Radio Science

RESEARCH ARTICLE
10.1029/2019RS007011

Key Points:
- Dispersive targets in synthetic aperture radar imaging can be detected by adding a scattering delay variable to signal processing
- Classification of targets into instantaneous versus delayed (dispersive) categories is hampered by the effects of speckle and noise
- Model-based statistical divergence measures can predict the misclassification rate without computationally heavy Monte Carlo simulations

Correspondence to:
M. Gilman,
mgilman@ncsu.edu

Citation:
Gilman, M., & Tsynkov, S. (2021). Divergence measures and detection performance for dispersive targets in SAR. *Radio Science*, *56*, e2019RS007011. https://doi.org/10.1029/2019RS007011

Received 11 OCT 2019
Accepted 19 NOV 2020
Accepted article online 8 DEC 2020

Divergence Measures and Detection Performance for Dispersive Targets in SAR

Mikhail Gilman[1] **and Semyon Tsynkov**[1]

[1]Department of Mathematics, North Carolina State University at Raleigh, Raleigh, NC, USA

Abstract When electromagnetic waves impinge on objects with complex geometries and/or internal structure, we can observe scattering that is distributed in time rather than instantaneous. To detect and characterize such targets, we build the coordinate-delay synthetic aperture radar (cdSAR) images by adding a delay term to the standard SAR matched filter. In order to apply this approach to the case of extended targets where the image intensity and phase are subject to strong and rapid variations (the phenomenon called speckle), we sample the cdSAR image at several coordinate-delay "points" in the vicinity of the scatterer location. The discrimination between the instantaneous and delayed targets is realized through autocorrelation analysis of this sample. Because of the statistical properties of speckle, misclassification errors are inevitable. Hence, prediction of the error rate as a function of system and target parameters becomes an important problem. While Monte Carlo simulations can generate the ensembles of data for direct calculation of the error rate, this approach is computationally demanding because of its slow convergence. In order to simplify the prediction of the error rate, we employ statistical divergence measures, namely, the Hellinger distance and Kullback-Leibler divergence. These divergence measures are calculated directly from the theoretical models of reflectivity of extended targets that we want to distinguish. We empirically establish a linear relation between the misclassification rate and the Hellinger distance for a certain class of simple target models. This relation allows us to make predictions of the error rate without performing the Monte Carlo simulations.

1. Introduction

Synthetic aperture radar (SAR) is an imaging technology that uses illumination of the target by a series of coherent electromagnetic (EM) pulses. Imaging due to coherent sources results in speckle (Goodman, 1984), which can be described as sharp noise-like variations of the intensity and phase of the image. These variations are uncorrelated on a pixel scale even when the essential parameters of the scatterer vary on a much bigger scale. For this reason, the latter variations are hard to detect, and even with the most advanced despeckling and segmentation methods (Lee & Pottier, 2009; Oliver & Quegan, 1998), errors in classification and/or parameter estimation are inevitable.

This paper considers the detection of dispersive targets, that is, targets with scattering delay. The total delay registered by the radar system, that is, the time interval between when the signal is emitted and response received, is the sum of a propagation delay and scattering delay (if any). The presence of scattering delay may be an indication of a complex internal geometry or structure of the scatterer, such as cavities or regular inhomogeneities. Depending on the context, this can provide a crucial piece of information for the classification of the scene and/or object; for example, in an urban environment, a delayed scatterer surrounded by a paved area may mean an open manhole. Physically, dispersive targets don't have to be distributed, that is, occupy more than one pixel, because the aforementioned cavities or inhomogeneities can be realized on a sub-resolution scale.

The main difficulty in the detection of scattering delay in SAR (Gilman & Tsynkov, 2019) is to separate it from the propagation delay, which is at the core of the SAR signal processing paradigm. There is a principal possibility to make this separation if the synthetic aperture is wide enough (Ferrara et al., 2017; Gilman & Tsynkov, 2019). However, the difference between these two types of delay in SAR images is typically small and manifests itself gradually. Hence, as a necessary part of any study involving dispersive targets, the detectability of scattering delay in the presence of speckle has to be evaluated.

A discrimination procedure between the instantaneous and dispersive targets in the presence of noise and speckle has been introduced in Gilman and Tsynkov (2019). The algorithm maximizes the probability density of the observed data given different parametrized models of the target. It was determined that even for a very simple model of the dispersive target, the discrimination quality depends on many parameters, such as the bandwidth, angular aperture width, maximum scattering delay, and target contrast. In order to assign a numerical characteristic to the discrimination quality, the Monte Carlo simulation was used in Gilman and Tsynkov (2019) to generate ensembles of images due to instantaneous and dispersive targets. The discrimination has been performed for each image in these ensembles, and the misclassification frequency was subsequently calculated. This frequency characterizes the quality of discrimination. The results of the numerical experiments described in Gilman and Tsynkov (2019) are in a good agreement with how one would intuitively expect the discrimination quality to behave as a function of the system and target parameters. The only unexpected observation was related to the so called homogeneous, or background, samples, and this issue has received some clarification in this work, as presented in sections 4 and 5.

