Sampling-based spotlight SAR image reconstruction from phase history data for speckle reduction and uncertainty quantification*

Victor Churchill† and Anne Gelb‡

Abstract. Spotlight mode airborne synthetic aperture radar (SAR) is a coherent imaging modality that is an important tool in remote sensing. Existing methods for spotlight SAR image reconstruction from phase history data typically produce a single image estimate which approximates the reflectivity of an unknown ground scene, and therefore provide no quantification of the certainty with which the estimate can be trusted. In addition, speckle affects all coherent imaging modalities causing a degradation of image quality. Many point estimate image reconstruction methods incorrectly treat speckle as additive noise resulting in an unnatural smoothing of the speckle that also reduces image contrast. The purpose of this paper is to address the issues of speckle and uncertainty quantification by introducing a sampling-based approach to SAR image reconstruction *directly from* phase history data. In particular, a statistical model for speckle as well as a corresponding sparsity technique to reduce it are directly incorporated into the model. Rather than a single point estimate, samples of the resulting joint posterior density are efficiently obtained using a Gibbs sampler, which are in turn used to derive estimates and other statistics which aid in uncertainty quantification. The latter information is particularly important in SAR, where ground truth images even for synthetically-created examples are typically unknown. While similar methods have been deployed to process formed images, this paper focuses on the integration of these techniques into image reconstruction from phase history data. An example result using real-world data shows that, when compared with existing methods, the sampling-based approach introduced provides parameter-free estimates with improved contrast and significantly reduced speckle, as well as uncertainty quantification information.

Key words. sampling-based image reconstruction, Bayesian uncertainty quantification, synthetic aperture radar

AMS subject classifications. 94A08, 68U10, 62F15, 65C05, 65F22, 62G07, 60J22.

1. Introduction.

Spotlight mode airborne synthetic aperture radar (SAR)[1] is a widely-used imaging technology for surveillance and mapping. Because SAR is capable of all-weather day-or-night imaging, it overcomes several challenges faced by optical imaging technologies and is an important tool in modern remote sensing, [42]. Applications where SAR imaging is important include areal mapping and analysis of ground scenes in environmental monitoring, remote mapping, and military surveillance, [1]. It is imperative in many of these applications to obtain practically artifact- and noise-free SAR images on which practitioners can rely. However, several issues with existing methods for SAR image reconstruction from phase history data pose challenges to this goal.

First, SAR image reconstruction is a large problem, requiring efficient storage and methods. Large image and data sizes prohibit the use of traditional matrix-based methods for

*Submitted to the editors DATE.

Funding: This research was funded in part by the NSF grants DMS #1502640 and DMS #1912685, the Air Force Office of Sponsored Research grant #FA9550-18-1-0316, and the Office of Naval Research MURI grant #N00014-20-1-2595.

†The Ohio State University, Columbus, OH (churchill.77@osu.edu)

‡Dartmouth College, Hanover, NH (annegelb@math.dartmouth.edu).

[1] To be concise, we will refer to spotlight mode airborne SAR as just SAR from now on.

linear inverse problems, as even storing dense matrices of the necessary size is problematic. Second, SAR is a coherent imaging system, meaning that both the collected data and the reflectivity image are complex-valued. While typically only the magnitude is viewed, the phase information should not be neglected in the image formation process, and is important for downstream tasks like interferometry, [48]. All coherent imaging modalities are affected by speckle, a multiplicative-noise-like phenomenon which causes grainy-looking images. For applications such as target identification, removing speckle as well as returns from non-targets in order to increase contrast around objects of interest is desirable. Existing methods for SAR image reconstruction from phase history data usually do not directly address speckle and post-processing operations like smoothing and filtering are typically necessary, [3, 46, 26, 53, 25, 22]. Critically, even if appropriate modeling is assumed in post-processing, information which was lost in the initial image formation process from data cannot be retrieved.

There are several common image formation methods for SAR. Basic, fast methods that rely on an inverse non-uniform fast Fourier transform (NUFFT), [40], provide no speckle reduction, while sparsity-based methods that rely on ℓ_1 regularization, [2, 19, 58, 57], disregard the physical meaning of speckle and instead choose to place penalties on approximate pixel magnitude values. Conflating speckle with the usual additive noise makes parameter selection for the ℓ_1 regularization penalty term very difficult (and essentially without physical meaning) in practice. Generally, using one of these existing methods results in a single image, typically a maximum likelihood or maximum *a posteriori* point estimate, that approximates the unknown ground truth. These predictions are statistics of a distribution and not probabilistic themselves, and therefore provide no information about the statistical confidence with which we can trust the features in the resulting images, e.g., which are more likely objects of interest and which are more likely attributed to speckle or noise. This makes forming reliable images difficult, particularly in SAR where even many synthetically-created examples have unknown true reflectivity.[2]

The purpose of this paper is to address the issues of speckle reduction and uncertainty quantification in SAR image reconstruction, while maintaining enough efficiency to enable working with image sizes typical in real-world applications. Significantly, in what follows we are able to address these problems within the process of reconstructing SAR images **directly** from phase history data, as opposed to relying on altering images that have already been reconstructed or otherwise processed. This is important because the phase history data is the primary source of information we have about the scene. When an image is formed, there is a loss of information. For example, when an operator is applied to the data in order to form an image, it may cause cancellation that an approximate inverse operator (if available) cannot retrieve. Any further processing, e.g. speckle reduction, performed on its pixel values therefore involves starting at an information deficit. Hence, it is advantageous to work directly from the data when possible. We achieve this direct reconstruction first and foremost by taking a more robust approach to estimation, sampling an entire posterior density estimate rather than just computing a point estimate. This will allow us to compute estimates and uncertainty

[2]We point out that there are some instances for which benchmarks for despeckling have been established, [28]. However, these tests operate on simulated magnitude-only images that have already been formed, while the focus of this paper is on reconstructing complex-valued images directly from real-world phase history data.

quantification information such as standard deviation and confidence intervals for all unknown parameters in the model. Our approach uses the hierarchical Bayesian prior structure from [60] and directly incorporates coherent imaging and speckle into the prior density. The prior density is also formed to encourage sparsity in order to reduce speckle and increase contrast. Conjugate priors are used so that the resulting posterior can be efficiently sampled by using a Gibbs sampler and a NUFFT. It is important to note that all parameters in the model are prescribed, requiring **no user input**. We note that sampling-based methods using this same prior structure have been developed to quantify uncertainty in basic real-valued linear inverse problems such as image reconstruction, see e.g. [8], and have also been applied to SAR imaging tasks such as moving target inference, [51], passive SAR image reconstruction, [62], and speckle noise model selection, [45]. Specifically, the goals and methods described here echo part of those developed in [51]. However, whereas [51] operates on SAR images that have already been formed, here the focus is on starting the problem earlier with collected SAR phase history data and integrating these techniques into the image reconstruction process itself rather than as a post-processing in the image domain for reasons aforementioned. In [62], a similar procedure is formed for image reconstruction from phase history data in the case of passive SAR. In [62], a spike-and-slab prior to encourage sparsity on the unknown image, while this paper utilizes an improper prior described below as in [60] and tests on spotlight mode airborne SAR data. In [45], an MCMC-based procedure is used to select an appropriate noise model for speckle. Here, the fully-developed speckle model [42] is used, although later it is discussed that this model is only appropriate in target-free regions. In addition, other Bayesian methods for a variety of SAR functions such as classification, [65], and image reconstruction, [30, 61, 63, 64], have been developed using the same prior as well as the deterministic estimation procedure from [60] known as sparse Bayesian learning (SBL) or Bayesian compressed sensing (BCS), [43]. In this paper, the focus is on using the prior model in a sampling-based framework for spotlight mode airborne SAR image reconstruction directly from phase history data.

The rest of this paper is organized as follows. Section 2.1 derives the hierarchical Bayesian prior of [60] from scratch, emphasizing the incorporation of coherent imaging, speckle, and sparsity using conjugate priors. Section 2.2 outlines an efficient sampling method for the resulting posterior based on the real-valued method of [8], highlighting the advantages over computing point estimates. Section 3 shows a real-world example using the Air Force Research Laboratory's GOTCHA Volumetric Data Set 1.0, [18]. In addition to the added benefit of uncertainty quantification information, the results suggest that the proposed method provides estimates with better contrast and reduced speckle when compared with other methods for reconstructing SAR images from phase history data. Some concluding remarks and ideas for future directions are provided in Section 4.

2. Methods.

2.1. Hierarchical Bayesian Model for SAR. This section begins by specifying the linear system used to model the relationship between SAR images and phase history data. We provide background on, as well as issues with, existing SAR image reconstruction methods and then describe our approach to address these issues. Next we re-derive the hierarchical prior structure from [60] in order to account for the speckle phenomenon, as well as to encourage sparsity. We work with the fully-developed speckle model to form a posterior density for all latent variables, which is analytically computed.

2.1.1. Discrete Model. We consider the process of reconstructing an image from collected SAR phase history data, or electromagnetic scattering data. To collect spotlight mode airborne SAR data,[3] an airborne sensor traverses a circular flight path, periodically transmitting an interrogating waveform in the form of high bandwidth pulses at equally-spaced azimuth angles θ toward an illuminated circular region of interest $\mathcal{D} = \{(x, y)|x^2 + y^2 \leq R^2\}$. The emitted energy pulses impinge on targets in the illuminated region that scatter electromagnetic energy back to the sensor. The sensor measures and processes the reflected signal. The demodulated data, called a phase history, is passed on to an image reconstruction processor. This paper concerns the image reconstruction step, which produces a reconstruction of the two-dimensional electromagnetic reflectivity function of the illuminated ground scene from SAR phase history data. For a detailed overview of SAR and basic image reconstruction techniques, see e.g. [34, 39, 42, 40]. Traditionally, SAR images are formed using back projection, see e.g. [40]. However, as back projection can produce streaking and sidelobe artifacts, we focus instead on the following linear model for reconstruction.

The measured SAR phase history data can be modeled as a continuous non-uniform Fourier transform of the reflectivity function. Given a constant elevation angle ϕ between the flight path and $\mathcal{D}$, the reflected waveforms are of the form

$$\hat{f}(\omega(t), \theta) = \int\int_{\mathcal{D}} f(x, y) \exp\left(-i\frac{4\pi\omega(t)\cos\phi}{c}(x, y)\cdot(\cos\theta, \sin\theta)\right) dxdy, \tag{2.1}$$

where c is the speed of light, [56]. Hence the phase history data $\hat{f}(\omega(t), \theta)$ are the two-dimensional Fourier transform of the reflectivity function $f(x, y)$. For details and assumptions relied upon to make this realization, see e.g. [42, 56].

