Unambiguous sparse recovery of migrating targets with a robustified Bayesian model

Bidon, Stéphanie; Lasserre, Marie; Le Chevalier, François

DOI
10.1109/TAES.2018.2848360

Publication date
2019

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Aerospace and Electronic Systems

Unambiguous Sparse Recovery of Migrating Targets with a Robustified Bayesian Model

Stéphanie Bidon, *Member, IEEE*, Marie Lasserre and François Le Chevalier

Abstract

The problem considered is that of estimating unambiguously migrating targets observed with a wideband radar. We extend a previously described sparse Bayesian algorithm to the presence of diffuse clutter and off-grid targets. A hybrid-Gibbs sampler is formulated to jointly estimate the sparse target amplitude vector, the grid mismatch and the (assumed) autoregressive noise. Results on synthetic and fully experimental data show that targets can be actually unambiguously estimated even if located in blind speeds.

Index Terms

Wideband radar, high range resolution, velocity ambiguities, range migration, Bayesian sparse recovery, Monte-Carlo Markov chain, Metropolis-adjusted Langevin algorithm.

I. Introduction

The primary functions of a radar system are often divided into two main categories: moving target detection (MTD) and radar imaging [1]. While the latter usually uses wideband waveform to obtain high resolution images (e.g., 1 GHz), MTD radars scan a search volume with a narrowband waveform (e.g., 20 MHz). Detection schemes are then based on the assumption that moving targets remain in their low range resolution cell during the coherent processing interval (CPI). A current limitation of such detection radars is the inherent presence of range and/or velocity ambiguities in the measurements. They not only prevent from unambiguously estimating target's features but also result in so-called blind zones where clutter folding makes detection impossible [2]. The unambiguous range and velocity are defined respectively as

$$R_a = \frac{cT_r}{2} \quad \text{and} \quad v_a = \frac{\lambda_c F_r}{2}$$

where c is the speed of light, $F_r = 1/T_r$ is the pulse repetition frequency (PRF), λ_c is the carrier wavelength. The resulting unambiguous range-velocity coverage $R_a v_a = c\lambda_c/4$ is independent of the PRF. Consequently, changing

This paper has been presented in part at the IEEE Radar Conferences, Cincinnati, OH, May 19–23, 2014 and Seattle, WA, May 8–12, 2017.

Stéphanie Bidon and Marie Lasserre are with the University of Toulouse, ISAE-Supaéro, Toulouse, France (firstname.lastname@isae-supaero.fr). The work of Marie Lasserre is supported by DGA/MRIS under grant 2014.60.0045.

François Le Chevalier is with the Delft University of Technology, The Netherlands. (e-mail: f.lechevalier@tudelft.nl).

the PRF to decrease one type of ambiguity inevitably results in an increase of the alternate. If one concentrates on the case of a low PRF mode (as done in this work), there are no range ambiguities but many velocity ambiguities and thus blind speeds.

A common approach to resolve ambiguities and enable detection in blind zones is to use a series of bursts that cycles through several carefully selected PRFs. Since the clear zone depends on the PRF, one may expect to detect a target at least for some of the PRFs [3], [4]. Nonetheless, this classical solution entails some drawbacks including ghosting (when the number of PRFs are less than the number of targets), a decrease of the time-on-target per PRF, etc. [4].

Another possible solution to obtain an unambiguous radar mode is to use a single low PRF wideband waveform for MTD radars [5]. The idea behind this approach is simple: benefit from the range walk of moving targets, significant then for high range resolution (HRR) radar, to resolve velocity ambiguity and enable detection in blind speeds. As a matter of fact, range migration constitutes an unambiguous measurement of the radial velocity unlike Doppler phase measurement. Conventional detection algorithms are not designed to handle range migration so that techniques have been implemented to compensate the phenomenon [6]–[8]. The interpolation method known as Keystone transform is certainly one of the most popular [8]. However, these algorithms do not exploit migration to either alleviate velocity ambiguity or detect targets in blind speeds. Alternatively, "wideband" algorithms have especially been developed for that purpose but are experiencing limitations [9]. A primary challenge is actually to distinguish the main peak response from its high velocity sidelobes specific to the wideband ambiguity function.

To achieve this, a sparse Bayesian algorithm has recently been proposed in [10]. Indeed, a sparse signal recovery (SSR) algorithm seems particularly relevant to this problem since it has the potential to represent a scatterer by a single peak deprived of sidelobes. Despite the encouraging results in [10], the algorithm is not designed to support either diffuse clutter or targets straddling range-velocity bins and may thus fail in practice.

In the literature, handling colored noise in a sparse recovery framework has been addressed mainly into two different ways. In the first approach, a two-step processing is recommended where data are firstly prewhitened so that a conventional SSR technique can then be applied in white noise scenario [11]–[13]. The first stage implies that the noise covariance matrix can be estimated from reliable secondary data and/or prior knowledge. In the second approach, a joint estimation of the sparse signal-of-interest and colored noise is favored [14]–[16].

Unlike the presence of colored noise in SSR, the problem of straddling targets (better known as the grid mismatch problem) has been widely covered in the SSR literature. Several strategies of robustification have been suggested including i) a local refinement of the grid [17]; ii) the joint estimation of the grid mismatch and the target amplitude vector [18]–[22]; iii) the development of grid-free based techniques [23].

In this work, we propose an extended version of the hierarchical Bayes model [10] that is robustified to both diffuse clutter and off-grid targets. This study is actually based on the preliminary results presented in [24], [25] while deepening the analysis (merging of both robustifications, details of calculation, more efficient sampling strategy, refined numerical analysis including new performance metrics and real data). More specifically, a primary-data-only approach is favored to avoid the presence of heterogeneous samples (likely to happen with an HRR radar).

Nonetheless, to regularize the noise estimation, an AR noise model is assumed [26], [27]; a common approach in radar [28]. Additionally, the grid mismatch is modeled as a perturbation that parametrizes nonlinearly the sparsifying dictionary. Both robustification leads to an original Bayesian model allowing for the joint estimation of the sparse target amplitude vector, the grid mismatch and the AR noise parameters. The resulting algorithm can then be reasonably applied on fully experimental data to unambiguously recover targets even when located in the blind speeds (at least to a certain extent).

The remaining of the paper is organized as follows. Section II introduces the observation model within an SSR framework. Sections III and IV describe the hierarchical model and its estimation procedure. Numerical simulations are provided in Sections V and VI on synthetic and fully experimental data while the last Section includes some concluding remarks.

II. Observation model in an SSR framework

In this Section we extend the observation model of [10] to robustify it towards off-grid targets as well as diffuse clutter.

A. Radar system and received signal

We consider a radar system transmitting a series of M pulses at a PRF F_r with carrier frequency F_c and bandwidth B. A low PRF is chosen meaning that radar ambiguities occur only in velocity. Additionally, a wideband waveform is assumed so that fast moving scatterers are not confined in a single range resolution cell but migrate during the CPI. To coherently sum the scatterer response and preserve its peak gain, the cell under test is thought of as a low range-resolution (LRR) segment of K range gates allowing for range walk. Applying a range transform to the LRR segment (a discrete Fourier transform), the signature of a single point scatterer can be expressed in the fast-frequency/slow-time domain as, for $k = 0, \ldots, K-1$ and $m = 0, \ldots, M-1$, [9]

$$[\boldsymbol{a}]_{m+kM} = \exp\left\{ j\ 2\pi[-f_r k + f_d m + \underbrace{f_d \mu k m}_{\text{migration term}}\ ] \right\} \tag{1}$$

where $\mu \triangleq B/(KF_c)$ is the fractional bandwidth per subband; f_r and f_d are the conventional (normalized) range and Doppler frequencies. They are defined by

$$f_r = \tau_0 \frac{B}{K} \quad \text{and} \quad f_d = \frac{v}{v_a}$$

where τ_0 and v are the initial round-trip delay and the supposedly constant radial velocity of the scatterer. Compared to a narrowband waveform, the scatterer response to a wideband waveform (1) entails cross-coupling terms that model range migration.

In what follows, the whole signal received on the LRR segment is thought of as the sum of point scatterers plus noise, i.e.,

$$\boldsymbol{y} = \sum_{\ell} \alpha_\ell \boldsymbol{a}_\ell + \boldsymbol{n} \tag{2}$$

where

$\boldsymbol{y}$	is the KM-length observation vector;
$\alpha_\ell, \boldsymbol{a}_\ell$	are the complex amplitude and signature of the ℓth scatterer;
$\boldsymbol{n}$	is the noise matrix modeling the receiver's internal noise and possibly diffuse clutter.

B. Likelihood function in an SSR framework

Our goal is to unambiguously recover the target scene represented by the scatterers even that in blind speeds. For that purpose, a sparse representation is favored as previously explained in Section I. The signal model (2) is then recast as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{3}$$

where $\boldsymbol{H}$ is the sparsifying dictionary that stems from a discretization of the range and velocity axes and $\boldsymbol{x}$ is the target amplitude vector to be estimated[1]. In practice the domain of reconstruction is selected as

$$[0, K) \times \left[-\frac{n_{va}}{2} v_a, \frac{n_{va}}{2} v_a\right) \tag{4}$$

with n_{va} the unfolding factor in velocity. The latter can be set by the radar operator to ensure that the range of all possible velocities is covered. The grid of analysis is then obtained by discretizing the domain (4) in $\bar{K} = n^r_{\text{zp}} K$ and $\bar{M} = n^v_{\text{zp}} n_{va} M$ equally spaced points respectively, n^r_{zp} and n^v_{zp} being oversampling factors (see Fig. 1). The dictionary $\boldsymbol{H}$ is thus of size $KM \times \bar{K}\bar{M}$ and its $\bar{\imath}$th colon is, at this modeling stage, defined as

$$[\boldsymbol{h}_{\bar{\imath}}]_i = \frac{1}{\sqrt{KM}} \exp\left(-j2\pi[\frac{\bar{k}}{\bar{K}}\, k + n_{va} \frac{\bar{\bar{m}}}{\bar{M}}\, m(1+\mu k)]\right) \tag{5}$$

where $i = m + kM$ and $\bar{\imath} = \bar{m} + \bar{k}\bar{M}$ with $\bar{k} = 0, \ldots, \bar{K}-1$ and $\bar{m} = 0, \ldots, \bar{M}-1$. In (5), it is worth noticing that the velocity index $\bar{\bar{m}}$ corresponds to a "signed version" of $\bar{m}$ to properly account for the range walk of receding targets. Typically, if $\bar{M}$ is even, one has $\bar{\bar{m}} = \bar{m}$ if $\bar{m} < \bar{M}/2$ otherwise $\bar{\bar{m}} = \bar{m} - \bar{M}$. The dictionary (5) is different from a simple 2D-Fourier matrix since it entails cross-coupling terms in conjunction with an unfolding factor allowing for velocity ambiguity removal.

1) Mismatch modeling: To robustify the model of [10] to off-grid scatterers, we parametrize the dictionary $\boldsymbol{H}$ by two perturbation vectors, denoted $\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r$, each modeling the mismatch in velocity and range respectively. Both are $\bar{K}\bar{M}$-length vectors and their $\bar{\imath}$th elements $\varepsilon^v_{\bar{\imath}}, \varepsilon^r_{\bar{\imath}}$ represent the grid mismatch of the $\bar{\imath}$th range-velocity bin. The $\bar{\imath}$th column (5) is consequently redefined as

$$\left[\boldsymbol{h}_{\bar{\imath}}\left(\varepsilon^v_{\bar{\imath}}, \varepsilon^r_{\bar{\imath}}\right)\right]_i = \frac{1}{\sqrt{KM}} \exp\left(-j2\pi[\frac{\bar{k} + \varepsilon^r_{\bar{\imath}}}{\bar{K}}\, k + n_{va} \frac{\bar{\bar{m}} + \varepsilon^v_{\bar{\imath}}}{\bar{M}}\, m(1+\mu k)]\right) \tag{6}$$

with $\varepsilon^v_{\bar{\imath}}, \varepsilon^r_{\bar{\imath}} \in [-.5, .5] \times [-.5, .5]$ (see Fig. 1).

[Fig. 1 about here.]

[1]In [10] the dictionary was designed from a compressed sensing point of view; it actually led to an approximated version of (5).