The disadvantage of the Monte Carlo approach is its high computational cost. Each individual discrimination involves solution of optimization problems over a multidimensional space of parameters that characterize the target. At the same time, the variance (i.e., mean square error) of the computed misclassification frequency decreases as slowly as the reciprocal size of the ensemble. Therefore, studying the dependence of the discrimination quality on problem parameters becomes computationally expensive.

This work provides a computationally efficient alternative to the Monte Carlo method for computing the misclassification frequency. In particular, we employ *statistical divergence measures as proxies of the classification quality*. A divergence measure deals with the probability density function of a sampled image rather than its individual realizations. Hence, we can bypass the Monte Carlo step and compute the divergence measure, and thus predict the misclassification frequency, directly from the target models.

We have conducted comprehensive numerical simulations to demonstrate the viability of our approach. We used two popular divergence measures: the Hellinger distance and Kullback-Leibler divergence (Pardo, 2006; Pollard, 2002). The former proved especially convenient because there appears to be a simple linear relation between the misclassification frequency and Hellinger distance. The main finding is that the variation of system and target parameters has only a very small effect on the relation between the divergence measure and misclassification frequency. This means that the error in predicting the misclassification frequency from a divergence measure will be small. In our simulations, it has never exceeded a few percent.

These observations allow us to study the performance of the classifier analytically. As a demonstration of the capabilities of this tool, we corroborate the effect observed earlier, namely, low sensitivity of misclassification rate to the so-called background samples (Gilman & Tsynkov, 2019). It is remarkable that despite their rather different formulations, the conclusions resulting from the analyses based on either the Hellinger distance or Kullback-Leibler divergence are identical.

Previous applications of divergence measures in the context of SAR include change detection (Inglada & Mercier, 2007) and segmentation assistance in despeckling (Torres et al., 2014). An assessment of the efficiency of change detection using various divergence measures can be found, for example, in Nascimento et al. (2010). However, we are not aware of any publication that uses divergence measures for the analysis of dispersive targets. A very good overview of the literature on dispersive targets in SAR can be found in Ferrara et al. (2017). The fundamentals of SAR imaging are presented in many excellent books, including Cheney and Borden (2009), Cumming and Wong (2005), and Jakowatz et al. (1996). Theoretical work on speckle is covered in Goodman (1984) and Oliver and Quegan (1998).

This article is organized as follows. In section 2, we give an overview of the approach to the detection of delayed return in SAR in the presence of speckle of noise that was developed in Gilman and Tsynkov (2019). In particular, section 2.1 presents the deterministic approach to dealing with the scattering delay; section 2.2 introduces speckle and describes the simple statistical models of instantaneous and delayed scatterers that are used to set up the discrimination problem; section 2.3 provides the relations between the second moments of the image and the parameters of the model; section 2.4 presents the discrimination algorithm; and section 2.5 defines the misclassification frequency. The rationale for using statistical divergence measures to predict the performance of the classifier is given in section 3. In section 4, we obtain an empirical relation between the Hellinger distance and the misclassification frequency for the classifier presented in