To discretize (2.1), consider $\hat{f}(\omega(t), \theta)$ for a discrete set of azimuth angles $\{\theta_j\}$, and a set of time steps corresponding to a discrete set of frequency values $\{\omega_k\}$, [56]. Then we have discretized (2.1) as the complex-valued linear system

$$\hat{\mathbf{f}} = \mathbf{F}\mathbf{f} + \mathbf{n}. \tag{2.2}$$

The objective is to infer a posterior density for $\mathbf{f}$ given $\hat{\mathbf{f}}$, where $\hat{\mathbf{f}} \in \mathbb{C}^M$ is the vertically-concatenated phase history data, $\mathbf{F} \in \mathbb{C}^{M\times N}$ is a two-dimensional discrete non-uniform Fourier transform matrix, and the vector $\mathbf{f} \in \mathbb{C}^N$ is the vertically-concatenated unknown reflectivity image matrix. Note that M is the length of the data and N the number of pixels in the image. Also note that by using the discrete Fourier transform in (2.2) we introduce both aliasing error and the Gibbs phenomenon. We note that (2.2) is a fairly simple model for the relationship between image and data in SAR. It is also common to modify (2.2) to include autofocusing for the purpose of phase error reduction, [58]. While such modifications are not a primary concern in this investigation, they can be incorporated into the proposed method in a straightforward manner. Finally, $\mathbf{n} \in \mathbb{C}^M$ represents model and measurement error.

[3]While we believe the method developed in this investigation may be suitably modified to fit other SAR modalities and corresponding models, further study and experimentation is required.

Throughout this paper we assume $\mathbf{n}$ is complex circularly-symmetric white Gaussian noise. That is, for each element i of $\mathbf{n}$, $\mathbf{n}_i \sim \mathcal{CN}(0, \beta^{-1})$ i.i.d., where $\beta^{-1} > 0$ is the noise variance. The assumption of white Gaussian noise is often made in SAR, but a diagonal covariance matrix implying independently distributed noise across each pixel with a potentially different variance can also be accommodated in the proposed sampling procedure. Nevertheless, we focus on the white Gaussian noise assumption. This yields the Gaussian likelihood function

$$p(\hat{\mathbf{f}}|\mathbf{f}, \beta) \propto \beta^M \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2\right), \tag{2.3}$$

which can be read as "the probability of $\hat{\mathbf{f}}$ given $\mathbf{f}$ and β," and measures the goodness of fit of the model (2.2). Note that $||\mathbf{g}||^2 := \mathbf{g}^H\mathbf{g}$ with $\mathbf{g}^H$ the conjugate transpose of $\mathbf{g}$. In the Bayesian approach to estimation, all quantities of interest are viewed as random variables, with probability distributions describing their behavior. Known quantities, e.g. SAR phase history data, are called observable variables, and unknown quantities, e.g. the reflectivity image, are called latent variables. The goal is to infer the latent variables from the observable variables. Encoding these quantities as random variables does not contradict that they are defined quantities, but rather expresses our lack of certainty about their values.

Recall that Bayes' theorem tells us that

$$\text{Posterior density} \propto \text{Likelihood function} \times \text{Prior density}.$$

The posterior density is built from the prior density on the latent variable, i.e. our belief about it before data has been considered, and the likelihood function, which governs how well the data fits the model. In this way the posterior is a synthesis of prior belief and information carried by the data, [13, 14, 44]. The deterministic approach to SAR image reconstruction, explained below in a digression, typically only obtains a single image that estimates the ground truth reflectivity of the scene. However, in a Bayesian formulation an entire posterior density function for the latent variables is sought. Hence, in an effort to better describe the unknown reflectivity, we take an approach to compute an entire posterior density from which samples can be drawn and statistics can be computed.

2.1.2. Digression on existing estimation techniques. Recall that the likelihood function is defined as the probability distribution of the observed variables conditional on the other variables. In this digression to explain the current state of SAR image reconstruction from phase history data, we will use the likelihood function in (2.3) to derive a few different inversion techniques used to find an estimate for the unknown quantity $\mathbf{f}$. As of now, the only model parameter is the noise variance β^{-1}, which in the deterministic methods is considered a known, or at least asserted, quantity (hence observable).

Perhaps the most straightforward way to estimate $\mathbf{f}$ from $\hat{\mathbf{f}}$ is to maximize the likelihood function. From the Gaussian likelihood defined above by (2.3), this estimate is

$$\mathbf{f}^*_{ML} = \arg\max_{\mathbf{f}} \left\{p(\hat{\mathbf{f}}|\mathbf{f}, \beta)\right\} = \arg\min_{\mathbf{f}} \left\{||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2\right\}. \tag{2.4}$$

For an overdetermined discrete non-uniform Fourier transform, $\mathbf{F}^H\mathbf{F} \approx \mathbf{I}$, hence we have that $\mathbf{f}^*_{ML} = \mathbf{F}^H\hat{\mathbf{f}}$. Due to the size of $\mathbf{F}$, storing and applying it as a matrix is not practical in real-world problems. Hence, a NUFFT, specifically the implementation in [33], is used to efficiently

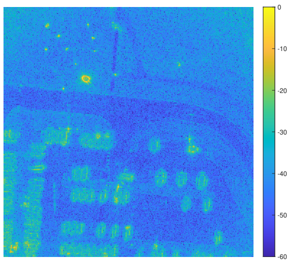


Figure 1: Parking lot SAR image reconstructed using the NUFFT from phase history data in GOTCHA dataset, [18].

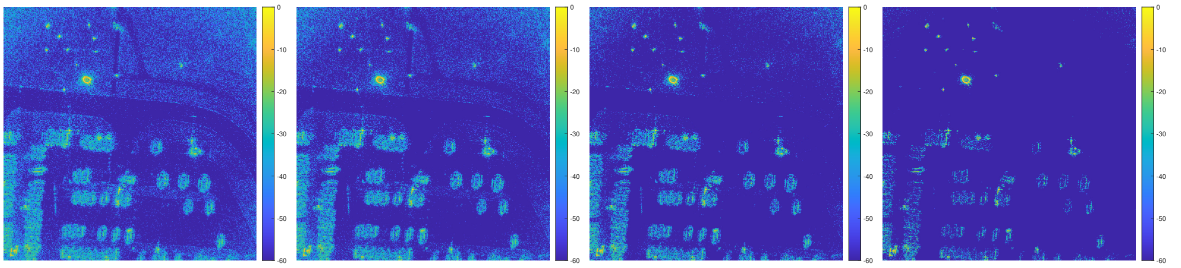
Figure 2: Optical images of parking lot being imaged in GOTCHA dataset, [18]. The scene contains a variety of calibration targets, such as primitive reflectors like the tophat shown, a Toyota Camry, forklift, and tractor.

apply the action of this matrix instead to avoid storage and accelerate the computation.[4] General information on NUFFTs can be found in [33, 41, 47]. The preceding estimate is therefore frequently referred to as a NUFFT reconstruction, as it only requires an inverse NUFFT application in order to invert the data. Generally speaking, the reflectivity image can be found by interpolating the typically polar grid of measured samples in frequency space to an equally spaced rectangular grid, then computing an inverse *uniform* fast Fourier transform, [1]. Although the NUFFT is computationally efficient, noisy data and model error can cause artifacts or a noisy image. An example is shown in Figure 1 using the GOTCHA parking lot data set, [18].[5] While these SAR images look much different than the optical images shown in Figure 2, key features from the parking lot can be recognized such as the roads, curbs, and cars. Note that only the magnitude of this complex-valued reflectivity image is viewed here. Observe also the very grainy appearance due to the speckle phenomenon, which we later discuss at length. Finally, since NUFFT reconstruction provides only this one image, we have no means of knowing whether or not all features shown are objects of interest.

To improve on the maximum likelihood (or NUFFT) estimate given by (2.4), the cost

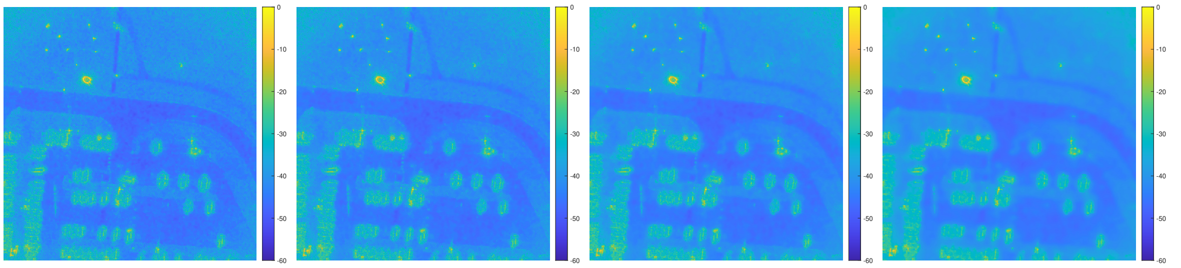
[4] For efficient use of space, we continue to use $\mathbf{F}$ and $\mathbf{F}^H$ notation despite using NUFFTs in their places for the actual implementation. Using a function as opposed to a matrix as a forward operator is commonly seen in nonlinear inverse problems, [7, 9, 10].

[5] The GOTCHA data used are fully specified in Section 3.



(a) (b) (c) (d)

Figure 3: Parking lot SAR images reconstructed with ℓ_1 regularization with regularization parameter $\lambda =$ (a) 1/80; (b) 1/60; (c) 1/40; (d) 1/20.



(a) (b) (c) (d)

Figure 4: Parking lot SAR images reconstructed with TV regularization with regularization parameter $\lambda =$ (a) 1/160; (b) 1/120; (c) 1/80; (d) 1/40.

function is frequently regularized by adding a penalty term on the ℓ_1 norm of $|\mathbf{f}|$ or a transform $\mathbf{T}|\mathbf{f}|$

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \left\{ \frac{\beta}{2} ||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2 + \lambda ||\mathbf{T}|\mathbf{f}|||_1 \right\}. \tag{2.5}$$

In addition to regularizing the ill-posed problem, where more than one $\mathbf{f}$ may satisfy the model equation, this formulation encourages sparsity in the magnitude $|\mathbf{f}|$. The phase, which is not modeled as sparse, [42], is governed only by the least squares fit term. Equation (2.5) in general has no direct solution and must be minimized using a convex optimization method like the alternating direction method of multipliers (ADMM), [11]. The ℓ_1 regularization term in (2.5) imposes the sparsity penalty on $\mathbf{f}$. In the field of compressive sensing, [16] the sparsity prior parameter λ and noise variance β^{-1} are often combined and relabeled as the regularization parameter, which balances the fidelity term, the sparsity penalty, and noise reduction. Figure 3 shows four reconstructed images employing (2.5) with $\mathbf{T} = \mathbf{I}$ using $\beta = 1$ and four different values of λ. Figure 4 also shows four reconstructed images employing (2.5), in this case with $\mathbf{T}$ defined as the total variation (approximate gradient) operator, again using $\beta = 1$ and four different values of λ. We see that the ℓ_1 regularization reconstruction does a decent job at sparsifying the image, i.e. drawing values to zero, especially along the road area which is fairly smooth and hence won't scatter much electromagnetic energy back to the sensor. This has the effect of increased contrast making targets like the cars more clearly visible. However, in the reconstructions using smaller λ values, there is still significant speckle causing an overall grainy appearance in some of the rough grassy areas where there are in fact no visible targets of interest. Moreover, reconstructions using larger values of λ are overly sparse – displaying an overall disconnected appearance in the car reflectivities. The TV regularization reconstructions on the other hand handles the grainy speckle issue very well, smoothing out much of the image. The issue here is the block-like appearance, which is an artifact known to occur in TV regularization.[6] These blocky regions occur to a varying degree depending on λ, and they increase background values potentially making it more difficult to determine the targets in the scene, particularly without any uncertainty quantification information. Both methods have been extensively applied in SAR (see e.g. [58, 29, 56, 2, 19]). In addition, methods that use weighted ℓ_1 or ℓ_2 norm regularization have had some success over standard ℓ_1 regularization, [17, 20, 23, 27], although there are issues with robustness, [23]. As is clear from Figures 3 and 4, the choice of the regularization parameter λ, which balances the model fidelity with a penalty on the magnitude of the sparsity domain, is critical. While the option to tune this parameter gives the user the ability to perhaps affect which magnitudes are large enough to be considered objects of interest, in the absence of ground truth this choice can be very difficult. These ℓ_1-norm-based methods also do not properly describe speckle and provide no uncertainty quantification. For these reasons we are motivated to use a probabilistic framework, which we now describe.