2) Noise modeling: To pursue our model robustification, we further take into account the presence of a diffuse clutter component that might arise in practice. In search of simplicity and to keep the computational complexity as low as possible, the following hypotheses are made. The noise $\boldsymbol{n}$ is assumed to be Gaussian distributed and represent simultaneously thermal noise and diffuse clutter. Since the latter usually corresponds to low velocity components, its range walk is neglected in our model, leading to a decoupled signature in range and velocity. Furthermore, the noise is assumed to be independent and identically distributed (iid) over range. Merging these hypotheses leads to

$$\boldsymbol{n}|\boldsymbol{R} \sim \mathcal{CN}_{KM}\left(\boldsymbol{0}, \boldsymbol{R}\right) \quad \text{with} \quad \boldsymbol{R} = \boldsymbol{I}_K \otimes \boldsymbol{\Gamma} \tag{7}$$

where $\mathcal{CN}\left(,\right)$ is the complex Gaussian distribution and $\otimes$ it the Kronecker product. $\boldsymbol{R}$ denotes the noise covariance matrix in range and velocity while $\boldsymbol{\Gamma}$ is the noise covariance matrix in velocity only. To model the corresponding correlation in slow-time, an autoregressive (AR) model of finite order $P < M$ is chosen so that the Cholesky factorization of $\boldsymbol{\Gamma}^{-1}$ can be expressed as [27]

$$\boldsymbol{\Gamma}^{-1} = \sigma_{\text{ar}}^{-2}(\boldsymbol{I}_M - \boldsymbol{\Phi})^H(\boldsymbol{I}_M - \boldsymbol{\Phi}) \tag{8}$$

where σ_{ar}^2 is the variance of the white input to the AR model and $\boldsymbol{\Phi}$ is a lower triangular Toeplitz matrix with zero diagonal elements

$$\boldsymbol{\Phi} = \text{Toeplitz}\left\{\left[0, \boldsymbol{\phi}^T, 0, \ldots, 0\right]\right\}$$

with $\boldsymbol{\phi} = \left[\phi_1, \ldots, \phi_P\right]^T$ the P-length vector containing the AR parameters. An equivalent useful notation is

$$\boldsymbol{\Phi} = \sum_{p=1}^{P} \phi_p \boldsymbol{L}_p \tag{9}$$

where $\boldsymbol{L}_p$ is the pth lower shift matrix. Note that σ_{ar}^2 represents the thermal noise power only in the case $P = 0$. In this work, P is assumed to be known.

3) Likelihood function: Using (6)-(7)-(8), the likelihood function of the observation vector (3) can be defined as

$$f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r, \sigma_{\text{ar}}^2, \boldsymbol{\phi}) = (\pi\sigma_{\text{ar}}^2)^{-KM} \exp\left\{-\sigma_{\text{ar}}^{-2}\|\breve{\boldsymbol{y}} - \breve{\boldsymbol{H}}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}\|_2^2\right\} \tag{10}$$

where whitened versions of the observation vector and the dictionary have been introduced, i.e.,

$$\breve{\boldsymbol{y}} = \left[\boldsymbol{I}_K \otimes (\boldsymbol{I}_M - \boldsymbol{\Phi})\right]\boldsymbol{y} \tag{11a}$$

$$\breve{\boldsymbol{H}}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r) = \left[\boldsymbol{I}_K \otimes (\boldsymbol{I}_M - \boldsymbol{\Phi})\right]\boldsymbol{H}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r) \tag{11b}$$

and where we have used $|\boldsymbol{R}^{-1}| = \sigma_{\text{ar}}^{-2KM}$. The parameters of interest in (10) are those describing the target scene, namely $\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r$, whereas σ_{ar}^2 and $\boldsymbol{\phi}$ can be seen as nuisance parameters. Unfortunately, estimating the target scene only from the observation is an ill-posed problem since the dictionary $\boldsymbol{H}$ is not invertible. Indeed, to unfold the velocity axis, one necessarily has $n_{va} > 1$ leading to $\bar{M} > M$ (in practice $\bar{M} \gg M$). To regularize the estimation problem additional information needs to be injected in the signal model. As explained in Section I a full Bayesian approach is chosen where each unknown parameter is considered as a random variable with a given prior pdf (probability density function).

III. Hierarchical Bayesian model

Herein we describe the priors selected in our robustified algorithm. The resulting hierarchical Bayesian model is represented graphically in Fig. 2. Compared to that of [10], the model incorporates two new branches due to each robustification. In what follows, we recall priors related to old branches and detail the new ones. Priors are selected to convey meaningful *a priori* knowledge while allowing for computationally tractable estimators.

[Fig. 2 about here.]

A. Target amplitude vector

As in [10], we assume that the elements of the target amplitude vector $x_{\bar{i}}$ are iid with a Bernoulli-Student-*t* distribution, i.e.,

$$\pi(\boldsymbol{x}) = \prod_{\bar{i}} \frac{1}{2} \left\{ \delta(x_{\bar{i}}) + \frac{\beta_0/\beta_1}{\pi} \left[1 + \frac{|x_{\bar{i}}|^2}{\beta_1} \right]^{\beta_0+1} \right\} \tag{12}$$

with $\delta()$ the Dirac delta function. This prior has a two-stage hierarchical structure useful for both computational tractability and interpretation, viz

$$\pi(\boldsymbol{x}) = \int \int \pi(\boldsymbol{x}|w, \sigma_x^2) \pi(w) \pi(\sigma_x^2) dw d\sigma_x^2 \tag{13}$$

with

$$\pi(x_{\bar{i}}|w, \sigma_x^2) = (1-w)\ \delta(x_{\bar{i}}) + w\text{CN}\left(x_{\bar{i}}|0, \sigma_x^2\right) \tag{14a}$$

$$\pi(w) = \mathbb{I}_{[0,1]}(w) \tag{14b}$$

$$\pi(\sigma_x^2) = \frac{\beta_1^{\beta_0}}{\Gamma(\beta_0)} \frac{e^{-\beta_1/\sigma_x^2}}{\sigma_x^{2(\beta_0+1)}} \mathbb{I}_{\mathbb{R}^+}(\sigma_x^2) \tag{14c}$$

where $\mathbb{I}_{\mathcal{A}}(.)$ is the indicator function on the set $\mathcal{A}$, $\Gamma(.)$ is the gamma function and $\text{CN}(.|.,.)$ is the complex Gaussian pdf with given mean and variance. Hence, the Bernoulli-Gaussian prior recognized in (14a), denoted $\mathcal{B}er\mathcal{CN}\left(w, 0, \sigma_x^2\right)$, indicates that for each range-velocity bin $\bar{i}$, a scatterer is *a priori* present with a probability w and, if so, its amplitude is Gaussian distributed with power σ_x^2. The uniform distribution of w (14b) expresses our lack of knowledge about the sparsity level in the target scene. Additionally, given that the amplitude of a scatterer may significantly vary from one to another, the inverse gamma distribution of σ_x^2 (14c) enables to cover a more or less wide range of target amplitudes according to the setting of the scale and shape parameters β_0, β_1.

The mixed-type nature of (12) actually allows one to decouple the sparsity level from the target power. Without the atom at zero, the lone hyperparameter σ_x^2 would have to monitor both criteria though opposite for high power targets. As a matter of fact, mixed-type distributions [29], [30], or in a similar way spike and slab priors [31], have been successfully used in the literature to induce sparse recovery, including the Bernoulli-Gaussian distribution [32]–[34]. Interestingly, when removing the atom at zero, the prior (12) turns into that encountered in the well known sparse Bayesian learning or relevance vector machine approach [35]–[37]. Furthermore, replacing (14c) by a gamma distribution leads to the famous Laplace prior [38], [39]

B. Off-grid vectors

In this work, we adopt a simple prior model to describe the grid mismatch. The latter is assumed iid between range-velocity bins so that

$$\pi(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r|\boldsymbol{x}) = \prod_{\bar{i}} \pi(\varepsilon^v_{\bar{i}}, \varepsilon^r_{\bar{i}}|x_{\bar{i}}). \tag{15}$$

In (15), we have intentionally conditioned the prior by the target amplitude $\boldsymbol{x}$ to enable, later on, the design of estimation scheme where grid mismatch is evaluated only if a scatterer is present at a given bin. Similar approaches have been suggested in [19]–[21]. Assuming further that the grid errors in velocity and range are independent, we propose

$$\pi(\varepsilon^v_{\bar{i}}, \varepsilon^r_{\bar{i}}|x_{\bar{i}}) = \begin{cases} \delta(\varepsilon^v_{\bar{i}})\delta(\varepsilon^r_{\bar{i}}) & \text{if } x_{\bar{i}} = 0 \quad (16a) \\ \mathbb{I}_{[-.5,.5]}(\varepsilon^v_{\bar{i}})\mathbb{I}_{[-.5,.5]}(\varepsilon^r_{\bar{i}}) & \text{if } x_{\bar{i}} \neq 0 \quad (16b) \end{cases}$$

where the last equation means that the location of a scatterer within its range-velocity bin is *a priori* equally likely.

C. Noise vector

To complete the prior model, we now turn to the statistical specification of the AR noise parameters. Conjugate priors to the likelihood are favored [40]. In particular, an inverse gamma pdf is chosen for σ^2_{ar}, i.e.,

$$\pi(\sigma^2_{\text{ar}}) = \frac{\gamma_1^{\gamma_0}}{\Gamma(\gamma_0)} \frac{e^{-\gamma_1/\sigma^2_{\text{ar}}}}{(\sigma^2_{\text{ar}})^{\gamma_0+1}} \mathbb{I}_{\mathbb{R}^+}(\sigma^2_{\text{ar}}). \tag{17}$$

In [10], an identical prior was selected to describe the thermal noise power (denoted σ^2) for the same mathematical reason. Nonetheless, in the general case $P > 0$, σ^2_{ar} and σ^2 have a different physical meaning so that σ^2_{ar} requires its own tuning regarding the scale and shape parameters γ_0, γ_1. Concerning the AR vector $\boldsymbol{\phi}$, we select a complex Gaussian prior pdf with mean $\bar{\boldsymbol{m}}_\phi$ and covariance matrix $\bar{\boldsymbol{R}}_\phi$, i.e.,

$$\pi(\boldsymbol{\phi}) = \frac{1}{\pi^P |\bar{\boldsymbol{R}}_\phi|} \exp\left\{ -\left[\boldsymbol{\phi} - \bar{\boldsymbol{m}}_\phi\right]^H \bar{\boldsymbol{R}}_\phi^{-1} \left[\boldsymbol{\phi} - \bar{\boldsymbol{m}}_\phi\right] \right\}. \tag{18}$$

Beyond mathematical convenience, we stress that both priors (17) and (18) can convey relevant prior information about noise provided an appropriate setting of the hyperparameters γ_0, γ_1 and $\bar{\boldsymbol{m}}_\phi, \bar{\boldsymbol{R}}_\phi$. Accordingly, they can be made very, moderately or non-informative. For example, if no prior knowledge is available about the AR noise, one may choose flat priors, namely (17) and (18) boils down to [41]

$$\pi(\sigma^2_{\text{ar}}) \propto \frac{1}{\sigma^2_{\text{ar}}} \mathbb{I}_{\mathbb{R}^+}(\sigma^2_{\text{ar}}) \quad \text{and} \quad \pi(\boldsymbol{\phi}) \propto 1 \tag{19}$$

where $\propto$ means proportional to. In the remainder of the paper, unless otherwise stated, we use (17) and (18) to design our estimator. To switch to the non informative priors (19), it suffices to set γ_0 and γ_1 to zero and/or the precision matrix $\bar{\boldsymbol{R}}_\phi^{-1}$ to the null matrix in the following expressions.

IV. Bayesian estimation

According to the previously described hierarchical model we propose a Bayesian estimation algorithm. Its outputs are samples from which the target scene is estimated. Neither at the input nor at the output of the algorithm a number of targets is either fixed or determined.

A. Posterior model

In a Bayesian framework, the posterior pdf merges information brought by the observation and the prior model. Particularly in this work, the posterior distribution of the target scene $\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r$ is after applying Bayes theorem

$$f(\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\pi(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r|\boldsymbol{x})\pi(\boldsymbol{x}) \tag{20}$$

where

$$f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r) = \iint\limits_{\boldsymbol{\phi}, \sigma^2_{\text{ar}}} f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r, \boldsymbol{\phi}, \sigma^2_{\text{ar}})\pi(\boldsymbol{\phi})\pi(\sigma^2_{\text{ar}})d\sigma^2_{\text{ar}}d\boldsymbol{\phi}. \tag{21}$$

The latter pdf is derived in Appendix A and leads to a posterior that is, to our knowledge, too complicated to analytically obtain standard Bayesian estimators. Alternatively, we use a numerical iterative approach that can sample intricate multivariate pdfs, especially that arising from hierarchical model. The method is computationally intensive; nonetheless with a sufficient number of samples, not only Bayesian estimators can be empirically obtained but also posterior pdfs which are more informative than a single point estimate. In what follows, we use the following notations: $\boldsymbol{\theta} = \left(\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r, w, \sigma^2_x, \sigma^2_{\text{ar}}, \boldsymbol{\phi}\right)$ is the random variable of all unknown variables, $\boldsymbol{\theta}_{-\boldsymbol{\zeta}}$ is the random variable $\boldsymbol{\theta}$ deprived of $\boldsymbol{\zeta}$ (e.g., $\boldsymbol{\theta}_{-\boldsymbol{x}}$), and $\boldsymbol{\theta}^{(t)} = \left(\boldsymbol{x}^{(t)}, \boldsymbol{\varepsilon}^{v\,(t)}, \boldsymbol{\varepsilon}^{r\,(t)}, w^{(t)}, \sigma_x^{2\,(t)}, \sigma^{2\,(t)}_{\text{ar}}, \boldsymbol{\phi}^{(t)}\right)$ are the samples simulated at the t-th iteration.

