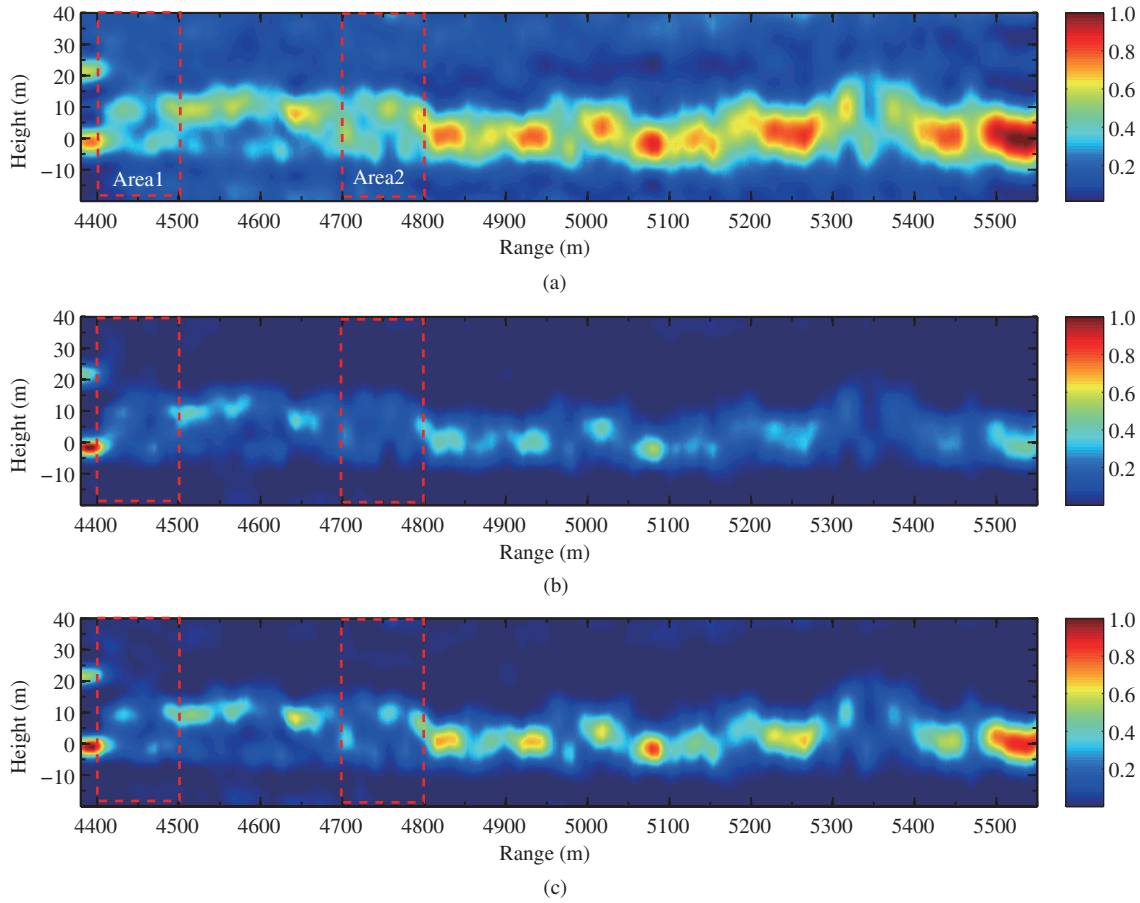


**Figure 3** Tomogram span as a function of range and elevation using a 21-by-21 window based on the real airborne dataset. Recovered results based on (a) actual six L-band 2D focused SAR complex image data using the conventional beamforing method, (b) actual six L-band 2D focused SAR complex image data using the wavelet-based CS method, and (c) the completed dataset of ten estimated optimal virtual baseline distribution using the wavelet-based CS method.

ground and canopy areas effectively because of limitations in the dataset, while the estimated optimal virtual baselines' data can successfully recover the targets in this area, and hence identify ground and canopy accurately.