From the cost function used in (2.5), the observation can be made that had the prior probability distribution

$$p(\mathbf{f}|\lambda) \propto \exp\left(-\lambda||\mathbf{T}|\mathbf{f}|||_1\right), \tag{2.6}$$

been invoked, the resulting posterior density would be

$$p(\mathbf{f}|\hat{\mathbf{f}}, \beta, \lambda) \propto p(\hat{\mathbf{f}}|\mathbf{f}, \beta)p(\mathbf{f}|\lambda) \propto \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2 - \lambda||\mathbf{T}|\mathbf{f}|||_1\right). \tag{2.7}$$

It is clear that maximizing (2.7) would yield (2.5), and hence (2.5) is known as a maximum a posteriori (MAP) estimate. This ℓ_1 prior is often chosen in the field of compressed sensing, where limited data is collected and sparsity is believed in $\mathbf{T}|\mathbf{f}|$. Of course this is not the only prior distribution that can be used and others would invoke other *a priori* beliefs. From this discussion above, it is clear that the regularization penalty term within the cost function imposes the *a priori* belief specified in the prior probability distribution. It is also evident that without prior information of λ or β, they will be difficult to choose. Hence we take the view that these parameters should also be estimated. We note that due to the difficulty of minimizing the ℓ_1 norm of the magnitude of a complex vector, in SAR, typically an approximation

[6] This is also often called the stair-casing effect due to the tendency of TV regularized solutions to be piecewise constant.

$|\mathbf{f}| = \Theta^*\mathbf{f}$ is made, where $\Theta_{j,j} = \text{angle}(\tilde{\mathbf{f}}_j)$ and $\tilde{\mathbf{f}}$ is an approximate cheaply computed solution such as the NUFFT image, [56]. This has two negative consequences. First, this method no longer regularizes the sparsity of the magnitude of $|\mathbf{f}|$, but just an approximation to the magnitude. In addition, the regularization term no longer corresponds to any prior distribution as the data was considered in order to form the initial estimate. The proposed method corrects both of these issues, working directly from phase history data and incorporating an appropriate prior on $\mathbf{f}$. Finally, there are a few issues with MAP estimates in general. The maximum is not a categorically strong representative of the posterior density, and in general sampling is a better way to interrogate a density than finding its maximum. A consequence of only estimating with the maximum is that once again we do not know the certainty with which we can trust the estimate or the features thereof. Hence we have no way of knowing which structures in the reflectivity estimate are truly there and which are noise or artifacts.

The above discussion inspires us to form a new approach which directly uses the SAR phase history data, which we describe in detail in the following sections. Considering each element of the model for SAR image reconstruction, a hierarchical Bayesian model [60] is constructed using conjugate priors, and a Gibbs sampler is used to sample the resulting posterior density. The result is a set of samples from the posterior density, from which a variety of statistics (including the sample mean and sample variance) can be computed and used not only to estimate the image but also the speckle and noise and in general to quantify uncertainty.

2.1.3. Hierarchical Prior. With the likelihood given by (2.3), the next step in computing an entire posterior density is to specify a prior density for the latent variable $\mathbf{f}$ as mentioned above in the digression on existing estimation. Recall that the prior expresses a belief about a quantity before observation. The hierarchical prior used in the proposed method is identical to the one used in [60], which was formulated for sparse regression and classification. Below it is re-derived with coherent imaging and speckle justifying its use. This prior has been used in MCMC-based methods for SAR image processing before, including moving target inference [51], passive SAR reconstruction [62], and noise model selection, [45], as well as in deterministic algorithms for image reconstruction, [64, 61, 30, 63].

We use the fact that SAR images are affected by the speckle phenomenon as a prior, effectively including an appropriate statistical characterization of speckle within our model. Speckle, which occurs in all coherent imaging and is often misidentified and mischaracterized as noise, causes a complicated granular pattern of bright and dark spots throughout an image, [42]. Although speckle is in fact signal and *not* noise, it nonetheless degrades the image quality by lowering the contrast, and hence when attempting to identify targets in a scene it is desirable to remove it. We note and recognize that in some applications, removing speckle is not desirable and the speckle is in fact leveraged via speckle-tracking for other tasks such as change detection. While speckle reduction is the goal in this paper, and hence priors are so-chosen, later in this section we describe how our model can in fact be easily adapted to simply model speckle and not reduce it. Speckle reduction is often tackled using denoising techniques, e.g. the TV scheme described in Section 2.1.2, or by filtering, [3], or by other post-processing techniques for speckle denoising, [26, 53, 25, 22].[7] Here instead we directly incorporate the speckle into the image reconstruction model, so that it is properly characterized as part of

[7]Note that the TV scheme described in this paper is incorporated directly into the inverse problem while

the data. Specifically we employ the fully-developed speckle model, [42, 57, 29, 52]. Although this model is really only appropriate when no dominant scatterers are present in a resolution cell, it has been previously invoked in spotlight mode airborne SAR, [42, 57]. In addition, we hypothesize that since we seek to reduce speckle specifically where there are no dominant scatterers (and the fully-developed speckle model applies), dominant scatterers will remain and thus we are able obtain the desired result of reducing speckle in regions without targets. We note that, unlike despeckling techniques that operate on magnitude-only images, [28], the product of the method will be despeckled complex-valued images. While further exploration is planned for future work, we expect this will be critical for coherent downstream tasks such as interferometry and change detection, where coherent images are required.

We now provide details of the fully-developed speckle model. Assume the real and imaginary parts of each image pixel i, $\mathrm{Re}(\mathbf{f}_i)$ and $\mathrm{Im}(\mathbf{f}_i)$, are respectively i.i.d. Gaussian with variance $\boldsymbol{\alpha}_i^{-1}$. That is, $\mathrm{Re}(\mathbf{f}_i), \mathrm{Im}(\mathbf{f}_i) \sim \mathcal{N}(0, \boldsymbol{\alpha}_i^{-1})$. By independence, $\mathrm{Re}(\mathbf{f}), \mathrm{Im}(\mathbf{f}) \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\alpha})^{-1})$. This is conveniently encoded by $\mathbf{f} \sim \mathcal{CN}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\alpha})^{-1})$ which means that $\mathbf{f}$ is circularly-symmetric complex Gaussian with density

$$p(\mathbf{f}|\boldsymbol{\alpha}) \propto \prod_{i=1}^{N} \boldsymbol{\alpha}_i \exp\left(-\frac{1}{2}||\sqrt{\boldsymbol{\alpha}} \odot \mathbf{f}||^2\right), \tag{2.8}$$

where $\odot$ is elementwise multiplication. Thus we see that the prior on the magnitude $|\mathbf{f}_i| = \sqrt{\mathrm{Re}(\mathbf{f}_i)^2 + \mathrm{Im}(\mathbf{f}_i)^2}$ is a Rayleigh probability distribution with mean proportional to $\boldsymbol{\alpha}_i^{-1}$. This is the standard specification for fully-developed speckle, [42, 57]. Because a change in the magnitude of each pixel $|\mathbf{f}_i|$ is proportional to a change in $\boldsymbol{\alpha}_i^{-1}$, the speckle phenomenon has also been modeled as a multiplicative noise, [4, 21]. As already mentioned, there are many techniques developed to reduce speckle, [3]. Of note here is that we address the speckle directly by including it in our model with the prior given by (2.8), and later estimating the associated speckle parameters $\boldsymbol{\alpha}_i^{-1}$. This is accomplished through sampling as opposed to attempting to quantify the remaining speckle via post-image-reconstruction techniques. Note that by parameterizing $\mathbf{f}$ with $\boldsymbol{\alpha}$ we are introducing another latent variable, which clearly provides a *computational* challenge (but not a methodological one), [50].

Since we now have a likelihood given by (2.3) and a prior defined in (2.8), we could compute a posterior for $\mathbf{f}$ if β and $\boldsymbol{\alpha}$ are specified. Specifically, by Bayes' theorem, the posterior density for $\mathbf{f}$ would be

$$p(\mathbf{f}|\hat{\mathbf{f}}, \boldsymbol{\alpha}, \beta) \propto p(\hat{\mathbf{f}}|\mathbf{f}, \beta) p(\mathbf{f}|\boldsymbol{\alpha}, \beta) \propto \beta^M \prod_{i=1}^{N} \boldsymbol{\alpha}_i \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2 - \frac{1}{2}||\sqrt{\boldsymbol{\alpha}} \odot \mathbf{f}||^2\right). \tag{2.9}$$

We could take the same approach as described in Section 2.1.2 of obtaining a MAP estimate leading to the optimization problem

$$\mathbf{f}^*_{MAP} = \arg\max_{\mathbf{f}} p(\mathbf{f}|\hat{\mathbf{f}}, \boldsymbol{\alpha}, \beta) = \arg\min_{\mathbf{f}} \left\{\frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2 + \frac{1}{2}||\sqrt{\boldsymbol{\alpha}} \odot \mathbf{f}||^2\right\}. \tag{2.10}$$

the methods in [26, 53, 25, 22] are used only *after* the image is formed.

We can dissect the components in (2.10) as a least-squares fidelity term coming from the likelihood function which measures the fit of the data to the proposed $\mathbf{f}$ followed by a regularization term which penalizes the ℓ_2 norm of $\mathbf{f}$ after being transformed by $\sqrt{\boldsymbol{\alpha}}$. Resulting from the Gaussian prior in (2.8), regularization with the ℓ_2 norm, known as Tikhonov regularization or ridge regression, [38], can be used to encourage smoothness in the solution.