B. Principle of estimation

The sampling procedure implemented is a Monte Carlo Markov Chain (MCMC) method resulting here in a so called hybrid Gibbs sampler [40, p.387]. Each unknown random variable $\boldsymbol{\zeta}$ in $\boldsymbol{\theta}$ is iteratively sampled according to its full conditional, namely $f(\boldsymbol{\zeta}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\zeta}})$. After discarding some first N_{bi} samples representing the burn in period, $\boldsymbol{\theta}^{(t)}$ is distributed according to the joint posterior $f(\boldsymbol{\theta}|\boldsymbol{y})$ and each sample $\boldsymbol{\zeta}^{(t)}$ is distributed according to its posterior $f(\boldsymbol{\zeta}|\boldsymbol{y})$. Hence, with enough samples (denoted N_r in total), Bayesian estimators can be obtained empirically. In this extended work, we concentrate on minimum mean square error (MMSE) estimation since it already led to satisfying target scene recovery in a simpler context [10]. The posterior mean $\mathcal{E}\{\boldsymbol{\zeta}|\boldsymbol{y}\} \triangleq \int \boldsymbol{\zeta} f(\boldsymbol{\zeta}|\boldsymbol{y})d\boldsymbol{\zeta}$ is practically assessed via

$$\hat{\boldsymbol{\zeta}}_{\text{mmse}} \triangleq \frac{1}{N_r}\sum_{t=1}^{N_r} \boldsymbol{\zeta}^{(t+N_{bi})}. \tag{22}$$

Not only the MMSE estimators of the parameters of interest $\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r$ can be obtained but also that of the nuisance parameters $\sigma^2_{\text{ar}}, \boldsymbol{\phi}$.

At this stage, each full conditional $f(\boldsymbol{\zeta}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\zeta}})$ remains to be determined. They are directly obtained by considering the joint posterior pdf

$$f(\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r, \sigma^2_{\text{ar}}, \boldsymbol{\phi}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r, \sigma^2_{\text{ar}}, \boldsymbol{\phi})\pi(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r|\boldsymbol{x})\pi(\boldsymbol{x}|w, \sigma^2_x)\pi(w)\pi(\sigma^2_x)\pi(\boldsymbol{\phi})\pi(\sigma^2_{\text{ar}}) \tag{23}$$

while fixing all but one parameter as discussed hereafter.

C. Sampling of w, σ_x^2, $\boldsymbol{x}$, and σ_{ar}^2

In the non-robustified approach [10], the full conditionals of w, σ_x^2, $\boldsymbol{x}$, and σ_{ar}^2 (denoted σ^2 with $P = 0$) were explicitly derived. With our augmented hierarchical model, it can straightforwardly be shown that the full conditionals keep the same functional forms as follows

$$w|\boldsymbol{y}, \boldsymbol{\theta}_{-w} \sim \mathcal{B}e\left(1 + \|\boldsymbol{x}\|_0, 1 + \bar{K}\bar{M} - \|\boldsymbol{x}\|_0\right) \tag{24a}$$

$$\sigma_x^2|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{x}} \sim \mathcal{IG}\left(\|\boldsymbol{x}\|_0 + \beta_0, \|\boldsymbol{x}\|_2^2 + \beta_1\right) \tag{24b}$$

$$x_{\bar{i}}|\boldsymbol{y}, \boldsymbol{\theta}_{-x_{\bar{i}}} \sim \mathcal{B}er\mathcal{CN}\left(\breve{w}_{\bar{i}}, \breve{\mu}_{\bar{i}}, \breve{\eta}_{\bar{i}}^2\right) \tag{24c}$$

$$\sigma_{\text{ar}}^2|\boldsymbol{y}, \boldsymbol{\theta}_{-\sigma_{\text{ar}}^2} \sim \mathcal{IG}\left(KM + \gamma_0, \left\|\breve{\boldsymbol{y}} - \breve{\boldsymbol{H}}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}\right\|_2^2 + \beta_1\right) \tag{24d}$$

where $\|.\|_0$ is the number of nonzero elements in a vector and the parameters in (24c) are

$$\begin{cases} \breve{w}_{\bar{i}} = \left[\dfrac{1-w}{w}\dfrac{\sigma_x^2}{\breve{\eta}_{\bar{i}}^2}\exp\left\{-\dfrac{|\breve{\mu}_{\bar{i}}|^2}{\breve{\eta}_{\bar{i}}^2}\right\} + 1\right]^{-1} \\ \breve{\eta}_{\bar{i}}^2 = \left\{\dfrac{1}{\sigma_x^2} + \dfrac{\|\breve{\boldsymbol{h}}_{\bar{i}}\|_2^2}{\sigma_{\text{ar}}^2}\right\}^{-1} \\ \breve{\mu}_{\bar{i}} = \dfrac{\breve{\eta}_{\bar{i}}^2}{\sigma_{\text{ar}}^2}\breve{\boldsymbol{h}}_{\bar{i}}^H\breve{\boldsymbol{e}}_{\bar{i}} \end{cases}$$

with $\breve{\boldsymbol{h}}_{\bar{i}}$ denoting the $\bar{i}$th colon of $\breve{\boldsymbol{H}}$ and

$$\breve{\boldsymbol{e}}_{\bar{i}} = \breve{\boldsymbol{y}} - \sum_{\bar{j}\neq\bar{i}} x_{\bar{j}}\breve{\boldsymbol{h}}_{\bar{j}}(\varepsilon_{\bar{j}}^v, \varepsilon_{\bar{j}}^r). \tag{25}$$

Interestingly enough, the parameters involved in (24c)-(24d) depends now on the whitened observation $\breve{\boldsymbol{y}}$ and the whitened-and-robustified dictionary $\breve{\boldsymbol{H}}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)$ of (11). The distributions in (24) are well known and easy to sample. We refer the interested reader to [10] for further details and rather focus on the technical novelty, namely the sampling of $\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r$ and $\boldsymbol{\phi}$.

D. Off-grid sampling

1) Full conditional: Similarly to the target amplitude vector $\boldsymbol{x}$, the conditional posterior pdf of the grid errors $\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r$ has not a known functional form (not explicited here) and seems hardly sampleable. As a result, we sample the grid errors on a range-velocity bin basis. Using (23), the conditional posterior of the grid errors at the $\bar{i}$th bin is

$$f(\varepsilon_{\bar{i}}^v, \varepsilon_{\bar{i}}^r|\boldsymbol{y}, \boldsymbol{\theta}_{-(\varepsilon_{\bar{i}}^v, \varepsilon_{\bar{i}}^r)}) \propto f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r, \sigma_{\text{ar}}^2, \boldsymbol{\phi})\pi(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r|\boldsymbol{x}).$$

Injecting the definition (25) of $\breve{\boldsymbol{e}}_{\bar{i}}$ in the likelihood (10) and using the off-grid prior (15)-(16), the joint posterior conditional of the grid mismatches becomes

$$f(\varepsilon_{\bar{i}}^v, \varepsilon_{\bar{i}}^r|\boldsymbol{y}, \boldsymbol{\theta}_{-(\varepsilon_{\bar{i}}^v, \varepsilon_{\bar{i}}^r)}; x_{\bar{i}} = 0) = \delta(\varepsilon_{\bar{i}}^v)\delta(\varepsilon_{\bar{i}}^r) \tag{26a}$$

$$f(\varepsilon_{\bar{i}}^v, \varepsilon_{\bar{i}}^r|\boldsymbol{y}, \boldsymbol{\theta}_{-(\varepsilon_{\bar{i}}^v, \varepsilon_{\bar{i}}^r)}; x_{\bar{i}} \neq 0) \propto \exp\left(-\sigma_{\text{ar}}^{-2}\|\breve{\boldsymbol{e}}_{\bar{i}} - x_{\bar{i}}\breve{\boldsymbol{h}}_{\bar{i}}(\varepsilon_{\bar{i}}^v, \varepsilon_{\bar{i}}^r)\|_2^2\right)\mathbb{I}_{[-.5,.5]}(\varepsilon_{\bar{i}}^v)\mathbb{I}_{[-.5,.5]}(\varepsilon_{\bar{i}}^r). \tag{26b}$$

To lighten the notations, the target distribution (26b) is also denoted later by

$$\tau(\boldsymbol{\varepsilon}_{\bar{i}}) \triangleq f(\varepsilon^v_{\bar{i}}, \varepsilon^r_{\bar{i}} | \boldsymbol{y}, \boldsymbol{\theta}_{-(\varepsilon^v_{\bar{i}}, \varepsilon^r_{\bar{i}})}; x_{\bar{i}} \neq 0)$$

with $\boldsymbol{\varepsilon}_{\bar{i}} \triangleq \begin{bmatrix} \varepsilon^v_{\bar{i}} & \varepsilon^r_{\bar{i}} \end{bmatrix}^T$. The pdf is represented in Fig. 3 in case of low and high signal-to-interference-plus-noise-ratios (SINR); see definition in (32). If a strong scatterer is present at the $\bar{i}$th bin, the target distribution (26b) becomes highly informative and thus peaked. As a result, since the latter does not belong to a known class of distribution, a sampling strategy remains to be determined.

Different methods can be thought of in practice: 1) ad-hoc techniques; 2) commonly used Bayesian methods [40]; 3) or more advanced strategies [45]. The former approach has been tested in [43]. The sampling scheme relies on deriving (26b) on a uniform grid that needs to become extremely fine for strong scatterer thereby rendering the method computationally inefficient if not unrealistic. Alternatively, a very common Bayesian method is the Metropolis-Hastings (MH) algorithm which is based on a two-step principle: propose and accept-reject a sample [40, Ch.7]. It has been tested in [42] in a simplified scenario where there is no-range migration and only white noise. The proposed sample is drawn from a Gaussian distribution resulting from an approximation of the target distribution. Unfortunately, the approximated pdf can undesirably depart from the target distribution if $|\varepsilon^v_{\bar{i}} - \varepsilon^r_{\bar{i}}| \to 1$. This leads to an inefficient sampling scheme where the proposed sample may be almost systematically rejected. Within the class of MH algorithms, an adequate proposal distribution is often obtained within a so-called Gaussian random walk [40, ch7.5]. The proposed sample is then drawn from a local Gaussian perturbation around the former accepted sample. Unfortunately, it still leads with (26b) to a slow sampler convergence. In light of these limitations, we opt in this work for the last strategy namely an advanced sampling technique as reviewed in [45]. We particularly choose an advanced MH approach known as Metropolis-Adjusted Langevin algorithm (MALA) [44]. So far, the MALA approach gave us the most satisfying results; though we do not claim any optimality here.

[Fig. 3 about here.]

2) MALA sampling: An MH algorithm is an iterative MCMC method that aims at sampling a target distribution τ. Here, it consists in drawing at the $\tilde{t}$-th iteration a candidate $\boldsymbol{c}$ from a so called proposal distribution $q(\boldsymbol{c}|\boldsymbol{\varepsilon}^{(\tilde{t}-1)})$ and accept this candidate as the new sample $\boldsymbol{\varepsilon}^{(\tilde{t})}$ with a given acceptance probability [40, ch.7]

$$\rho = \min\left(1, \frac{\tau(\boldsymbol{c})}{\tau(\boldsymbol{\varepsilon}^{(\tilde{t}-1)})} \frac{q(\boldsymbol{\varepsilon}^{(\tilde{t}-1)}|\boldsymbol{c})}{q(\boldsymbol{c}|\boldsymbol{\varepsilon}^{(\tilde{t}-1)})}\right).$$

In case of MALA, the proposal distribution is chosen according to a local quadratic approximation to $\log \tau$ around $\boldsymbol{\varepsilon}^{(\tilde{t}-1)}$, i.e., $q(\boldsymbol{c}|\boldsymbol{\varepsilon}^{(\tilde{t}-1)})$ such as

$$q(\boldsymbol{c}|\boldsymbol{\varepsilon}) = \mathrm{N}\left(\boldsymbol{c}|\boldsymbol{\varepsilon} + \frac{1}{2}\boldsymbol{\Sigma}_\varepsilon(\boldsymbol{\varepsilon})\boldsymbol{d}_\varepsilon(\boldsymbol{\varepsilon}), \boldsymbol{\Sigma}_\varepsilon(\boldsymbol{\varepsilon})\right) \tag{27}$$

where $\mathrm{N}(.|.,.)$ is the Gaussian pdf, $\boldsymbol{d}_\varepsilon$ is the gradient of $\log \tau$ and $\boldsymbol{\Sigma}_\varepsilon$ is a positive definite matrix. As underlined in [45], the matrix $\boldsymbol{\Sigma}_\varepsilon$ ought to account for the correlation structure of τ. A self-evident choice for $\boldsymbol{\Sigma}_\varepsilon$ is the inverse of the Hessian matrix. Nonetheless, given the form of (26b), its positiveness is not always ensured. To bypass this

problem, we have followed the path of [46] by selecting $\boldsymbol{\Sigma}_\varepsilon$ as an inverse Fisher information matrix (FIM); other strategies are reviewed in [45]. Expressions of $\boldsymbol{d}_\varepsilon$ and $\boldsymbol{\Sigma}_\varepsilon$ are given in Appendix B. The proposed MH sampling scheme is summarized in Fig. 5 and incorporated in the main algorithm of Fig. 4. It is worth noticing that, to ensure convergence of the so obtained hybrid Gibbs sampler, only one MH occurrence suffices [40, ch.10], i.e., $N_{MH} = 1$ in Fig. 5.