## 5 Conclusion

In this paper, a novel coherence of measurement matrix-based baseline distribution optimization criterion is proposed, along with an $L_1$ regularization missing data completion technique for unobserved baselines (not belonging to the actual baseline distribution), to facilitate wavelet-based CS-TomoSAR in forested areas. Compared to using only the data of actual baselines, the reconstructed elevation image obtained via the dataset estimated by our proposed method shows higher quality, e.g., higher recovery accuracy, better super-resolving ability. Thus, the proposed method can estimate the optimal baseline distribution and reconstruct the data of unobserved baselines efficiently in wavelet-based CS-TomoSAR tailored to forested areas, which is very helpful for TomoSAR inversion based on fewer baselines. Experimental results using simulated and real airborne data have validated the proposed method.

In addition, it should be noted that our proposed coherence-based baseline distribution optimization criterion and missing data completion technique are also appropriate for other sparsity based-TomoSAR imaging methods. These methods can also achieve image performance improvements as above wavelet-based $L_1$ regularization CS-TomoSAR imaging technique based on the completed dataset estimated by our proposed method compared with results obtained using only data from actual baselines.

**Acknowledgements** This work was supported by Chinese Academy of Sciences/State Administration of Foreign Experts Affairs International Partnership Program Creative Research Team and National Natural Science Foundation of China (Grant No. 61571419). The authors would like to thank Dragon 3 Project (ID10609) and Prof. Erxue Chen and Prof. Stefano Tebaldini for providing the Biomass dataset.

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1 Reigber A, Moreira A. First demonstration of airborne SAR tomography using multibaseline L-band data. IEEE Trans Geosci Remote Sens, 2000, 38: 2142–2152
2 Fornaro G, Serafino F, Lombardini F. Three-dimensional multipass SAR focusing: experiments with long-term spaceborne data. IEEE Trans Geosci Remote Sens, 2005, 43: 702–714
3 Zhu X, Bamler R. Tomographic SAR inversion by $L_1$-norm regularization-the compressive sensing approach. IEEE Trans Geosci Remote Sens, 2010, 48: 3839–3846
4 Bi H, Zhang B C, Hong W. Matrix completion-based distributed compressive sensing for polarimetric SAR tomography. Sci China Inf Sci, 2015, 58: 119301
5 Nannini M, Scheiber R, Moreira A. Estimation of the minimum number of tracks for SAR tomography. IEEE Trans Geosci Remote Sens, 2009, 47: 531–543
6 Bi H, Zhang B, Hong W. $L_q$ regularization-based unobserved baselines' data estimation method for tomographic synthetica aperture radar inversion. J Appl Remote Sens, 2016, 10: 035014
7 Donoho D. Compressed sensing. IEEE Trans Inf Theory, 2006, 52: 1289–1306
8 Candès E, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Commun Pure Appl Math, 2006, 59: 1207–1223
9 Nyquist H. Certain topics in telegraph transmission theory. Trans Am Inst Electr Eng, 1928, 47: 617–644
10 Shannon C. Communication in the presence of noise. Proc Inst Radio Eng, 1949, 37: 10–21
11 Zhu X, Bamler R. Very high resolution SAR tomography via compressive sensing. In: Proceedings of Fringe 2009, Frascati, 2009. 1–7
12 Budillon A, Evangelista A, Schirinzi G. SAR tomography from sparse samples. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Cape Town, 2009. 865–868
13 Aguilera E, Nannini M, Reigber A. Wavelet-based compressed sensing for SAR tomography of forested areas. IEEE Trans Geosci Remote Sens, 2013, 51: 5283–5295
14 Candès E, Tao T. Near-optimal signal recovery from random projections: universal encoding strategies. IEEE Trans Inf Theory, 2006, 52: 5406–5425
15 Tropp J. Greed is good: alogrithmic results for sparse approximation. IEEE Trans Inf Theory, 2004, 50: 2231–2242
16 Bi H, Jiang C, Zhang B, et al. Track distribution optimization for tomographic synthetic aperture radar imaging. J Sys Eng Electron, 2015, 37: 1787–1792
17 Granville V, Krivanek M, Rasson P. Simulated annealing: a proof of convergence. IEEE Trans Pattern Anal Mach Intell, 1994, 16: 652–656