As mentioned above, using a MAP estimate like (2.10) as the solution is limiting – first because it may not be representative of the posterior and second because it provides no information about the statistical confidence of the estimate of each recovered pixel value, or in any other recovered features of the image, [49]. Finally, the regularization parameters for both the cost function and prior in the MAP estimate approach (analogous to β and $\boldsymbol{\alpha}$ here) are *user-specified*. Yet they are truly unknown and therefore should be inferred from the data. For these reasons we take a different approach than (2.10) and seek the *joint* posterior $p(\mathbf{f}, \boldsymbol{\alpha}, \beta|\hat{\mathbf{f}})$. Significantly, we will not only be estimating an entire density for the complex image, but also the speckle parameter $\boldsymbol{\alpha}$, which will lend clarity when determining whether or not the speckle reduction techniques are actually working,[8] as well as the noise parameter β. In order to calculate $p(\mathbf{f}, \boldsymbol{\alpha}, \beta|\hat{\mathbf{f}})$, we must define prior densities on $\boldsymbol{\alpha}$ and β. In general, we have no intuition for the values of $\boldsymbol{\alpha}$ and β, and we can encode that uncertainty by choosing uninformative priors to allow as much variation as possible and let the data choose.

Although there is no theoretical constraint on the type of prior used for β, in order to obtain an analytical form of the posterior, we follow [60] and choose a conjugate Gamma prior. That is, $\beta \sim \Gamma(c, d)$ with probability density function

$$p(\beta|c, d) \propto \beta^{c-1} \exp(-d\beta). \tag{2.11}$$

Similarly a conjugate Gamma prior is invoked on each element of $\boldsymbol{\alpha}$, i.e. $\boldsymbol{\alpha}_i \sim \Gamma(a, b)$ for each element $i = 1, \ldots, N$. By independence, $\boldsymbol{\alpha} \sim \Gamma(a, b)$ with

$$p(\boldsymbol{\alpha}|a, b) \propto \prod_{i=1}^{N} \boldsymbol{\alpha}_i^{a-1} \exp\left(-b \sum_{i=1}^{N} \boldsymbol{\alpha}_i\right). \tag{2.12}$$

Because the Gamma prior is conjugate to the Gaussian in (2.8), the prior and the posterior are from the same distribution family. That is, the individual posterior densities for β or $\boldsymbol{\alpha}$ will be Gamma. Note the dependence of (2.11) and (2.12) on parameters a, b, c, and d, which as in [8, 60] are chosen rather than inferred. In [8], analogous parameters in a real-valued model are chosen to reflect the uncertainty in the latent variable, making the prior uninformative. Specifically, $a, c = 1$ and $b, d = 10^{-4}$. While our focus here is on speckle reduction, our tests in this direction indicate that these parameters are appropriate for SAR image reconstruction in applications where speckle reduction is neither required nor desired, producing an estimate similar in appearance to an NUFFT image. On the other hand in [60], $a, b, c, d := 0$, resulting in an improper prior $p(\mathbf{f}_i) \sim 1/|\mathbf{f}_i|$, which is peaked at zero and hence encourages sparsity.[9]

[8] Without a reference ground truth image, speckle statistics are typically only estimated from small regions of images post-reconstruction, [3].

[9] To ensure numerical robustness in our implementation, we choose these parameters to be machine precision rather than 0.

Importantly, choosing a, b, c and d in this way removes any need for user-defined parameters in this model. Our previous work used this prior to perform edge detection from data similar to that seen in SAR, [24]. Nevertheless, the derivation below is done for general a, b, c, and d. We stress that a, b, c, and d, are the only parameters required to be defined in this model, and were not tuned beyond what is mentioned above.

2.1.4. Posterior Computation.

The form of the joint posterior density is achieved through the hierarchical Bayesian model described above, [8, 15, 12, 60], where the likelihood parameters $\mathbf{f}$ and β are given priors (with prior parameters $\boldsymbol{\alpha}$, c, and d), referred to as hyperparameters. The hyperparameter $\boldsymbol{\alpha}$ is also given a prior (called a hyperprior) with hyperhyperparameters a and b. By Bayes' theorem, the joint posterior for $\mathbf{f}$, $\boldsymbol{\alpha}$, and β is

$$\begin{aligned} p(\mathbf{f},\boldsymbol{\alpha},\beta|\hat{\mathbf{f}},a,b,c,d) &\propto p(\hat{\mathbf{f}}|\mathbf{f},\beta)p(\beta|c,d)p(\mathbf{f}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}|a,b) \\ &\propto \beta^{M+c-1}\prod_{i=1}^{N}\boldsymbol{\alpha}_i^a \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}}-\mathbf{F}\mathbf{f}||^2-\frac{1}{2}||\sqrt{\boldsymbol{\alpha}}\odot\mathbf{f}||^2-d\beta-b\sum_{i=1}^{N}\boldsymbol{\alpha}_i\right), \end{aligned} \tag{2.13}$$

where we recall that M and N are defined in (2.2). The algorithm in Section 2.2 for sampling (2.13) will require the individual posteriors for each latent variable. Because of the conjugate priors used, these can be found analytically. The posterior for $\mathbf{f}$ is Gaussian, for $\boldsymbol{\alpha}$ is a product of independent Gammas, and for β is Gamma. We have

$$p(\mathbf{f}|\hat{\mathbf{f}},\boldsymbol{\alpha},\beta) \propto \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}}-\mathbf{F}\mathbf{f}||^2-\frac{1}{2}||\sqrt{\boldsymbol{\alpha}}\odot\mathbf{f}||^2\right) \tag{2.14a}$$

$$p(\boldsymbol{\alpha}|\hat{\mathbf{f}},\mathbf{f},\beta,a,b) \propto \prod_{i=1}^{N}\boldsymbol{\alpha}_i^a \exp\left(-\frac{1}{2}||\sqrt{\boldsymbol{\alpha}}\odot\mathbf{f}||^2-b\sum_{i=1}^{N}\boldsymbol{\alpha}_i\right) \tag{2.14b}$$

$$p(\beta|\hat{\mathbf{f}},\mathbf{f},\boldsymbol{\alpha},c,d) \propto \beta^{M+c-1}\exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}}-\mathbf{F}\mathbf{f}||^2+d\beta\right). \tag{2.14c}$$

Therefore each latent variable can be sampled from the following distributions

$$\mathbf{f}|\hat{\mathbf{f}},\boldsymbol{\alpha},\beta \sim \mathcal{CN}\left((\beta\mathbf{F}^H\mathbf{F}+\mathrm{diag}(\boldsymbol{\alpha}))^{-1}\beta\mathbf{F}^H\hat{\mathbf{f}},(\beta\mathbf{F}^H\mathbf{F}+\mathrm{diag}(\boldsymbol{\alpha}))^{-1}\right) \tag{2.15a}$$

$$\boldsymbol{\alpha}|\hat{\mathbf{f}},\mathbf{f},\beta,a,b \sim \Gamma\left(1+a,\frac{1}{2}\mathbf{f}\odot\bar{\mathbf{f}}+b\right) \tag{2.15b}$$

$$\beta|\hat{\mathbf{f}},\mathbf{f},\boldsymbol{\alpha},c,d \sim \Gamma\left(M+c,\frac{1}{2}||\hat{\mathbf{f}}-\mathbf{F}\mathbf{f}||^2+d\right). \tag{2.15c}$$

In [60], the same posterior density is reached. However, rather than sampling the posterior, [60] takes the approach of computing a deterministic estimate in a method known as sparse

Bayesian learning (SBL) which we describe now for comparison purposes later. From (2.14a), the conditional posterior of $\mathbf{f}$ given values for $\boldsymbol{\alpha}$ and β is Gaussian with mean and variance

$$\boldsymbol{\mu} = \beta \boldsymbol{\Sigma} \mathbf{F}^H \hat{\mathbf{f}} \tag{2.16a}$$

$$\boldsymbol{\Sigma} = \left(\beta \mathbf{F}^H \mathbf{F} + \operatorname{diag}(\boldsymbol{\alpha})\right)^{-1}. \tag{2.16b}$$

If $\boldsymbol{\alpha}$ and β are estimated then (2.16) can be evaluated. Closed form estimates are not available, so the following update rules are used[10]

$$\boldsymbol{\alpha}_i^{(new)} = \frac{1 - \boldsymbol{\alpha}_i \boldsymbol{\Sigma}_{ii}}{|\boldsymbol{\mu}|_i^2}, \quad i = 1, \ldots, N, \tag{2.17a}$$

$$\beta^{(new)} = \frac{M - N + \sum_{i=1}^N \boldsymbol{\alpha}_i \boldsymbol{\Sigma}_{ii}}{||\hat{\mathbf{f}} - \mathbf{F}\boldsymbol{\mu}||^2}. \tag{2.17b}$$

Iterating between updates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (2.16) and $\boldsymbol{\alpha}$ and β in (2.17) until a convergence criterion has been reached, the mean $\boldsymbol{\mu}$ is used as the final image estimate. Although this algorithm provides a full density for $\mathbf{f}$, only point estimates are achieved for $\boldsymbol{\alpha}$ and β. We note that this algorithm has been used for reconstructing spotlight SAR images from phase history data before, [64], as well as other types of SAR image reconstruction, [30, 61, 63]. Figure 5(g) shows the parking lot scene reconstructed using SBL. It is evident that for this GOTCHA dataset, the image looks very similar to the ℓ_1 regularization reconstruction using a heavy penalty shown in Figure 5(f).

2.2. Sampling-based SAR Image Reconstruction. Now that the joint posterior has been specified (2.13), it remains to be defined how to learn information about and to interrogate it by efficiently gathering samples and later developing statistics. In this section, a sampling-based image reconstruction procedure based on that of the real-valued method in [8] is used to obtain approximate samples from each latent variable in (2.13). From these samples, various estimates and confidence statistics can be retrieved. Clearly (2.13) is not described by a known family of probability distributions. In fact, it is essentially the product of two Gaussian and two Gamma distributions. Therefore, it cannot be efficiently sampled directly. While a standard MCMC implementation like the Metropolis-Hastings algorithm could be used to obtain approximate samples, because of the conjugate prior structure, we can apply a Gibbs sampler, [37], which obtains approximate samples from the joint posterior (2.13) by sequentially sampling the individual posteriors for each latent variable given in (2.15a), (2.15b), and (2.15c). As with other Markov chain Monte Carlo (MCMC) methods, Gibbs sampling creates a Markov chain of samples, each of which is correlated with the other samples.