E. AR parameters sampling

It remains to sample $\boldsymbol{\phi}$ according to its full conditional. Using (23), the latter can be expressed as

$$\begin{aligned} f(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\phi}}) &\propto f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r, \sigma^2_{\text{ar}}, \boldsymbol{\phi})\pi(\boldsymbol{\phi}) \\ &\propto \exp\left\{-\sigma^{-2}_{\text{ar}} \| \left[\boldsymbol{I}_K \otimes (\boldsymbol{I}_M - \boldsymbol{\Phi})\right] \boldsymbol{\nu} \|_2^2\right\} \exp\left\{ - \left[\boldsymbol{\phi} - \bar{\boldsymbol{m}}_\phi\right]^H \bar{\boldsymbol{R}}_\phi^{-1} \left[\boldsymbol{\phi} - \bar{\boldsymbol{m}}_\phi\right]\right\} \end{aligned} \tag{28}$$

where we have set $\boldsymbol{\nu} \triangleq \boldsymbol{y} - \boldsymbol{H}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}$. The pdf (28) actually represents a complex Gaussian distribution. To see it, it suffices to develop the squared norm in (28) as

$$\begin{aligned} \| \left[\boldsymbol{I}_K \otimes (\boldsymbol{I}_M - \boldsymbol{\Phi})\right] \boldsymbol{\nu} \|_2^2 &= \sum_{k=0}^{K-1} \| (\boldsymbol{I}_M - \boldsymbol{\Phi}) \boldsymbol{\nu}_k \|_2^2 \\ &= \sum_{k=0}^{K-1} \boldsymbol{\nu}_k^H \left[\boldsymbol{I}_M + \boldsymbol{\Phi}^H \boldsymbol{\Phi} - \Re\{\boldsymbol{\Phi}\}\right] \boldsymbol{\nu}_k \end{aligned}$$

where $\boldsymbol{\nu}_k$ is the M-length-vector such that

$$\begin{bmatrix} \boldsymbol{\nu}_0^T & \dots & \boldsymbol{\nu}_{K-1}^T \end{bmatrix}^T \triangleq \boldsymbol{y} - \boldsymbol{H}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}.$$

Then replacing $\boldsymbol{\Phi}$ by its expression (9), one gets

$$\| \left[\boldsymbol{I}_K \otimes (\boldsymbol{I}_M - \boldsymbol{\Phi})\right] \boldsymbol{\nu} \|_2^2 = \|\boldsymbol{\nu}\|^2 + \boldsymbol{\phi}^H \boldsymbol{P}_\phi \boldsymbol{\phi} - 2\Re\left\{\boldsymbol{\phi}^H \boldsymbol{m}_\phi\right\} \tag{29}$$

where $\boldsymbol{P}_\phi$ is a $P \times P$ matrix and $\boldsymbol{m}_\phi$ is a P-length vector defined as, for $p, p' \in \{1, \dots, P\}$,

$$\begin{aligned} [\boldsymbol{P}_\phi]_{p-1,p'-1} &= \sum_{k=0}^{K-1} \boldsymbol{\nu}_k^H \boldsymbol{L}_p^H \boldsymbol{L}_{p'} \boldsymbol{\nu}_k \\ [\boldsymbol{m}_\phi]_{p-1} &= \sum_{k=0}^{K-1} \boldsymbol{\nu}_k^H \boldsymbol{L}_p \boldsymbol{\nu}_k. \end{aligned} \tag{30}$$

Finally, injecting (29) in (28) yields

$$\begin{aligned} f(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\phi}}) &\propto \exp\left\{-\sigma^{-2}_{\text{ar}} \left[\boldsymbol{\phi}^H \boldsymbol{P}_\phi \boldsymbol{\phi} - 2\Re\left\{\boldsymbol{\phi}^H \boldsymbol{m}_\phi\right\}\right] - \boldsymbol{\phi}^H \bar{\boldsymbol{R}}_\phi^{-1} \boldsymbol{\phi} + 2\Re\left\{\boldsymbol{\phi}^H \bar{\boldsymbol{R}}_\phi^{-1} \bar{\boldsymbol{m}}_\phi\right\}\right\} \\ &\propto \exp\left\{ - \left[\boldsymbol{\phi} - \boldsymbol{\mu}_\phi\right]^H \boldsymbol{\Sigma}_\phi^{-1} \left[\boldsymbol{\phi} - \boldsymbol{\mu}_\phi\right]\right\} \end{aligned}$$

where, in the last line, we have completed the square with $\boldsymbol{\Sigma}_\phi = \left[\sigma^{-2}_{\text{ar}} \boldsymbol{P}_\phi + \bar{\boldsymbol{R}}_\phi^{-1}\right]^{-1}$ and $\boldsymbol{\mu}_\phi = \boldsymbol{\Sigma}_\phi \left[\sigma^{-2}_{\text{ar}} \boldsymbol{m}_\phi + \bar{\boldsymbol{R}}_\phi^{-1} \bar{\boldsymbol{m}}_\phi\right]$. We recognize the complex Gaussian distribution

$$\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\phi}} \sim \mathcal{CN}_P\left(\boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}_\phi\right) \tag{31}$$

which is easy to sample. This completes the description of our hybrid Gibbs sampler that is summarized in Fig. 4.

[Fig. 4 about here.]

[Fig. 5 about here.]

V. Numerical simulations

In this Section, performance of the proposed hybrid Gibbs sampler is illustrated on synthetic data. Numerical values used to run the simualtions can be found in the caption of each depcited Figure.

A. Parameter setting

Data are generated according to the model (1), (2) and (7). The AR parameters are selected to approximately mimic the disturbance estimated from experimental data (see Section VI-B). In the same spirit scatterers are also added at zero velocity. Other scatterers injected in the scene represent moving targets. The SINR of a scatterer is defined as

$$\text{SINR} = \mathcal{E}\left\{|\alpha|^2\right\} \boldsymbol{a}^H \boldsymbol{R}^{-1} \boldsymbol{a}. \tag{32}$$

We consider that no information is available about the AR components and thus select the noninformative priors (19) for σ_{ar}^2 and $\boldsymbol{\phi}$. Concerning the average target power σ_x^2, we use the results of [47] and accordingly choose a prior centered around high power values (with respect to the scatterers). Concretely, we set the mean $\text{m}_{\sigma_x^2}$ and standard deviation $\text{std}_{\sigma_x^2}$ of the prior of σ_x^2 (14c) to some finite desired values and convert them in terms of scale and shape parameters via $\beta_0 = \text{m}_{\sigma_x^2}^2/\text{std}_{\sigma_x^2}^2 + 2$ and $\beta_1 = \text{m}_{\sigma_x^2}\left[\text{m}_{\sigma_x^2}^2/\text{std}_{\sigma_x^2}^2 + 1\right]$.

To assess the benefits of our robustified approach (identified by the tag 'AROFF'), we additionally run its initial version [10] that assumes a White noise with ON-grid targets ('WON') as well as two partially robustified versions; one with respect to OFF-grid targets only ('WOFF') and the other with respect to an AR disturbance only ('ARON'). Outputs of a simple coherent integration, defined as $\boldsymbol{a}^H\boldsymbol{y}/\sqrt{\boldsymbol{a}^H\boldsymbol{a}}$, is also depicted to indicate the location of conventional sidelobes.

B. Example on a single run

Representative range-velocity maps recovered by the four algorithms are depicted in Fig. 6. Firstly, when range-velocity bin straddling occurs, estimating grid mismatch (WOFF and AROFF) prevents from i) misestimation of small discretes; ii) false estimation due to splitting of strong discretes. Note that split occurs in range-velocity bins corresponding to the columns of the (whitened) dictionary sharing a high coherence. Secondly, estimation schemes based on a white noise model (WON and WOFF) fails at describing clutter properly, namely numerous false estimations arise in blind velocities. By nature, diffuse clutter cannot be represented with a finite number of (slow) discretes. The remaining part is then absorbed by discretes sharing a high coherence in the dictionary, i.e., that located in the clutter velocity sidelobes. Thirdly, discretes located in blind velocities can also be well recovered especially when both grid mismatch and AR noise are estimated. Overall, we clearly see the advantage of our doubly robustified algorithm (AROFF) to obtain an unambiguous mode.

To reinforce this view, AR spectrums associated with the estimated AR vectors are depicted in Fig.7, i.e.,

$$S_{\text{ar}}(v) \triangleq \sigma^2_{\text{ar}} \left| 1 - \sum_{p=1}^{P} \phi_p e^{-j2\pi p \frac{v}{v_a}} \right|^{-2} .$$

AROFF spectrum is very close to that of the true AR spectrum enabling a near-optimal noise whitening within the algorithm. On the other hand, ARON spectrum may falsely identify spectral components located at the velocities of range-off-grid targets whose mismatch cannot be estimated. These peaks may appear aliased since our model assumes a slow moving disturbance (cf. Section II-B2). In any event, the associated ARON whitening filter will entail undesired notches at these velocities.

Finally, empirical pdfs estimated from the hybrid Gibbs sampler (AROFF) are depicted in Fig. 8. Even in case of flat priors, posteriors are highly peaked indicating that the measurement significantly informs the model. Additionally, MMSE estimates are close to the true parameter values.

[Fig. 6 about here.]

[Fig. 7 about here.]

[Fig. 8 about here.]

C. Monte-Carlo runs

Monte-Carlo simulations are conducted to confirm trends observed on single outputs. A Swerling 0 point target is simulated amid AR noise. Three main zones in the range-velocity domain are then defined as illustrated in Fig. 1: the target's bin ($\mathcal{T}$), the blind speeds ($\mathcal{B}$) and the white noise zone ($\mathcal{W}$). The power of each reconstructed zone after whitening is chosen as a performance metrics, namely

$$P_{\mathcal{Z}} \triangleq \mathcal{E} \left\{ \| \hat{\boldsymbol{R}}^{-1/2} \sum_{\bar{j} \in \mathcal{Z}} \boldsymbol{h}_{\bar{j}}(\widehat{\varepsilon^v_{\bar{j}}}, \widehat{\varepsilon^r_{\bar{j}}}) \hat{x}_{\bar{j}} \|^2 \right\}$$

where $\mathcal{Z} \in \{\mathcal{T}, \mathcal{B}, \mathcal{W}\}$ is the index set of one of the three considered zone and the notation $\hat{\zeta}$ refers to the parameter ζ either estimated or assumed by the algorithm (e.g., for WON and ARON $\widehat{\varepsilon^v_{\bar{j}}} = \widehat{\varepsilon^r_{\bar{j}}} = 0$).

In the first numerical example the target is placed on the grid in the first blind speed; performance of the WON and ARON algorithms is then illustrated in Fig. 9. The estimated power $P_{\mathcal{B}}$ confirms that significant false estimations arise in the blind speeds when the diffuse component is ignored[2]. Alternatively, the estimated power $P_{\mathcal{T}}$ shows that a strong enough target is well recovered (i.e., for SINR $\gtrapprox$ 14 dB) provided that diffuse clutter is estimated otherwise the latter contributes to overestimating the target. As for the estimated power $P_{\mathcal{W}}$, it shows that false estimation in the white noise zone might be very rare whether or not diffuse clutter is estimated.

In the second numerical example the target is still in the blind speed but placed off the grid; performance of the ARON and AROFF is then illustrated in Fig 10. Ignoring a strong off-grid leads to tremendous loss. Hence, both examples illustrate the benefit of the proposed robustification.

[2]For the ARON technique, $P_{\mathcal{B}}$ is very low and would appear smoother with a very large number of Monte-Carlo runs.

[Fig. 9 about here.]

[Fig. 10 about here.]

VI. Results on fully experimental data

Herein, we process *fully experimental radar data* to show, in practice, the feasibility of estimating unambiguously targets using range migration even that located in blind velocities.

A. Experimental setup

Data are collected with the Software-Defined Radio (SDR) radar 2400AD2 from ©Ancortek Inc. A linear frequency modulated continuous waveform (LFMCW) is selected on transmit with a 2 GHz bandwidth operating at 25 GHz with a PRF of 250 Hz. The radar delivers IQ samples at baseband after a mixing operation. A conventional inverse-range-transform is then applied and a LRR segment of interest is selected to apply our wideband algorithms. The radar scene consists of 1) an empty corridor with plants located in the LRR 2) a remote control car endorsing the role of target (its length is ca. 15 cm). Additionally a fan hidden from the radar can be activated to simulate a windy vegetation. Data sets and videos of the measurement can be downloaded from URL [48].