In terms of computational efficiency, an issue occurs in sampling the individual posterior for $\mathbf{f}$ given by (2.15a), where in general a large linear system determined by (2.14a) would need to be solved for $\mathbf{f}$. As previously mentioned, even storing the dense matrices $\mathbf{F}$ and $\mathbf{F}^H$ in real-world problems is not practical. However, because $\mathbf{F}$ is a non-uniform discrete Fourier transform matrix, we can utilize existing libraries to quickly apply a non-uniform fast Fourier

[10] For details, we refer the reader to Appendix A of [60].

transform (NUFFT), [33]. Broadly speaking, the NUFFT is performed by interpolating non-uniform Fourier mode quantities to a uniform grid so that a uniform FFT can be used, [33]. This is not without error of course, which mainly comes from the error accumulated when "gridding" non-uniform to uniform Fourier modes. We note that improving the accuracy of the NUFFT is also a widely studied topic, [59, 54, 32, 35], and further work will be needed to meaningfully quantify this error for this application. For the current investigation, in order to apply $\mathbf{F}$ efficiently, we employ a unitary operation (the uniform FFT). This means that the covariance matrix in (2.15a) can be approximately diagonalized as

$$(\beta \mathbf{F}^H \mathbf{F} + \operatorname{diag}(\boldsymbol{\alpha}))^{-1} \approx (\beta \mathbf{I} + \operatorname{diag}(\boldsymbol{\alpha}))^{-1}, \tag{2.18}$$

which can be very efficiently inverted using elementwise division on the diagonal,[11] yielding

$$\mathbf{f} \sim \mathcal{CN}\left((\beta \mathbf{I} + \operatorname{diag}(\boldsymbol{\alpha}))^{-1} \beta \mathbf{F}^H \hat{\mathbf{f}}, (\beta \mathbf{I} + \operatorname{diag}(\boldsymbol{\alpha}))^{-1}\right), \tag{2.19}$$

where $\mathbf{F}^H \hat{\mathbf{f}}$ can be precomputed and repeatedly reused for efficiency. Clearly using the right hand side in (2.18) introduces additional error, along with that from modifying the non-uniform modes in order to make them conform with a uniform grid, oscillations due to the Gibbs phenomenon, and model and measurement error. A potentially more accurate method would be to use elementwise division by $\beta + \boldsymbol{\alpha}$ as a preconditioner in a conjugate gradient descent scheme, however this would be far less efficient.

By combining (2.14), (2.15), and (2.18) we arrive at Algorithm 2.1, which produces K samples for $\mathbf{f}$, $\boldsymbol{\alpha}$, and β, each of which are approximately drawn from the joint posterior. Notice that each sample requires one NUFFT application.

Algorithm 2.1 An efficient MCMC method for sampling from $p(\mathbf{f}, \boldsymbol{\alpha}, \beta | \hat{\mathbf{f}}, a, b, c, d)$

Initiate $\mathbf{f}^0$, $\boldsymbol{\alpha}^0$, β^0. Choose a, b, c, d. Let $k = 0$;

Compute $\tilde{\mathbf{f}} = \mathbf{F}^H \hat{\mathbf{f}}$;

for $k = 1$ to K **do**

Compute $\mathbf{f}^{k+1} \sim \mathcal{CN}\left((\beta^k \mathbf{I} + \operatorname{diag}(\boldsymbol{\alpha}^k))^{-1} \beta^k \tilde{\mathbf{f}}, (\beta^k \mathbf{I} + \operatorname{diag}(\boldsymbol{\alpha}^k))^{-1}\right)$;

Compute $\boldsymbol{\alpha}^{k+1} \sim \Gamma\left(1 + a, \frac{1}{2}|\mathbf{f}^{k+1}|^2 + b\right)$;

Compute $\beta^{k+1} \sim \Gamma\left(M + c, \frac{1}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}^{k+1}||^2 + d\right)$;

end for

[11]In creating comparison images, this technique is also used to efficiently evaluate (2.16).

2.2.1. Chain convergence. The convergence rate for the Markov chain formed in Algorithm 2.1 is generally unknown, but how to determine chain convergence can be described as follows. First, a trace plot is often generated to display the history of a parameter's samples, showing where the chain has been exploring. These time series of the individually sampled parameters can then be used to gauge chain convergence, [13]. In particular, the average value of a converged chain should have no long term trend, and samples should look like random noise. Colloquially this is referred to as "mixing well." Since in our case there are $\sim 5 \times 10^5$ latent variables, displaying trace plots is not practical. Hence instead we adopt the following statistic from [8, 36] to determine chain convergence. In this case multiple chains are computed using randomly chosen starting points based on the observation that the variance within a single chain will converge faster than the variance between chains. A statistic is computed for each element of each latent variable, the value of which is a measure of convergence for that individual parameter. The derivation of this statistic described below closely follows [8].

Compute n_r chains (in our implementation this is done in parallel) each of length $2n_s$, keeping only the latter n_s samples. Let ψ_{ij} denote the ith sample from the jth chain for a single parameter, and define

$$B = \frac{n_s}{n_r - 1} \sum_{j=1}^{n_r} \left(\bar{\psi}_{\cdot j} - \bar{\psi}_{\cdot\cdot}\right)^2,$$

where $\bar{\psi}_{\cdot j}$ is the mean of the samples in the chain j, $\bar{\psi}_{\cdot\cdot}$ is the mean of the samples in every chain, and

$$W = \frac{1}{n_r} \sum_{j=1}^{n_r} s_j^2, \quad \text{with} \quad s_j^2 = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} \left(\psi_{ij} - \bar{\psi}_{\cdot j}\right)^2.$$

Hence B is a measure of the variance between the chains while W is a measure of the variance within each individual chain. The marginal posterior variance $\text{var}(\psi|\hat{\mathbf{f}})$ is then estimated by

$$\widehat{\text{var}}^+(\psi|\hat{\mathbf{f}}) = \frac{n_s - 1}{n_s} W + \frac{1}{n_s} B, \tag{2.20}$$

which is an unbiased estimate under stationarity, [36]. From this variance estimate, we compute the desired statistic

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\psi|\hat{\mathbf{f}})}{W}}, \tag{2.21}$$

which tends to 1 from above as $n_s \to \infty$. Once $\hat{R}$ dips below 1.1 for all sampled parameters, the $n_s n_r$ samples can together be considered samples from the posterior (2.13), [36]. Note that other values can also be chosen as a tolerance for $\hat{R}$, [8], but using 1.1 seems reasonable when accounting for additional numerical errors. We also note that this is not the only statistic used to determine chain convergence. From the resulting $n_s n_r$ samples of $\mathbf{f}$, $\boldsymbol{\alpha}$, and β, a variety of sample statistics can be computed which describe the joint posterior density as well as help to quantify the uncertainty in the data, which we describe in the next section.

3. Results. We now provide a real-world example that demonstrates the accuracy, efficiency, and robustness of the proposed method for SAR image reconstruction from phase history data. Note that the ground truth reflectivity image is unknown, preventing the computation of standard error statistics such as the relative error. This is the case even in synthetically-created SAR examples, where the true reflectivity is still unknown. Therefore, the uncertainty quantification information the proposed method provides is all the more valuable, as it is able to quantify how much we should trust pixel values and structures in the image even in the absence of ground truth. Throughout, all reflectivity images $\mathbf{f}$ are displayed in decibels (dB):

$$20\log_{10}\left(\frac{|\mathbf{f}|}{\max|\mathbf{f}|}\right), \tag{3.1}$$

with a minimum of −60 dB and maximum of 0 dB. Lesser or greater values are assigned the minimum or maximum. We begin with a specification of the data used in the image reconstruction example that follows.

3.1. Data. The GOTCHA Volumetric SAR Data Set 1.0 consists of SAR phase history data of a parking lot scene collected at X-band with a 640 MHz bandwidth with full azimuth coverage at 8 different elevation angles with full polarization, [18]. This is a real-world SAR dataset captured by the Air Force Research Laboratory. A plane carrying a sensor flew a roughly circular measurement flight around a parking lot near the Sensors Directorate Building at Wright-Patterson Air Force Base in Dayton, Ohio, and collected SAR phase history data. The parking lot contains various targets including civilian vehicles, construction vehicles, calibration targets, primitive reflectors, and military vehicles. Figure 2 shows optical images of the targets. Note that because this is real-world data, the elevation angle is not perfectly constant, and the path is not perfectly circular. The center frequency is 9.6GHz and bandwidth is 640MHz. This public release data has been used extensively for testing new SAR image reconstruction methods, [6, 5, 31, 56]. It is available from **https://www.sdms.afrl.af.mil/index.php?collection=gotcha**.

3.2. Computing Statistics from Samples. After running Algorithm 2.1, we obtain a group of samples of $\mathbf{f}$, $\boldsymbol{\alpha}$, and β, from the joint posterior density (2.13). Now we can use these samples to form statistics to summarize that complicated density. Perhaps the most obvious statistics to compute from the samples are the mean of $\mathbf{f}$, β, and $\boldsymbol{\alpha}$. For $\mathbf{f}$ and $\boldsymbol{\alpha}$, these will be images that can give information about objects and features and their locations within the image. For β, the mean will be scalar. Indeed there are many other ways to form estimates for these quantities, e.g. sorting the samples by pixel value and looking at the image formed by the median pixel value can also provide an estimate. Computing the variance or standard deviation of the samples can be useful in determining the range of possible values for each pixels, which can in turn be used to quantify uncertainty. In addition to the mean and variance, computing confidence intervals for each of the sampled parameters can aid in uncertainty quantification as well. Specifically, we sort the samples by pixel values from lowest to highest and form a confidence interval for each pixel. The interval between the 0.025 percentile pixel value and the 0.975 percentile pixel value represents a 95% confidence interval for the value of that parameter. In order to display this information, samples are drawn uniformly from this interval for each pixel and displayed in a GIF, called a confidence image, [49]. As there is not yet a seamless way to integrate videos into PDFs, here we simply display the lower and upper bounds. While not thoroughly explored in this paper, we anticipate that these samples and their confidence images can offer more information and answer downstream questions, e.g. about the support of the scene, [8].