B. Diffuse and correlated clutter components

To begin with, target-free datasets are used. The mean vector and covariance matrix of the observation $\boldsymbol{y}$ are estimated via a simple empirical mean

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_t \quad \text{and} \quad \hat{\boldsymbol{R}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{y}_t \boldsymbol{y}_t^H - \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^H$$

with T the length of the training interval and $\boldsymbol{y}_t$ the measurement collected in the LRR from sweeps $tM, \ldots, (t+1)M-1$. To represent them in a comprehensible manner the two following metrics are chosen

$$\hat{A} \triangleq |\boldsymbol{a}^H \hat{\boldsymbol{\mu}}|^2 \quad \text{and} \quad \hat{L} \triangleq \boldsymbol{a}^H \hat{\boldsymbol{R}}^{-1} \boldsymbol{a} / (\boldsymbol{a}^H \boldsymbol{a})$$

which can be viewed as an adapted pattern and a SINR-loss respectively. Results are depicted in Fig. 11. Turning the fan on actually produces a diffuse clutter component otherwise the environment can be approximately assimilated to a white noise with clutter discretes. Furthermore, the experimental SINR-loss obtained in case of strong diffuse clutter indicates that estimating weak target in the first blind speed may remain challenging (though SINR-loss in blind velocities can be reduced by increasing the CPI).

[Fig. 11 about here.]

C. Target estimation

A dataset containing the target is now considered when the latter is about to exit the first blind speed $-v_a$ at range-gate ca. 61. Range-velocity maps and AR spectrums obtained are represented in Figs. 12-13. Trends observed on synthetic data are confirmed: robustifing the initial algorithm of [10] enables estimation of a target in the blind speed with fully experimental data. (Outputs of consecutive bursts can be also seen in a movie at the URL [48] where the target recedes within the blind speed).

[Fig. 12 about here.]

[Fig. 13 about here.]

VII. Conclusion and perspectives

We have described an extended version of a sparse Bayesian algorithm that unambiguously recovers migrating targets from real wideband radar data. Two robustifications have been brought to the hierarchical model to handle both diffuse clutter (in absence of secondary data) and off-grid targets. The disturbance is modeled as an AR noise while grid mismatch is defined as a nonlinear perturbation in the sparsifying dictionary. The resulting estimation algorithm is an hybrid Gibbs sampler that generates samples according to the posterior distributions of both the nuisance and parameters of interest (i.e., target amplitude vector, mismatch, AR noise). Performance of the MMSE estimators is assessed on synthetic and fully experimental data. Modeling diffuse clutter reduces greatly false estimation in blind speeds. Modeling grid mismatch reduces significantly target splits (thus false or missed estimation). Our robustification actually enabled unambiguous estimation of a target located in a blind speed from real data. As in conventional processing, a sufficiently high SINR remains necessary to ensure an appropriate recovery.

Extensions to this work include i) developing alternative algorithms to decrease the complexity of our sampler ii) refining the target and noise models as observed from some experimental datasets iii) incorporating the hierarchical model into a detection scheme.

Appendix A
Posterior pdf

In this Appendix, we develop the expression of the posterior (20) of $\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r | \boldsymbol{y}$. To that end we derive the pdf (21) of $\boldsymbol{y} | \boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r$

$$
\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r) &= \int_{\boldsymbol{\phi}} \int_{\sigma^2_{\text{ar}}} f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r, \sigma^2_{\text{ar}}, \boldsymbol{\phi}) \pi(\sigma^2_{\text{ar}}) \pi(\boldsymbol{\phi}) d\sigma^2_{\text{ar}} d\boldsymbol{\phi} \\
&\propto \int_{\boldsymbol{\phi}} \pi(\boldsymbol{\phi}) \int_{\sigma^2_{\text{ar}}} \frac{e^{-\left[\|\breve{\boldsymbol{y}} - \breve{\boldsymbol{H}}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}\|_2^2 + \gamma_1\right]/\sigma^2_{\text{ar}}}}{\sigma^{2\,(KM+\gamma_0+1)}_{\text{ar}}} \mathbb{I}_{\mathbb{R}^+}(\sigma^2_{\text{ar}}) d\sigma^2_{\text{ar}} d\boldsymbol{\phi} \\
&\propto \int_{\boldsymbol{\phi}} \frac{\pi(\boldsymbol{\phi})}{\|\breve{\boldsymbol{y}} - \breve{\boldsymbol{H}}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}\|_2^2 + \gamma_1} d\boldsymbol{\phi}.
\end{aligned}
$$

In the special case of a noninformative prior for $\boldsymbol{\phi}$, as defined in (19), the pdf can be further simplified observing that its denominator becomes

$$\begin{aligned} d(\boldsymbol{\phi}) &\triangleq \gamma_1 + \|\breve{\boldsymbol{y}} - \breve{\boldsymbol{H}}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}\|_2^2 \\ &\underset{(29)}{=} \gamma_1 + \|\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}\|_2^2 + \boldsymbol{\phi}^H \boldsymbol{P}_\phi \boldsymbol{\phi} - 2\Re\left\{\boldsymbol{\phi}^H \boldsymbol{m}_\phi\right\} \\ &= \gamma_1 + \|\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}\|_2^2 - \boldsymbol{m}_\phi^H \boldsymbol{P}_\phi^{-1} \boldsymbol{m}_\phi + \left[\boldsymbol{\phi} - \boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi\right]^H \boldsymbol{P}_\phi^{-1}\left[\boldsymbol{\phi} - \boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi\right] \\ &= d(\boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi) + \left[\boldsymbol{\phi} - \boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi\right]^H \boldsymbol{P}_\phi^{-1}\left[\boldsymbol{\phi} - \boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi\right] \end{aligned}$$

where $\boldsymbol{P}_\phi$ and $\boldsymbol{m}_\phi$ are defined in (30). Hence, we obtain

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r) &\propto \int_{\boldsymbol{\phi}} \left[d(\boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi) + \left[\boldsymbol{\phi} - \boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi\right]^H \boldsymbol{P}_\phi^{-1}\left[\boldsymbol{\phi} - \boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi\right]\right]^{P+KM-P+\gamma_0} d\boldsymbol{\phi} \\ &\propto |\boldsymbol{P}_{\boldsymbol{\phi}}|^{-1} d(\boldsymbol{P}_\phi^{-1}\boldsymbol{m}_\phi)^{-(KM-P+\gamma_0)} \end{aligned} \tag{33}$$

where we have used the definition of the complex multivariate *t* distribution with shape parameter ν, location vector $\boldsymbol{\mu}$, and scale matrix $\boldsymbol{\Sigma}$ [50], i.e.,

$$\mathrm{T}(\boldsymbol{\phi}|\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\Gamma(P+\nu)}{|\pi\boldsymbol{\Sigma}|\Gamma(\nu)\nu^P}\left\{1 + [\boldsymbol{\phi} - \boldsymbol{\mu}]^H \boldsymbol{\Sigma}^{-1}[\boldsymbol{\phi} - \boldsymbol{\mu}]/\nu\right\}^{-(P+\nu)}.$$

Otherwise in case of an informative prior, deriving (21) in closed form seems intractable. Nonetheless, a closed-form could be similarly obtained as in (33) while redesigning the AR vector prior as $\pi(\sigma_{\text{ar}}^2|\boldsymbol{\phi}) = \mathrm{CN}_P\left(\boldsymbol{\phi}|\bar{\boldsymbol{m}}_\phi, \sigma_{\text{ar}}^2\bar{\boldsymbol{R}}_\phi\right)$.

APPENDIX B
GRADIENT AND FISHER INFORMATION MATRIX IN PROPOSED MALA

In this Appendix, we derive the analytical expressions of $\boldsymbol{d}_\varepsilon$ and $\boldsymbol{\Sigma}_\varepsilon$ used in the MALA proposal distribution (27). To that end, we firstly notice that the target distribution (26b) can be interpreted as the posterior distribution of $\boldsymbol{\varepsilon}_{\bar{i}}|\boldsymbol{e}_{\bar{i}}$ with respect to the "observation" $\boldsymbol{e}_{\bar{i}}$ and the prior (16b) while considering σ_{ar}^2 and $x_{\bar{i}}$ known, viz

$$\tau(\boldsymbol{\varepsilon}_{\bar{i}}) \propto \mathrm{CN}\left(\boldsymbol{e}_{\bar{i}}|x_{\bar{i}}\boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}}), \boldsymbol{R}\right)\pi(\boldsymbol{\varepsilon}_{\bar{i}}|x_{\bar{i}})$$

where $\boldsymbol{R}$ has been defined in (7)-(8) and where we have redefined $\boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}}) \triangleq \boldsymbol{h}_{\bar{i}}(\varepsilon_{\bar{i}}^v, \varepsilon_{\bar{i}}^r)$. Hence, noting that the prior of $\boldsymbol{\varepsilon}_{\bar{i}}|x_{\bar{i}}$ is uniform and following [46], $\boldsymbol{d}_\varepsilon$ and $\boldsymbol{\Sigma}_\varepsilon$ represent respectively the gradient of the log-likelihood and the FIM inverse associated with the observation model $\boldsymbol{e}_{\bar{i}}|\boldsymbol{\varepsilon}_{\bar{i}} \sim \mathcal{CN}\left(x_{\bar{i}}\boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}}), \boldsymbol{R}\right)$. Since $\boldsymbol{\varepsilon}_{\bar{i}}$ is real-valued, we can directly apply results from [51, App.15C], i.e.,

$$\begin{aligned} \boldsymbol{d}_\varepsilon(\boldsymbol{\varepsilon}_{\bar{i}}) &= 2\Re\left\{x_{\bar{i}}^* \left[\frac{\partial \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})}{\partial \varepsilon_{\bar{i}}^v} \quad \frac{\partial \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})}{\partial \varepsilon_{\bar{i}}^r}\right]^H \boldsymbol{R}^{-1}\left[\boldsymbol{e}_{\bar{i}} - x_{\bar{i}}\boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})\right]\right\} \\ \boldsymbol{\Sigma}_\varepsilon^{-1}(\boldsymbol{\varepsilon}_{\bar{i}}) &= 2|x_{\bar{i}}|^2\Re\left\{\left[\frac{\partial \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})}{\partial \varepsilon_{\bar{i}}^v} \quad \frac{\partial \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})}{\partial \varepsilon_{\bar{i}}^r}\right]^H \boldsymbol{R}^{-1}\left[\frac{\partial \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})}{\partial \varepsilon_{\bar{i}}^v} \quad \frac{\partial \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})}{\partial \varepsilon_{\bar{i}}^r}\right]\right\} \end{aligned}$$

where, using the definition of $\boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})$ in (6), it follows

$$\frac{\partial \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})}{\partial \varepsilon_{\bar{i}}^v} = j\boldsymbol{\delta}^v \odot \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}}) \quad \text{and} \quad \frac{\partial \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})}{\partial \varepsilon_{\bar{i}}^r} = -j\boldsymbol{\delta}^r \odot \boldsymbol{h}_{\bar{i}}(\boldsymbol{\varepsilon}_{\bar{i}})$$

with $[\boldsymbol{\delta}^v]_{m+kM} = 2\pi n_{va}(1+\mu k)m/\bar{M}$ and $[\boldsymbol{\delta}^r]_{m+kM} = 2\pi k/\bar{K}$. It is worth noticing that to avoid the sampling chain to be stuck in a local minima, we restart sampling randomly in $[-.5,.5]\times[-.5,.5]$ as soon as the Hessian matrix at $\boldsymbol{\varepsilon}_{\bar{i}}$ is nonpositive definite.

REFERENCES

[1] W. L. Melvin and J. A. Scheer, "Overview: Advanced techniques in modern radar," in *Principle of Modern Radar. Advanced Techniques*, W. L. Melvin and J. A. Scheer, Eds. Edison, NJ: SciTech Publishing, 2013, ch. 1.

[2] M. A. Richards, "Doppler processing," in *Principle of Modern Radar. Basic Principles*, M. A. Richards, J. A. Scheer, and W. A. Holm, Eds. Edison, NJ: SciTech Publishing, 2010, ch. 17.

[3] G. Trunk and S. Brockett, "Range and velocity ambiguity resolution," in *Proc. IEEE Radar Conf. 1993*, 1993, pp. 146 – 149.

[4] A. Partizian, "Airborne pulse-doppler radar," in *Principle of Modern Radar. Radar Applications*, W. L. Melvin and J. A. Scheer, Eds. Edison, NJ: SciTech Publishing, 2014, ch. 5.

[5] F. Le Chevalier, "Radar non ambigü à large bande," French Patent 9 608 509, 1996.

[6] D. E. Iverson, "Coherent processing of ultra-wideband radar signals," *Proc. IEE Radar, Sonar Navig.*, vol. 141, no. 3, pp. 171–179, Jun. 1994.

[7] N. Jiang, R. Wu, and J. Li, "Super resolution feature extraction of moving targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 3, pp. 781–793, Jul. 2001.

[8] R. P. Perry, R. C. DiPietro, and R. L. Fante, "SAR imaging of moving targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 188–200, Jan. 1999.

[9] F. Deudon, S. Bidon, O. Besson, and J.-Y. Tourneret, "Velocity dealiased spectral estimators of range migrating targets using a single low-PRF wideband waveform," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 1, pp. 244–265, Jan. 2013, http://oatao.univ-toulouse.fr/8304/.

[10] S. Bidon, J.-Y. Tourneret, L. Savy, and F. Le Chevalier, "Bayesian sparse estimation of migrating targets for wideband radar," *IEEE Transactions on Aerospace and Electronic Systems (Special Section on Compressive Sensing Applied to Radar)*, vol. 50, no. 2, pp. 871–886, Apr. 2014.

[11] J. T. Parker, M. A. Ferrara, and L. C. Potter, "Radar applications of sparse reconstruction and compressed sensing," in *Principle of Modern Radar. Advanced Techniques*, W. L. Melvin and J. A. Scheer, Eds. Edison, NJ: SciTech Publishing, 2013, ch. 5.

[12] K. Fyhn, T. L. Jensen, T. Larsen, and S. H. Jensen, "Compressive sensing for spread spectrum receivers," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2334–2343, May 2013.

[13] M. A. Atassi and I. Abou-Faycal, "A reconstruction algorithm for noisy compressed sensing; the UWB channel estimation test case," in *2012 19th International Conference on Telecommunications (ICT)*, April 2012, pp. 1–6.

[14] J. J. Fuchs, "DOA estimation in the presence of unknown colored noise, the global matched filter approach," in *2010 18th European Signal Processing Conference*, Aug 2010, pp. 1369–1373.

[15] I. W. Selesnick, S. U. Pillai, K. Y. Li, and B. Himed, "Angle-Doppler processing using sparse regularization," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, March 2010, pp. 2750–2753.

[16] L. Ning, T. T. Georgiou, and A. Tannenbaum, "Separation of system dynamics and line spectra via sparse representation," in *49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 473–478.

[17] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[18] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, 2011.

[19] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3809–3822, Jul. 2012.

[20] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[21] L. Hu, J. Zhou, Z. Shi, and Q. Fu, "A fast and accurate reconstruction algorithm for compressed sensing of complex sinusoids," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5744–5754, Nov. 2013.

[22] Z. Tan and A. Nehorai, "Sparse direction of arrival estimation using co-prime arrays with off-grid targets," *IEEE Signal Process. Lett.*, vol. 21, no. 1, January 2014.

[23] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, Aug. 2013.

[24] S. Bidon, O. Besson, J.-Y. Tourneret, and F. Le Chevalier, "Bayesian sparse estimation of migrating targets in autoregressive noise for wideband radar," in *Proc. IEEE Int. Radar Conf. (RADARCON)*, Cincinnati, OH, May 19–23, 2014.

[25] M. Lasserre, S. Bidon, and F. Le Chevalier, "An unambiguous radar mode with a single PRF wideband waverform," in *Proceedings IEEE Radar Conference (RadarConf)*, Seattle, WA, May 8-12 2017.

[26] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Ann. Statist.*, vol. 36, no. 1, pp. 199–227, 2008.

[27] A. Wiesel, O. Bibi, and A. Globerson, "Time varying autoregressive moving average models for covariance estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2791–2801, Jun. 2013.

[28] Y. I. Abramovich, N. K. Spencer, and M. D. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1298–1311, Apr. 2007.

[29] P. Schnitter, L. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Proceedings of the Workshop on Information Theory and Applications*, La Jolla, CA, Jan. 2008.

[30] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059–2070, Sep. 2009.

[31] T. J. Mitchell and J. J. Beauchamp, "Bayesian variable selection in linear regression," *Journal of the American Statistical Association*, vol. 83, no. 404, pp. 1023–1032, 1988.

[32] J. J. Kormylo and J. M. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Transactions on Information Theory*, vol. 28, no. 3, pp. 482–488, May 1982.

[33] G. Z. Dai and J. M. Mendel, "Maximum a posteriori estimation of multichannel Bernoulli-Gaussian sequences," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 181–183, Jan. 1989.

[34] Q. Cheng, R. Chen, and T. H. Li, "Simultaneous wavelet estimation and deconvolution of reflection seismic signals," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 377–384, Mar. 1996.

[35] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[36] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, Jun. 2001.

[37] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[38] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.

[39] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian group-sparse modeling and variational inference," *IEEE Transactions on Signal Processing*, vol. 62, no. 11, pp. 2906–2921, June 2014.

[40] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY: Springer Science, 2004.

[41] S. Ni and D. Sun, "Bayesian estimates for vector autoregressive models," *Journal of Business and Economic Statistics*, vol. 23, no. 1, Jan. 2005.

[42] S. Bidon, M. Lasserre, O. Besson, and F. Le Chevalier, "Bayesian sparse estimation of targets with range-Doppler grid mismatch," in *Proc. IEEE Int. Radar Conf. (RADARCON)*, Washington, DC, May 11-15, 2015.

[43] M. Lasserre, S. Bidon, and F. Le Chevalier, "Velocity ambiguity mitigation of off-grid range migrating targets via Bayesian sparse recovery," in *Proceedings Statistical Signal Processing Workshop (SSP)*, Palma de Mallorca, Jun. 26-29 2016.

[44] G. O. Roberts and O. Stramer, "Langevin diffusions and Metropolis-Hastings algorithms," *Methodology And Computing In Applied Probability*, vol. 4, no. 4, pp. 337–357, 2002.

[45] M. Pereyra, P. Schniter, E. Chouzenoux, J. C. Pesquet, J. Y. Tourneret, A. O. Hero, and S. McLaughlin, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 224–241, March 2016.

[46] M. Girolami and B. Calderhead, “Riemann manifold Langevin and Hamiltonian Monte Carlo methods,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 2, pp. 123–214, 2011.

[47] M. Lasserre, S. Bidon, and F. L. Chevalier, “New sparse-promoting prior for the estimation of a radar scene with weak and strong targets,” *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4634–4643, Sept 2016.

[48] [Online]. Available: https://sourceforge.isae.fr/projects/ralf/wiki

[49] S. Bidon, F. Deudon, O. Krasnov, and F. Le Chevalier, “Coherent integration for wideband LFMCW applied to PARSAX experimental data,” in *Proceedings European Radar Conference (EURAD)*, Manchester, Oct. 12–14, 2011.

[50] S. Kotz and S. Nadarajah, *Multivariate t-distributions and their applications*. Cambridge University Press, 2004.

[51] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

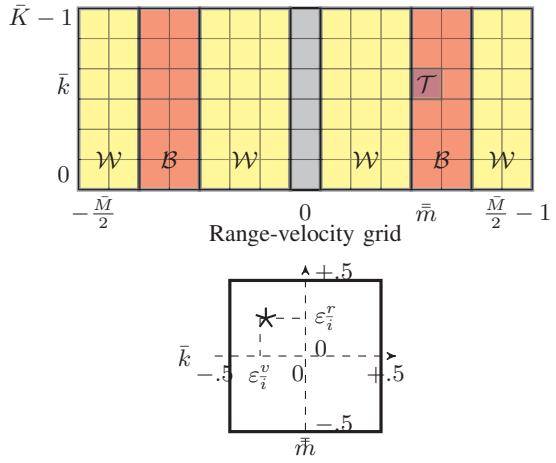


Fig. 1. Grid of analysis in the range-velocity domain and grid mismatch phenomenon. $\mathcal{T}$, $\mathcal{B}$, $\mathcal{W}$ designate the set of indices of the target, the blind zone, and the white noise zone respectively. Star marker represents the true target location.

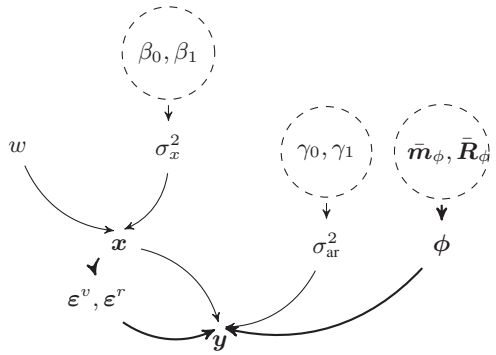
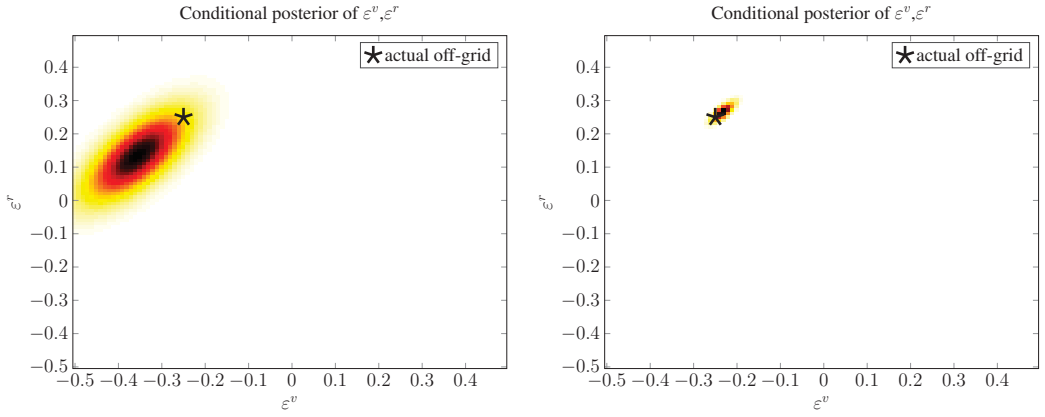


Fig. 2. Graphical representation of the proposed hierarchical Bayesian model. Arrows represent statistical dependence. Parameters in dotted circle are set by radar operator according to prior knowledge about radar scene. Compared to the model of [10], new branches are depicted in thick line.



(a) (b)

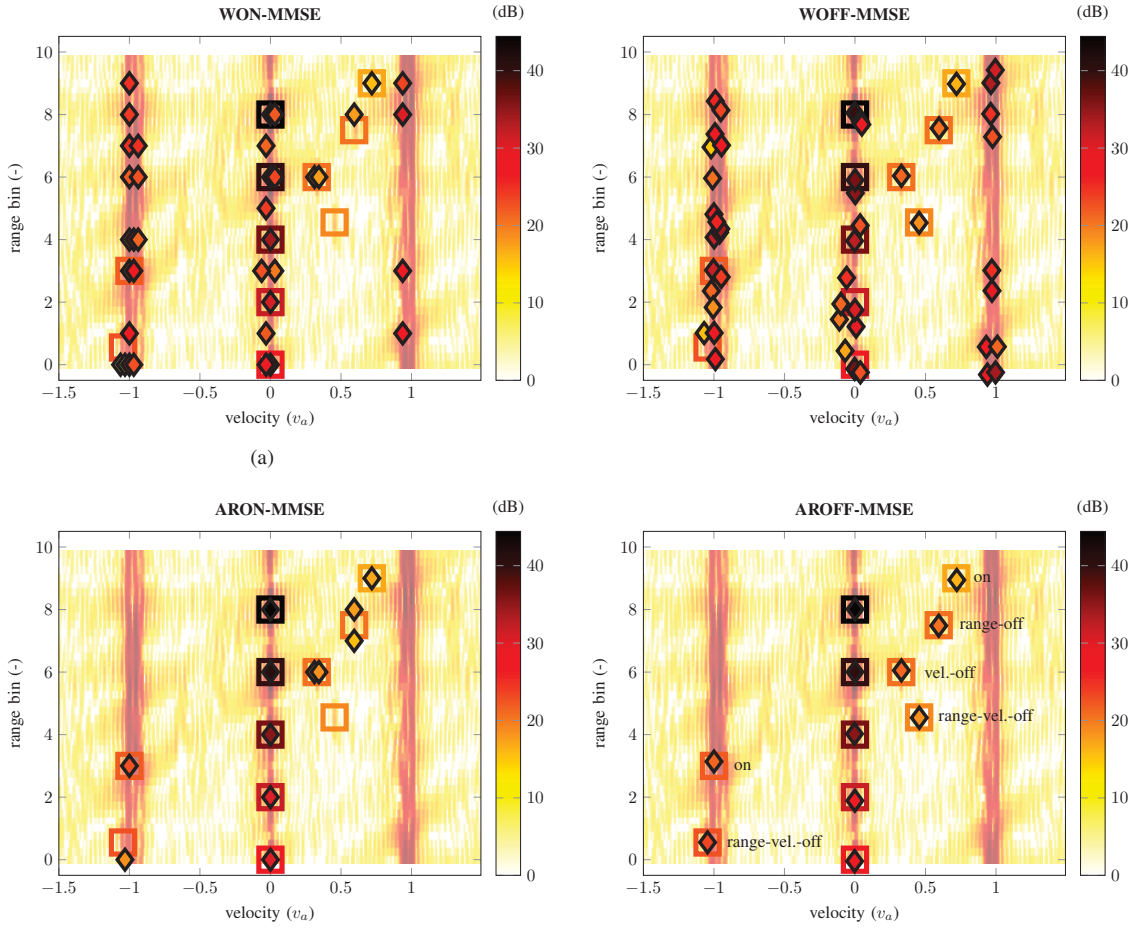
Fig. 3. Conditional posterior distribution (26b) of the grid mismatch: influence of the SINR. Scenario: $M = 16$, $K = 6$, $n_{va} = 2$, $B/F_c = 10\%$, $\sigma^2_{\text{ar}} = 2.5$, $\boldsymbol{\phi} = [-0.9e^{-j0.05}]$, single point scatterer with $\bar{i} = 40$ (i.e., $\bar{m} = 8$, $\bar{k} = 3$), $\varepsilon^v_{40} = -.25$, $\varepsilon^r_{40} = .25$, variable SINR. Processing: $n^v_{\text{zp}} = n^r_{\text{zp}} = 1$. (a) $\text{SINR}_{\bar{i}} = 10$ dB. (b) $\text{SINR}_{\bar{i}} = 25$ dB.