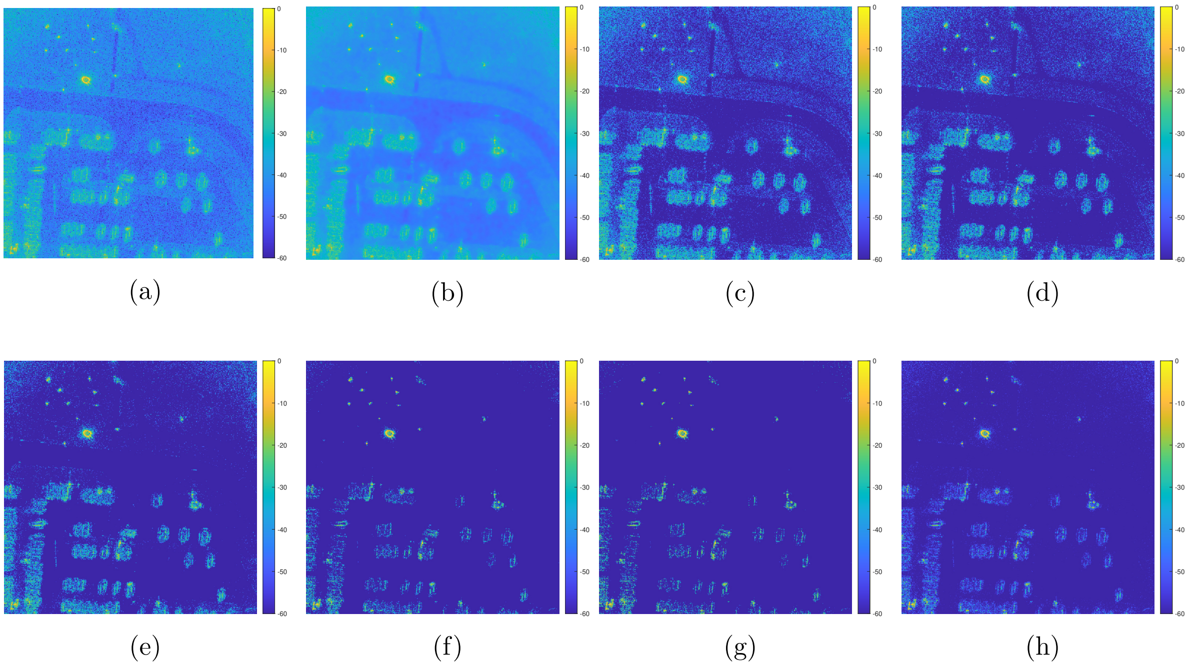


Figure 5: Full images formed with particulary sparsifying methods: (a) NUFFT; (b) TV regularization with $\lambda = 1/80$; (c-f) ℓ_1 regularization with $\lambda = 1/80, 1/60, 1/40, 1/20$; (g) SBL; (h) proposed method.

3.3. Example Estimates. Figure 5(h) shows the mean of the samples generated by Algorithm 2.1 for the GOTCHA parking lot scene, which is used as the image estimate for comparison purposes. Figure 5 compares full images of the GOTCHA parking lot scene using a NUFFT, TV regularization, ℓ_1 regularization, and the proposed method using the mean of the samples as an estimate. The full images shown are square with $N = 512^2$. Code from [55] was used to perform image formation for the comparison methods, as well as to wrangle the GOTCHA data. Figure 6 zooms in on two smaller subregions of the illuminated scene in order to see how each image formation method compares when localizing particular targets. The inverse NUFFT image corresponds to a maximum likelihood estimate, minimizing a least squares cost function. This does little to reduce speckle and noise and serves as a benchmark image. The ℓ_1 regularization scheme encourages sparsity (more zero values) in the estimate, yet it is evident that much of the speckle remains unless the regularization parameter λ is made so large that only a grainy image remains. It is indeed apparent the ℓ_1 method is differ-

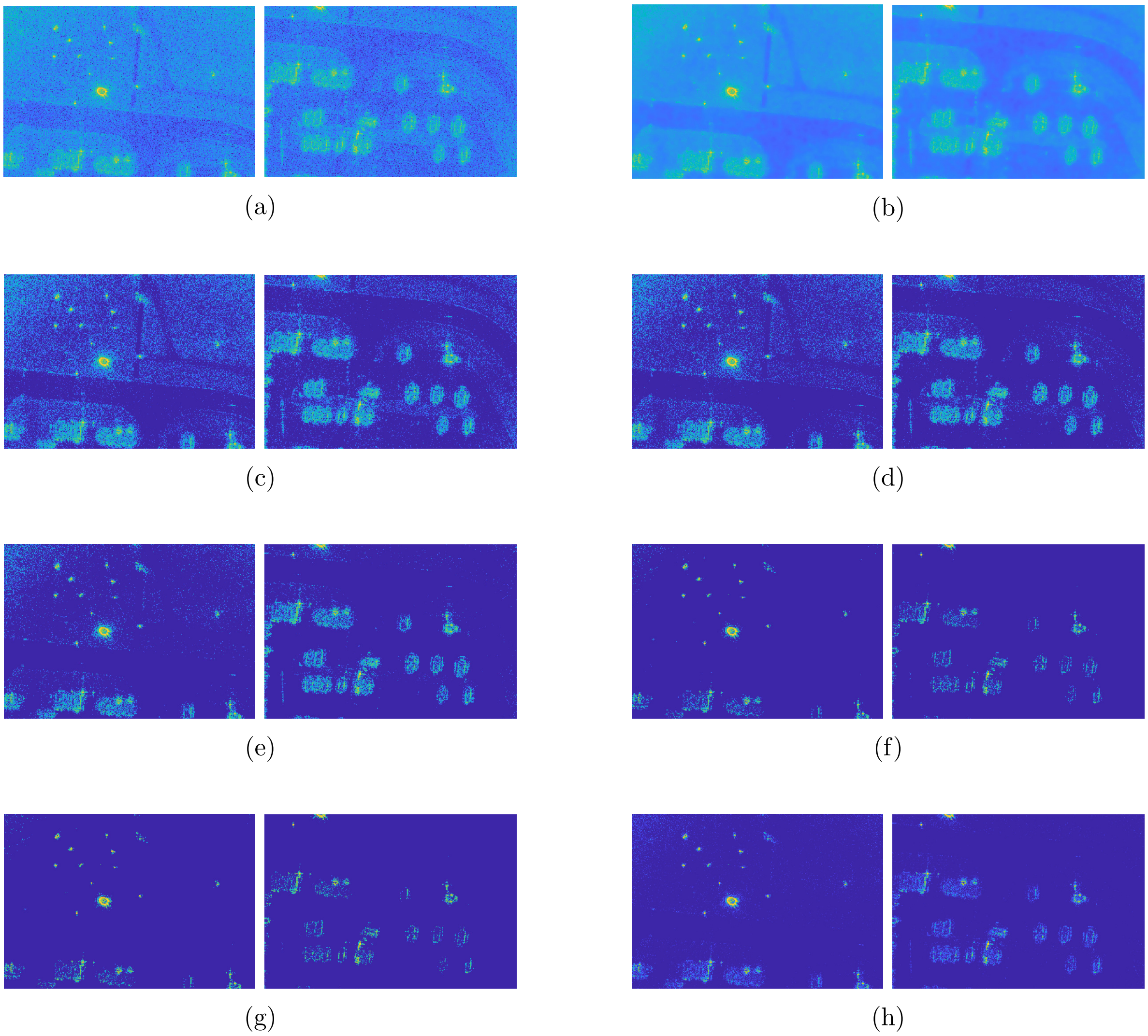


Figure 6: Two subregions of images from Fig. 5 formed with: (a) NUFFT; (b) TV regularization with $\lambda = 1/80$; (c-f) ℓ_1 regularization with $\lambda = 1/80, 1/60, 1/40, 1/20$; (g) SBL; (h) proposed method.

ent, regardless of the choice of λ, than the sampling method proposed here in its handling of speckle, which follows from its global penalty on magnitudes. The TV regularization removes much of the speckle, however it leaves block-like artifacts in its place. Recall that TV regularization is essentially an image denoising model – it aims to recover a piecewise constant image and also does not distinguish speckle from noise – which may explain the results. These comparison methods have been extensively applied in SAR. See, e.g., [2, 19, 40, 56, 58, 57]. The sampling-based method, which recall also uses sparsity-encouraging parameters, retrieves an estimate with noise and speckle reduction, as well as improved contrast, while maintaining visibly clear targets. There also appear to be no new artifacts, such as the block-like artifacts in the TV reconstruction. Note that these images are shown on the same color scale (MATLAB 'parula'), highlighting the contrast gains of the proposed method. One visible difference is the potential presence of roadside curbs in imagery from all methods aside from SBL and the proposed method. Importantly, while we have advantageously created an appropriate model for SAR imaging without any parameters to tune, this means that the learning algorithm decides which features are important. We hypothesize that the specific reason the curbs are left out is that the curbs produce anisotropic scattering, meaning that the reflection from all azimuth angles is not the same. Therefore, when the full azimuth (wide angle) aperture is used in this direct imaging, signal from these areas is not strong enough to withstand the more sparsifying prior used in both SBL and the proposed method. While for target identification, it may in fact be desirable not to clutter the scene images with returns from non-targets (e.g. grass and curbs), we recognize that in other applications this may be a critical drawback. In future work, we will focus on a composite approach which appropriately treats anisotropic scatterers by combining many small angle apertures, [56], rather than using a full azimuth as is done here.

	NUFFT	ℓ_1 regularization				TV regularization				SBL	Proposed
λ	N/A	1/20	1/40	1/60	1/80	1/40	1/80	1/120	1/160	N/A	N/A
Variance	51.28	0.59	7.42	31.00	51.28	0.91	1.39	2.32	3.65	0.63	0.59

Table 1: Variance for a small homogeneous subregion with each algorithm for various values of regularization parameter λ.

To quantify the improvement and speckle reduction, Table 1 shows the variance of each image in a small (50 pixel by 50 pixel) homogeneous region containing no targets to the left of the top hat reflector. This type of measurement is commonly used to evaluate speckle reduction, [3]. The ℓ_1 regularization method with $\lambda = 1/20$, the SBL algorithm, and the proposed method show the lowest variance, implying the best speckle reduction. However, the reduction from ℓ_1 regularization is not a targeted reduction as it simply comes from applying a global magnitude penalty. In addition, we see that the TV reconstructions for various λ also exhibit strong speckle reduction.

Table 2 gives the runtime for each algorithm. Each method was performed on Polaris, a shared memory computer operated by Dartmouth Research Computing with 40 cores, 64-bit Intel processors, and 1 TB of memory. Using such a large machine was necessary in order to store the samples (here $n_r n_s = 5 \cdot 1322$ for each of $2 \times 512^2 + 1$ parameters). While only images with $N = 512^2$ are shown throughout this paper, converged chains were computed for other values and the required chain lengths are shown in Table 3. In particular, Tables 2 and 3 show that convergence takes significantly more samples for larger images, and hence significantly more time. It is interesting to note that the required chain length appears to be roughly linear, although more examination is clearly needed.

Similar to $\mathbf{f}$, recall that the sampling-based image reconstruction method also produces

NUFFT	ℓ_1 regularization	Algorithm 2.1
.03s	5.8s	203s

Table 2: Runtimes for each algorithm with $N = 512^2$.

N	Algorithm 2.1
128^2	517
256^2	896
512^2	1322

Table 3: Required chain length n_s for various N.

samples of $\boldsymbol{\alpha}$, the parameter governing speckle, as well as β, the inverse noise variance, from the joint posterior. The mean of the $\boldsymbol{\alpha}$ samples is shown in Figure 7, while a histogram for the β samples is shown in Figure 8. Several observations can be made from these images. First, many features that were in the reflectivity image are also visible in this Figure 7. In particular, by comparing the results of our sampling method to the MAP estimate in (2.10), it is evident that, as desired, we predominantly regularize *away* from the large magnitude features, that is, presumably where there are no prominent targets. In addition to providing heuristics about the success of this algorithm through the lens of deterministic regularization, we also have that the reciprocal values of this image provide an estimate for the mean speckle parameter. Recall that the magnitude of each pixel $|\mathbf{f}_i|$ is Rayleigh distributed with mean proportional to $\boldsymbol{\alpha}_i^{-1}$, hence changes in the magnitude of each pixel $|\mathbf{f}_i|$ are proportional to $\boldsymbol{\alpha}_i^{-1}$. We see from Figure 7 that there is practically no speckle (most pixels are on the order of 10^{-14}) except at the various large magnitude target reflectivities, matching the speckle reduction we saw in Figures 5 and 6. In addition, this matches our earlier hypothesis that the validity of fully-developed speckle only for cells with no dominant scatterers is inconsequential. Indeed, the dominant scatterers remain and the speckle is reduced specifically in regions where there are no targets, i.e. where the fully-developed speckle model holds. This confirms that effectively using sparsity-encouraging measures will successfully reduce speckle.[12]

3.4. Visualizing Uncertainty Quantification. With the samples having been drawn, and estimates computed, we now seek to visualize uncertainty quantification information in order to inform the trustworthiness of these estimates. This additional information is intended to help human as well as potentially machine actors further interrogate a scene. Because this is an imaging application, any such useful information must be displayed in a visibly tractable way. We present several options below.

[12]Moreover, we anticipate that modifications to the model with a different specific intent would also be confirmed by evidence from the samples themselves.

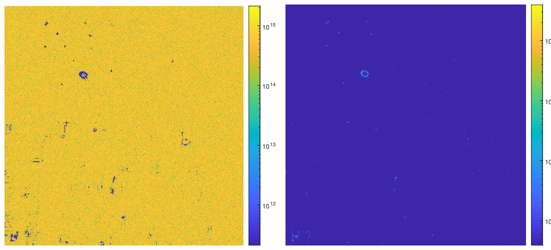


Figure 7: (left) Sample mean of $\boldsymbol{\alpha}$; (right) Sample mean of $\boldsymbol{\alpha}^{-1}$.

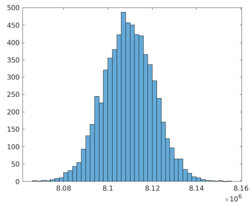


Figure 8: Histogram of samples of β. The sample mean is $\sim 8.11 \times 10^6$.

One way to quantify uncertainty in the above estimates, therefore fully benefiting from computing the entire joint posterior density, is to look at the sample variance (or sample standard deviation) at each pixel. This can be helpful in forming a confidence estimate by acknowledging that roughly 2 standard deviations from the mean contains 95% of samples in a Gaussian distribution. Figure 9 shows the sample variance of $\mathbf{f}$ for the example from Section 3.3. Notice that the variance is significantly lower for pixels of small magnitude. This is exactly what would be expected with multiple scales in a scene – large magnitude pixels tend to vary more than small magnitude pixels do. We can perform the same analysis for the $\boldsymbol{\alpha}$, and hence Figure 9 also shows the sample variance of the prior precision (or regularization matrix) $\boldsymbol{\alpha}$.

Another way to visualize uncertainty is to use confidence images, which can also provide visual information and insight into the uncertainties in the estimates, i.e. which features in the image can be trusted. Visualizing samples of a one-dimensional signal can be done using, e.g., confidence intervals with error bars on the mean estimate, trace plots represented as error bars at each point of the signal, or histograms. For example, the aforementioned histogram for samples of the one-dimensional β is shown in Figure 8. In many applications, a trace plot of the sample chain is used to show a cursory level of convergence. However, for two-dimensional images the visualization of the chain variance is less obvious. A tool to visualize 2D confidence images called Twinkle was developed in [49]. In Twinkle samples are sorted in increasing order and the 0.025 percentile value and the 0.975 percentile value are chosen as the lower and upper bounds for a 95% confidence interval at a particular pixel. Such an interval is computed for every pixel. Figure 10 shows the lower and upper bounds for the confidence images. We see that the particularly bright features occur in both the lower and upper confidence bounds, indicating relatively high confidence in these targets. Meanwhile away from the very bright

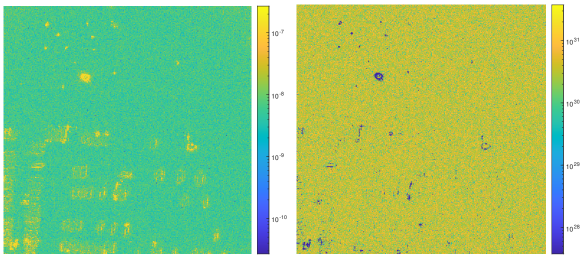


Figure 9: (left) Sample variance of $\mathbf{f}$; (right) Sample variance of $\boldsymbol{\alpha}$.

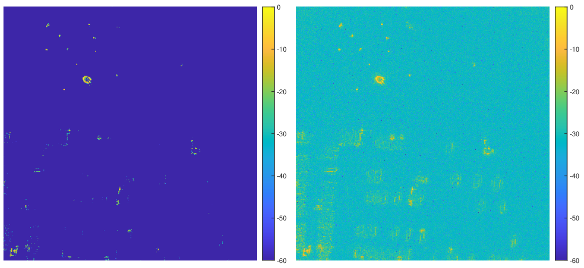


Figure 10: Comparison of the 95% confidence images for samples of $\mathbf{f}$. (left) 0.025 percentile image; (right) 0.975 percentile image.

targets, there is much more variation, indicating uncertainty. In Twinkle, new image samples are formed by drawing pixel values uniformly at random from within the confidence interval. A GIF or short movie can then be created from the image samples, showing them in quick succession for a fraction of a second each. The heuristic is that we can be more confident in features that persist in the image throughout the video, and less confident in features or pixel values in the image that flicker or twinkle. The latter could be an object of interest or attributable to an artifact or noise. In addition to Twinkle, another reasonable way to view this type of information is to simply display the posterior samples themselves in a GIF or short movie. Once again, similar analysis can be performed for $\boldsymbol{\alpha}$ (and $\boldsymbol{\alpha}^{-1}$), with similar conclusions drawn from Figure 11 as with the associated variance images for these quantities.

4. Conclusions.

In this paper we developed a procedure for sampling-based spotlight mode airborne SAR image reconstruction from phase history data. This task is challenging due to the problem size and the speckle phenomenon. Our framework uses a hierarchical Bayesian model with conjugate priors [60] to directly incorporate fully-developed speckle. A parameter-free sparsity-encouraging sampling method is introduced to provide estimates of the image, the speckle, and the noise directly from phase history data rather than through the processing of formed images. The GOTCHA data set example realizes this modeling, and demonstrates that our method reduces speckle and noise and improves contrast compared with other commonly used methods in real world problems. Uncertainty quantification information unavailable to other methods is also provided in the form of variance and confidence images, indicating when the pixel values and features shown in an estimate can be trusted. We

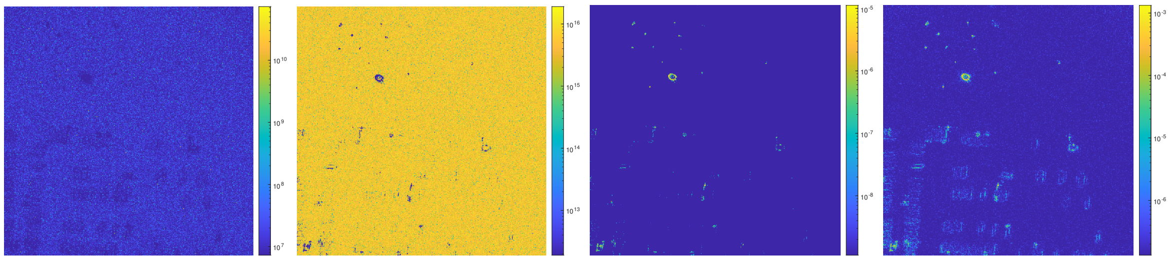


Figure 11: Comparison of the 95% confidence images for $\boldsymbol{\alpha}$: (left) 0.025 percentile image; (center left) 0.975 percentile image; and for $\boldsymbol{\alpha}^{-1}$: (center right) 0.025 percentile image; (right) 0.975 percentile image.

also quantify the uncertainty for the speckle and noise. Such information is of particular importance in SAR, where ground truth images even for synthetically-created phase history data sets are typically unknown.

Future work will focus on further accelerating the sampling method, as well as decreasing storage and memory requirements. This will enable image reconstruction with more pixels, as well as multi-pass and three-dimensional imaging. It will also allow composite image formation for wide angle SAR to complement the direct imaging results of this paper, for example addressing the issue of curbs addressed earlier. In addition, we hope to apply this sampling framework to other SAR modalities, as well as include coherent downstream processes such as interferometry and change detection.

REFERENCES

[1] F. Andersson, R. Moses, and F. Natterer, *Fast Fourier methods for synthetic aperture radar imaging*, IEEE Transactions on Aerospace and Electronic Systems, 48 (2012), pp. 215–229.

[2] R. Archibald, A. Gelb, and R. B. Platte, *Image reconstruction from undersampled Fourier data using the polynomial annihilation transform*, Journal of Scientific Computing, 67 (2016), pp. 432–452.

[3] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, *A tutorial on speckle reduction in synthetic aperture radar images*, IEEE Geoscience and Remote Sensing Magazine, 1 (2013), pp. 6–35.

[4] G. Aubert and J.-F. Aujol, *A variational approach to removing multiplicative noise*, SIAM Journal on Applied Mathematics, 68 (2008), pp. 925–946.

[5] C. D. Austin, *Sparse methods for model estimation with applications to radar imaging*, PhD thesis, The Ohio State University, 2012.

[6] C. D. Austin, E. Ertin, and R. L. Moses, *Sparse multipass 3D SAR imaging: Applications to the GOTCHA data set*, in Algorithms for Synthetic Aperture Radar Imagery XVI, vol. 7337, International Society for Optics and Photonics, 2009, p. 733703.

[7] J. Bardsley and T. Cui, *Optimization-based MCMC methods for nonlinear hierarchical statistical inverse problems*, arXiv preprint arXiv:2002.06358, (2020).

[8] J. M. Bardsley, *MCMC-based image reconstruction with uncertainty quantification*, SIAM Journal on Scientific Computing, 34 (2012), pp. A1316–A1332.

[9] J. M. Bardsley and T. Cui, *A Metropolis-Hastings-within-Gibbs sampler for nonlinear hierarchical-Bayesian inverse problems*, in 2017 MATRIX Annals, Springer, 2019, pp. 3–12.