Require: $\boldsymbol{y}$, n_{va}, $\bar{K}, \bar{M}$, (γ_0, γ_1),(β_0, β_1), $\bar{\boldsymbol{m}}_\phi, \bar{\boldsymbol{R}}_\phi$
Ensure: $\boldsymbol{\theta}^{(t)} \sim f(\boldsymbol{\theta}|\boldsymbol{y})$, $\boldsymbol{\zeta}^{(t)} \sim f(\boldsymbol{\zeta}|\boldsymbol{y})$
{Initialization}
$\boldsymbol{x}^{(0)}$, $\boldsymbol{\varepsilon}^{v\,(0)}$, $\boldsymbol{\varepsilon}^{r\,(0)}$, $\boldsymbol{\phi}^{(0)}$
{Iterations}
for $t = 1$ to $N_{bi} + N_r$ **do**
$w^{(t)}|\boldsymbol{y}, \boldsymbol{\theta}_{-w} \sim \mathcal{B}e\left(1 + \|\boldsymbol{x}\|_0, 1 + \bar{K}\bar{M} - \|\boldsymbol{x}\|_0\right)$
$\sigma_x^{2\,(t)}|\boldsymbol{y}, \boldsymbol{\theta}_{-\sigma_x^2} \sim \mathcal{IG}\left(\|\boldsymbol{x}\|_0 + \beta_0, \|\boldsymbol{x}\|_2^2 + \beta_1\right)$
for $\bar{i} \in \{0, \ldots, \bar{K}\bar{M} - 1\}$ **do**
$x_{\bar{i}}^{(t)}|\boldsymbol{y}, \boldsymbol{\theta}_{-x_{\bar{i}}} \sim \mathcal{B}er\mathcal{CN}\left(\breve{w}_{\bar{i}}, \breve{\mu}_{\bar{i}}, \breve{\eta}_{\bar{i}}^2\right)$
$\begin{cases} \boldsymbol{\varepsilon}_{\bar{i}}^{(t)} = \mathbf{0} & \text{if } x_{\bar{i}} = 0 \\ \boldsymbol{\varepsilon}_{\bar{i}}^{(t)}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\varepsilon}_{\bar{i}}} \sim \tau(\boldsymbol{\varepsilon}_{\bar{i}}) & \text{otherwise} \end{cases}$
end for
$\sigma_{\text{ar}}^{2\,(t)}|\boldsymbol{y}, \boldsymbol{\theta}_{-\sigma_{\text{ar}}^2} \sim \mathcal{IG}\left(KM + \gamma_0, \left\|\breve{\boldsymbol{y}} - \breve{\boldsymbol{H}}(\boldsymbol{\varepsilon}^v, \boldsymbol{\varepsilon}^r)\boldsymbol{x}\right\|_2^2 + \beta_1\right)$
$\boldsymbol{\phi}^{(t)}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\phi}} \sim \mathcal{CN}_P\left(\boldsymbol{\mu}_\phi, \boldsymbol{\Sigma}_\phi\right)$
end for
{Estimators}
$\hat{\boldsymbol{\zeta}}_{\text{mmse}} = \frac{1}{N_r}\sum_{t=1}^{N_r} \boldsymbol{\zeta}^{(t+N_{bi})}$

Fig. 4. Proposed hybrid Gibbs sampler to estimate off-grid migrating targets in AR noise. Conditional terms actually depends on the most updated samples in the iteration though not explicitly written due to space limitation. Convergence of the sampler can be reached independently of the initial conditions [40]. $\boldsymbol{\zeta}$ represents any of the sampled variables.

Ensure: $\varepsilon_{\bar{i}}^{(t)}|\boldsymbol{y}, \boldsymbol{\theta}_{-\boldsymbol{\varepsilon}_{\bar{i}}} \sim \tau(\boldsymbol{\varepsilon}_{\bar{i}})$
  {Initialization}
  $\boldsymbol{\varepsilon}^{(0)} = \boldsymbol{\varepsilon}_{\bar{i}}^{(t-1)}$
  {Iterations}
  for $\tilde{t} = 1$ to N_{MH} **do**
    $\boldsymbol{c} \sim \mathcal{N}\left(\boldsymbol{\varepsilon}^{(\tilde{t}-1)} + \frac{1}{2}\boldsymbol{\Sigma}_{\varepsilon}(\boldsymbol{\varepsilon}^{(\tilde{t}-1)})\boldsymbol{d}_{\varepsilon}(\boldsymbol{\varepsilon}^{(\tilde{t}-1)}), \boldsymbol{\Sigma}_{\varepsilon}(\boldsymbol{\varepsilon}^{(\tilde{t}-1)})\right)$
    $\rho = \min\left(1, \frac{\tau(\boldsymbol{c})}{\tau(\boldsymbol{\varepsilon}^{(\tilde{t}-1)})} \frac{q(\boldsymbol{\varepsilon}^{(\tilde{t}-1)}|\boldsymbol{c})}{q(\boldsymbol{c}|\boldsymbol{\varepsilon}^{(\tilde{t}-1)})}\right)$
    $u \sim \mathcal{U}_{[0,1]}$
    if $u \leq \rho$ **then**
      $\boldsymbol{\varepsilon}^{(\tilde{t})} = \boldsymbol{c}$
    else
      $\boldsymbol{\varepsilon}^{(\tilde{t})} = \boldsymbol{\varepsilon}^{(\tilde{t}-1)}$
    end if
  end for

Fig. 5. Proposed MALA algorithm to sample the conditional posterior of the grid errors at the $\bar{i}$th bin $\tau(\boldsymbol{\varepsilon}_{\bar{i}})$. N_{MH} is the number of iterations of the algorithm. When MALA algorithm incorporated in hybrid Gibbs sampler of Fig. 4 then $N_{MH} = 1$.



(a)

Fig. 6. Range-velocity map (modulus of the complex amplitude only). Data parameters: $F_c = 10$ GHz, $B = 1$ GHz, $T_r = 1$ ms, $M = 32$, $K = 10$ ($v_a = 15$ m/s and $c/(2B) = 15$ cm), $\sigma^2_{\text{ar}} = 1.7$, $\boldsymbol{\phi} = [0.5, 0.3, 0.25]^T$. Processing parameters: $n_{va} = 3$, $n^v_{\text{zp}} = n^r_{\text{zp}} = 1$, β_0, β_1 such that $\text{m}_{\sigma^2_x} = 45$ dB and $\text{std}_{\sigma^2_x} \approx 15$ dB, $\gamma_0 = \gamma_1 = 0$, $\bar{\boldsymbol{R}}^{-1}_{\phi} = \mathbf{0}$, $P = 3$, $N_{bi} = 1E^{+3}$ and $N_r = 200$. Square markers indicate true scatterer location. Diamond markers indicate estimated scatterers. Coherent integration depicted as a transparent background.

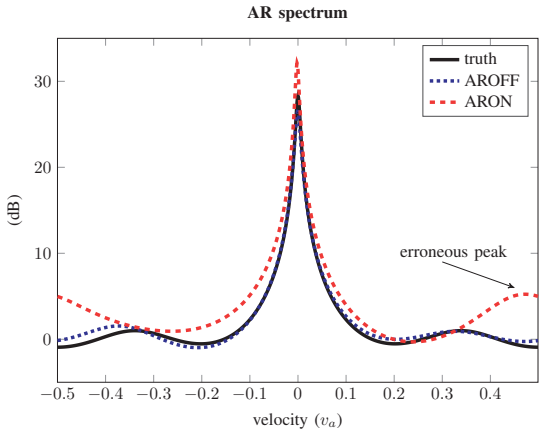


Fig. 7. AR spectrum. Scenario is that of Fig. 6.

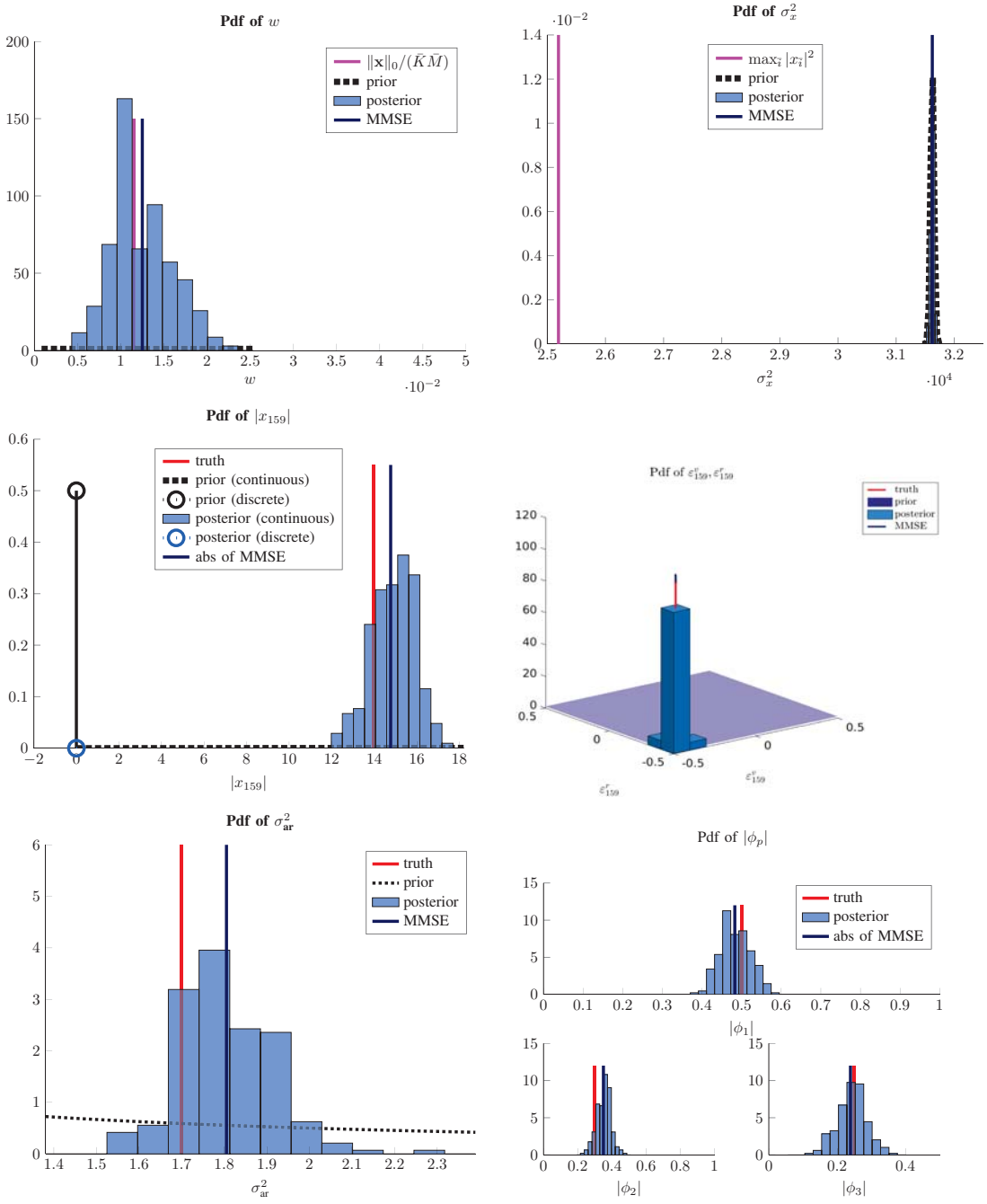


Fig. 8. Empirical posterior distributions estimated with the hybrid Gibbs sampler (AROFF). Scenario is that of Fig. 6. Prior of absolute value $|x_{\bar{i}}|$ can be found in [47, Eq. (5)]. $\bar{i} = 159$ is the index of the closest range-velocity-off-grid target in Fig. 6.

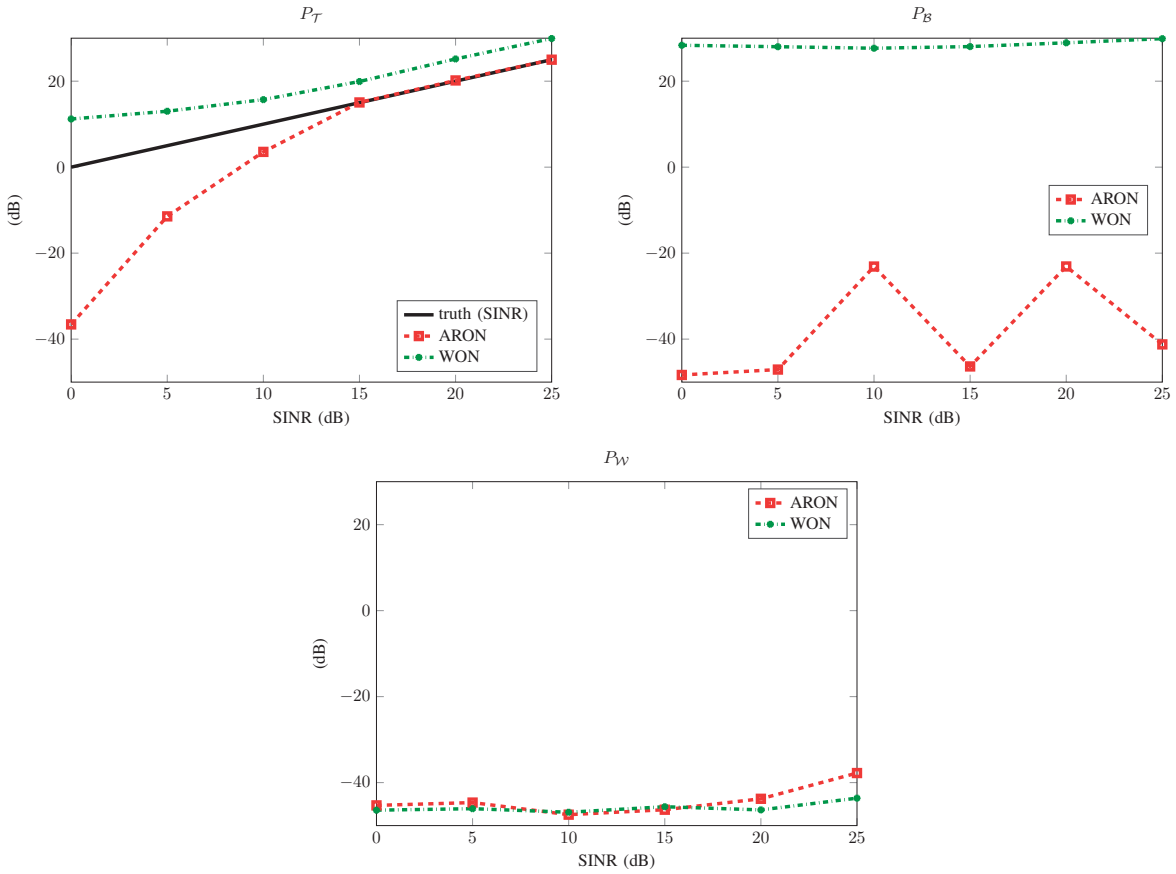


Fig. 9. Power of the reconstructed scene after whitening. Scenario is that of Figure 6 except: $K = 6$, a single point scatterer in first blind speed with varying SINR and $\bar{\imath} = 318$ (i.e., $\bar{m} = 30$, $\bar{k} = 3$), $\varepsilon^v_{318} = \varepsilon^r_{318} = 0$). 500 Monte-Carlo runs.

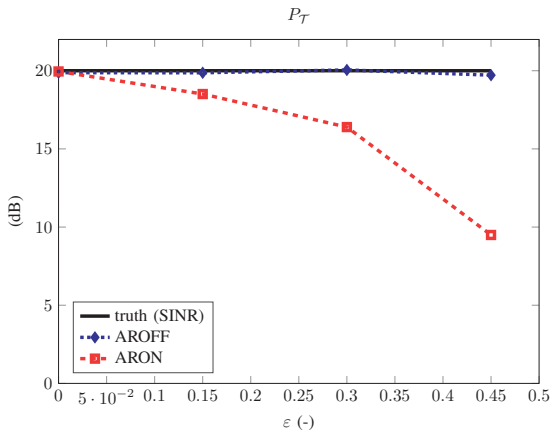


Fig. 10. Power of the reconstructed scene after whitening. Scenario is that of Fig. 9 except single point scatterer with varying off-grid $\varepsilon^v_{318} = \varepsilon^r_{318} = \varepsilon$ and fixed SINR=20 dB. $\mathcal{T}$ includes the nearest range-velocity bins to that of the scatterer.

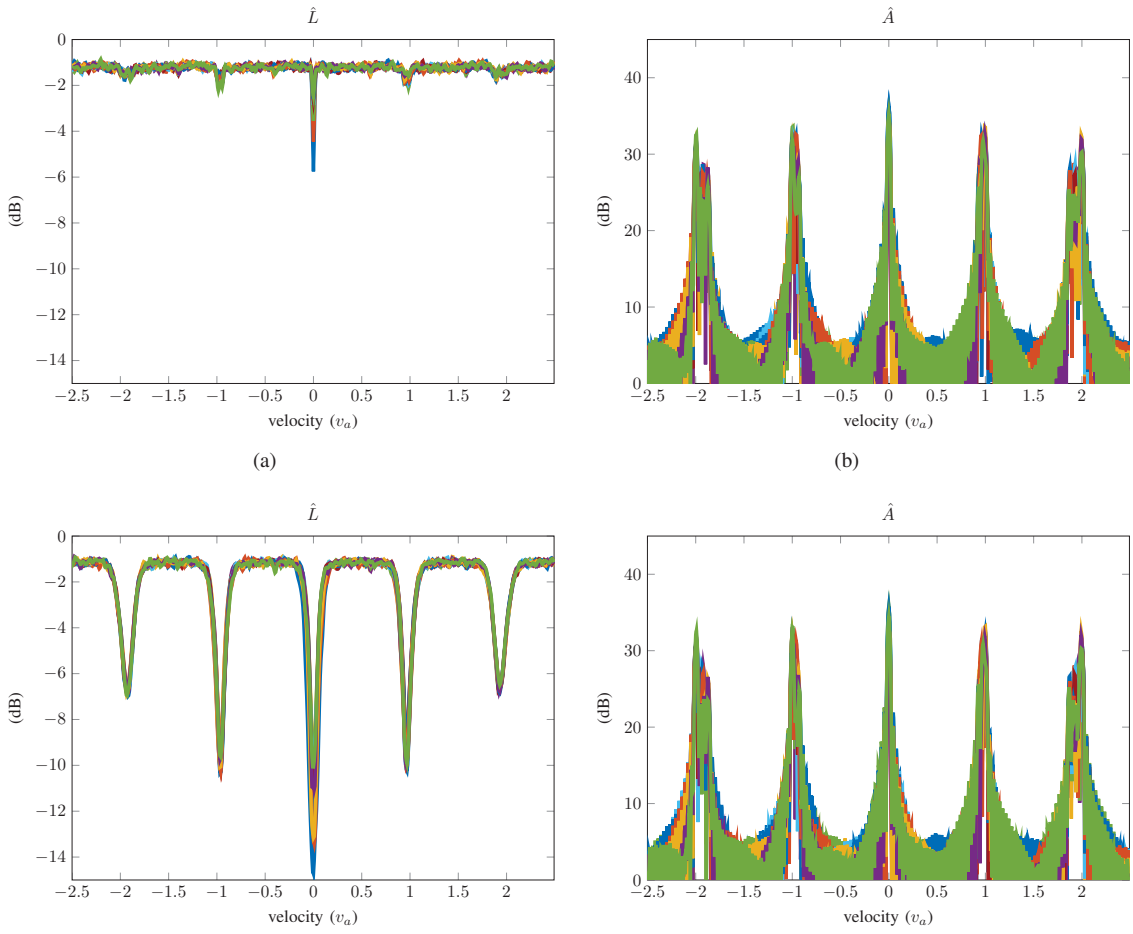


Fig. 11. Estimated SINR-loss from experimental radar data (velocity cut plot at each range-gate). Data parameters: $F_c = 25$ GHz, $B = 2$ GHz, $T_r = 4$ ms, $M = 32$, $K = 6$ (range-gate: 58–64) [according to [49] $\mu = B/(F_c - B/2)/K$, $v_a = 2T_r(F_c - B/2)/c$; $\mu = 0.0104$, $v_a \approx 1.56$ m/s $c/(2B) = 7.5$ cm]. Processing parameters: $n_{va} = 5$, $T \approx 2.4KM$. (a)-(b) Free-target data set 2017-07-27-14-05-00 with fan off. (c)-(d) Free-target data set 2017-07-27-13-57-54 with fan on.

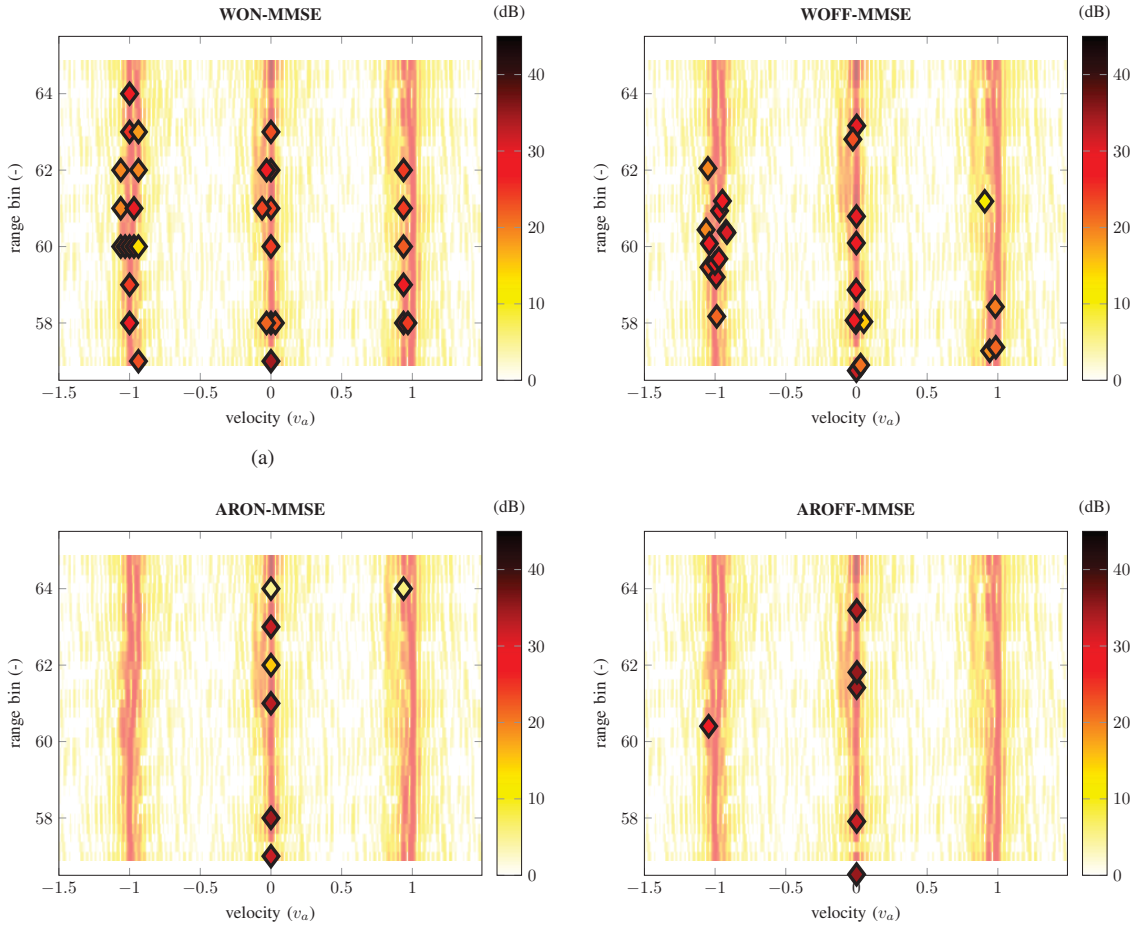


Fig. 12. Range-velocity map (modulus of the complex amplitude only). Data set: 2017-07-27-13-42-41. Data parameters: $F_c = 25$ GHz, $B = 2$ GHz, $T_r = 4$ ms, $M = 32$, $K = 8$ ($v_a = 1.56$ m/s and $c/(2B) = 7.5$ cm). Processing parameters: $n_{va} = 3$, $n_{\text{zp}}^v = n_{\text{zp}}^r = 1$, initial pulse: 1178), initial range-gate: 57, β_0, β_1 such that $\text{m}_{\sigma_x^2} = 45$ dB and $\text{std}_{\sigma_x^2} \approx 15$ dB, $\gamma_0 = \gamma_1 = 0$, $\bar{\boldsymbol{R}}_{\phi}^{-1} = \mathbf{0}$, $P = 3$, $N_{bi} = 2E^{+3}$ and $N_r = 200$. Diamond markers indicate estimated scatterers. Coherent integration depicted as a transparent background.

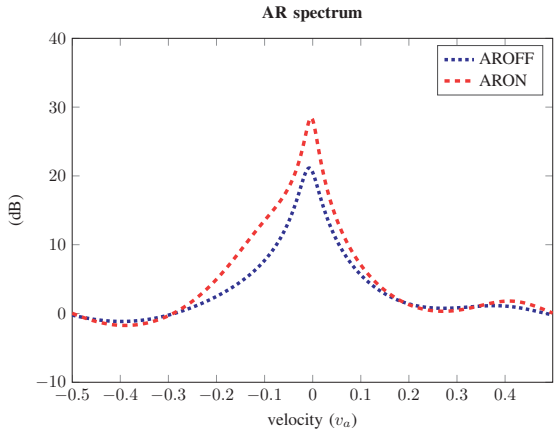


Fig. 13. AR spectrum. Scenario is that of Fig. 12.