[10] J. M. Bardsley, A. Seppanen, A. Solonen, H. Haario, and J. Kaipio, *Randomize-then-optimize for sampling and uncertainty quantification in electrical impedance tomography*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 1136–1158.

[11] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al., *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends® in Machine learning, 3 (2011), pp. 1–122.

[12] D. Calvetti, M. Pragliola, E. Somersalo, and A. Strang, *Sparse reconstructions from few noisy data: Analysis of hierarchical Bayesian models with generalized gamma hyperpriors*, Inverse Problems, (2019).

[13] D. Calvetti and E. Somersalo, *An introduction to Bayesian scientific computing: Ten lectures on subjective computing*, vol. 2, Springer Science & Business Media, 2007.

[14] D. Calvetti and E. Somersalo, *Hypermodels in the Bayesian imaging framework*, Inverse Problems, 24 (2008), p. 034013.

[15] D. Calvetti, E. Somersalo, and A. Strang, *Hierachical Bayesian models and sparsity: ℓ_2-magic*, Inverse Problems, 35 (2019), p. 035003.

[16] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory, 52 (2006), pp. 489–509.

[17] E. J. Candes, M. B. Wakin, and S. P. Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, Journal of Fourier Analysis and Applications, 14 (2008), pp. 877–905.

[18] C. H. Casteel Jr, L. A. Gorham, M. J. Minardi, S. M. Scarborough, K. D. Naidu, and U. K. Majumder, *A challenge problem for 2D/3D imaging of targets from a volumetric data set in an urban environment*, in Algorithms for Synthetic Aperture Radar Imagery XIV, vol. 6568, International Society for Optics and Photonics, 2007, p. 65680D.

[19] M. Çetin and W. C. Karl, *Feature-enhanced synthetic aperture radar image formation based on non-quadratic regularization*, IEEE Transactions on Image Processing, 10 (2001), pp. 623–631.

[20] R. Chartrand and W. Yin, *Iteratively reweighted algorithms for compressive sensing*, in IEEE International Conference on Acoustics, Speech and Signal Processing, 2008., IEEE, 2008, pp. 3869–3872.

[21] D.-Q. Chen and L.-Z. Cheng, *Spatially adapted total variation model to remove multiplicative noise*, IEEE Transactions on Image Processing, 21 (2012), pp. 1650–1662.

[22] G. Chierchia, M. El Gheche, G. Scarpa, and L. Verdoliva, *Multitemporal sar image despeckling based on block-matching and collaborative filtering*, IEEE Transactions on Geoscience and Remote Sensing, 55 (2017), pp. 5467–5480.

[23] V. Churchill, R. Archibald, and A. Gelb, *Edge-adaptive ℓ_2 regularization image reconstruction from non-uniform Fourier data*, Inverse Problems and Imaging, 13 (2019), pp. 931–958.

[24] V. Churchill and A. Gelb, *Detecting edges from non-uniform Fourier data via sparse Bayesian learning*, Journal of Scientific Computing, (2019), pp. 1–22.

[25] D. Cozzolino, S. Parrilli, G. Scarpa, G. Poggi, and L. Verdoliva, *Fast adaptive nonlocal sar despeckling*, IEEE Geoscience and Remote Sensing Letters, 11 (2013), pp. 524–528.

[26] A. Daoui, M. Yamni, H. Karmouni, M. Sayyouri, H. Qjidaa, et al., *Stable computation of higher order charlier moments for signal and image reconstruction*, Information Sciences, 521 (2020), pp. 251–276.

[27] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, *Iteratively reweighted least squares minimization for sparse recovery*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 63 (2010), pp. 1–38.

[28] G. Di Martino, M. Poderico, G. Poggi, D. Riccio, and L. Verdoliva, *Benchmarking framework for sar despeckling*, IEEE Transactions on geoscience and remote sensing, 52 (2013), pp. 1596–1615.

[29] X. Dong and Y. Zhang, *SAR image reconstruction from undersampled raw data using maximum a posteriori estimation*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8 (2014), pp. 1651–1664.

[30] H. Duan, L. Zhang, J. Fang, L. Huang, and H. Li, *Pattern-coupled sparse bayesian learning for inverse synthetic aperture radar imaging*, IEEE Signal Processing Letters, 22 (2015), pp. 1995–1999.

[31] E. Ertin, C. D. Austin, S. Sharma, R. L. Moses, and L. C. Potter, *GOTCHA experience report: Three-dimensional SAR imaging with complete circular apertures*, in Algorithms for Synthetic Aperture Radar Imagery XIV, vol. 6568, International Society for Optics and Photonics, 2007, p. 656802.

[32] H. G. Feichtinger, K. Gröchenig, and T. Strohmer, *Efficient numerical methods in non-uniform sampling theory*, Numer. Math., 69 (1995), pp. 423–440, https://doi.org/10.1007/s002110050101,

http://dx.doi.org/10.1007/s002110050101.
[33] J. A. Fessler and B. P. Sutton, *Nonuniform fast Fourier transforms using min-max interpolation*, IEEE Transactions on Signal Processing, 51 (2003), pp. 560–574.
[34] G. Franceschetti and R. Lanari, *Synthetic aperture radar processing*, CRC press, 1999.
[35] A. Gelb and G. Song, *A frame theoretric approach to the non-uniform fast Fourier transform*, Siam J. Numer. Anal., 52 (2014), pp. 1222–1242.
[36] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, Chapman and Hall/CRC, 2013.
[37] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (1984), pp. 721–741.
[38] G. H. Golub, P. C. Hansen, and D. P. O'Leary, *Tikhonov regularization and total least squares*, SIAM journal on matrix analysis and applications, 21 (1999), pp. 185–194.
[39] W. G. Goodman, R. S. Carrara, and R. M. Majewski, *Spotlight synthetic aperture radar signal processing algorithms*, Artech House, (1995), pp. 245–285.
[40] L. A. Gorham and L. J. Moore, *SAR image formation toolbox for MATLAB*, in Algorithms for Synthetic Aperture Radar Imagery XVII, vol. 7699, International Society for Optics and Photonics, 2010, p. 769906.
[41] L. Greengard and J.-Y. Lee, *Accelerating the nonuniform fast Fourier transform*, SIAM Review, 46 (2004), pp. 443–454.
[42] C. V. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-mode synthetic aperture radar: A signal processing approach*, Springer Science & Business Media, 2012.
[43] S. Ji, Y. Xue, and L. Carin, *Bayesian compressive sensing*, IEEE Transactions on Signal Processing, 56 (2008), pp. 2346–2356.
[44] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, vol. 160, Springer Science & Business Media, 2006.
[45] O. Karakuş, E. E. Kuruoğlu, and M. A. Altinkaya, *Generalized bayesian model selection for speckle on remote sensing images*, IEEE Transactions on Image Processing, 28 (2018), pp. 1748–1758.
[46] J.-S. Lee, L. Jurkevich, P. Dewaele, P. Wambacq, and A. Oosterlinck, *Speckle filtering of synthetic aperture radar images: A review*, Remote sensing reviews, 8 (1994), pp. 313–340.
[47] J.-Y. Lee and L. Greengard, *The type 3 nonuniform FFT and its applications*, Journal of Computational Physics, 206 (2005), pp. 1–5.
[48] L. J. Moore, B. D. Rigling, and R. P. Penno, *Characterization of phase information of synthetic aperture radar imagery*, IEEE Transactions on Aerospace and Electronic Systems, 55 (2018), pp. 676–688.
[49] J. G. Nagy and D. P. O'Leary, *Image restoration through subimages and confidence images*, Electronic Transactions on Numerical Analysis, 13 (2002), pp. 22–37.
[50] R. M. Neal, *Bayesian learning for neural networks*, vol. 118, Springer Science & Business Media, 2012.
[51] G. Newstadt, E. Zelnio, and A. Hero, *Moving target inference with bayesian models in sar imagery*, IEEE Transactions on Aerospace and Electronic Systems, 50 (2014), pp. 2004–2018.
[52] C. Oliver and S. Quegan, *Understanding synthetic aperture radar images*, SciTech Publishing, 2004.
[53] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, *A nonlocal sar image denoising algorithm based on llmmse wavelet shrinkage*, IEEE Transactions on Geoscience and Remote Sensing, 50 (2011), pp. 606–616.
[54] D. Potts, G. Steidl, and M. Tasche, *Fast Fourier transforms for nonequispaced data: a tutorial*, in Modern sampling theory, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Boston, MA, 2001, pp. 247–270.
[55] T. Sanders, *MATLAB imaging algorithms: Image reconstruction, restoration, and alignment, with a focus in tomography*. http://www.toby-sanders.com/software, https://doi.org/10.13140/RG.2.2.33492.60801, https://doi.org/10.13140/RG.2.2.33492.60801. Accessed: 2016-19-08.
[56] T. Sanders, A. Gelb, and R. B. Platte, *Composite SAR imaging using sequential joint sparsity*, Journal of Computational Physics, 338 (2017), pp. 357–370.
[57] T. Scarnati, *Recent Techniques for Regularization in Partial Differential Equations and Imaging*, Arizona State University, 2018.
[58] T. Scarnati and A. Gelb, *Joint image formation and two-dimensional autofocusing for synthetic aper-*

ture radar data, Journal of Computational Physics, 374 (2018), pp. 803–821.
[59] H. Sedarat and D. G. Nishimura, *On the optimality of the gridding reconstruction algorithm*, IEEE Trans. Med. Imaging, 19 (2000), pp. 306–317.
[60] M. E. Tipping, *Sparse Bayesian learning and the relevance vector machine*, Journal of Machine Learning Research, 1 (2001), pp. 211–244.
[61] J. Wu, F. Liu, L. Jiao, and X. Wang, *Compressive sensing sar image reconstruction based on bayesian framework and evolutionary computation*, IEEE Transactions on Image Processing, 20 (2011), pp. 1904–1911.
[62] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, *High-resolution passive sar imaging exploiting structured bayesian compressive sensing*, IEEE Journal of Selected Topics in Signal Processing, 9 (2015), pp. 1484–1497.
[63] J.-p. Xu, Y.-n. Pi, and Z.-j. Cao, *Bayesian compressive sensing in synthetic aperture radar imaging*, IET Radar, Sonar & Navigation, 6 (2012), pp. 2–8.
[64] M. Xue, E. Santiago, M. Sedehi, X. Tan, and J. Li, *Sar imaging via iterative adaptive approach and sparse bayesian learning*, in Algorithms for Synthetic Aperture Radar Imagery XVI, vol. 7337, International Society for Optics and Photonics, 2009, p. 733706.
[65] X. Zhang, J. Qin, and G. Li, *Sar target classification using bayesian compressive sensing with scattering centers features*, Progress In Electromagnetics Research, 136 (2013), pp. 385–407.