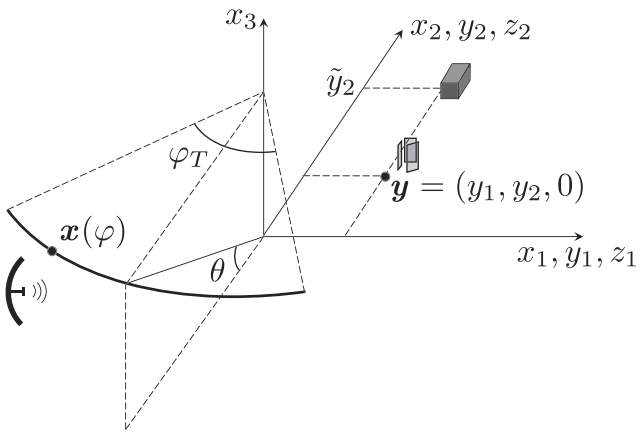

Figure 1. Schematic of SAR imaging geometry. The antenna trajectory is an arc of a circle with angular extent φ_T at some constant elevation above the ground. A group of reflectors in the neighborhood of $\mathbf{y} = (y_1, y_2, 0)$ with sub-resolution spacing may exhibit a delayed return due to a complex scattering geometry. This delayed return can overlap with an immediate return from an object near $\tilde{\mathbf{y}} = (y_1, \tilde{y}_2, 0)$ with $\tilde{y}_2 > y_2$, see (9). When the delay exceeds $\Delta_R/(2c)$, where Δ_R is the range resolution, see (6), the points $\mathbf{y}$ and $\tilde{\mathbf{y}}$ will correspond to different pixels. Note that angles, distances, and sizes in this illustration are not to scale; in particular, the dimensions of the displayed objects (both vertical and horizontal) and their respective distances to the pixel centers $\mathbf{y}$ and $\tilde{\mathbf{y}}$ are assumed to be much smaller than the resolution size, the size of the scene is much smaller than the distance to the antenna (e.g., $|\mathbf{y}| \sim |\tilde{\mathbf{y}}| \ll |\mathbf{x}|$), and the aperture is narrow: $\varphi_T \ll 1$.

section 2.4. Then, using this relation, we study analytically the effect of background samples on the performance of the classifier. In section 5, we compare and contrast two divergence measures, the Kullback-Leibler divergence and the Hellinger distance, reproducing some of the results of section 4 using the former. The summary of the work is given in section 6.

2. Detection of Scattering Delay in SAR

In this section, we briefly review the approach to the detection of scattering delay in SAR that was developed in Gilman and Tsynkov (2019).

2.1. Coordinate-Delay SAR Image and Point Spread Function

The goal of SAR imaging is to build an approximate reconstruction of the reflectivity function of the target. In the standard SAR, the reflectivity, which is assumed instantaneous, is a function of the coordinates: $\nu_{\text{inst}}(\mathbf{z})$, where $\mathbf{z} = (z_1, z_2, 0)$; see Figure 1. This function is unknown, and the reconstruction is provided by a convolution of the received signal with the matched filter (Cheney & Borden, 2009; Cumming & Wong, 2005; Gilman et al., 2017). In such a filter, the propagation delay is presented in the form $2|\mathbf{x}-\mathbf{y}|/c$, where c is the speed of light, $\mathbf{x} = \mathbf{x}(\varphi)$ is a point on the antenna trajectory parametrized by the look angle φ, and $\mathbf{y} = (y_1, y_2, 0)$ is the image coordinate (henceforth, we will ignore the third coordinate of $\mathbf{y}$ and $\mathbf{z}$). The resulting standard SAR image $I_{\text{std}}(\mathbf{y})$ is an approximate reconstruction of the instantaneous reflectivity function $\nu_{\text{inst}}(\mathbf{z})$. If such an image exhibits features that could be attributed to the presence of a delayed scatterer (e.g., the streaks in the range direction, as mentioned in section 2.2), then a different kind of formulation for reflectivity, as well as signal processing, is warranted.

In particular, we define the new reflectivity ν as a function of not only the coordinates but also the scattering delay $t_{\mathbf{z}}$: $\nu = \nu(t_{\mathbf{z}}, \mathbf{z})$, and the corresponding scattering delay $t_{\mathbf{y}}$ is added to the propagation delay $2|\mathbf{x}-\mathbf{y}|/c$ in the matched filter (Ferrara et al., 2017; Gilman & Tsynkov, 2019). The resulting coordinate-delay SAR (cdSAR) image $I(t_{\mathbf{y}}, \mathbf{y})$ is an approximate reconstruction of $\nu(t_{\mathbf{z}}, \mathbf{z})$ and is related to the latter by convolution with the imaging kernel W:

$$I(t_{\mathbf{y}}, \mathbf{y}) = \int_0^\infty dt_{\mathbf{z}} \int d\mathbf{z}\, \nu(t_{\mathbf{z}}, \mathbf{z}) W(t_{\mathbf{y}}, \mathbf{y}; t_{\mathbf{z}}, \mathbf{z}). \tag{1}$$

The imaging kernel $W = W(t_{\mathbf{y}}, \mathbf{y}; t_{\mathbf{z}}, \mathbf{z})$ is obtained by applying the chosen signal processing algorithm to the received field that is modeled using the appropriate EM scattering and propagation techniques.

The properties of the imaging kernel define the quality of reconstruction. For the formulation in (1), the ideal kernel is $W \sim \delta(t_{\mathbf{y}} - t_{\mathbf{z}}, \mathbf{y} - \mathbf{z})$. Due to technical limitations though, it cannot be achieved. The expression for the imaging kernel derived in Gilman and Tsynkov (2019) depends on the properties of the imaging system and observation conditions. We present it below using the same notations as in the original publication; in particular, we assume that ω_0 is the carrier frequency of the radar pulse, B is the bandwidth, φ_T is the angular extent of the synthetic aperture (i.e., $|\varphi| \leqslant \varphi_T/2$), θ is the incidence angle, and the lower indices 1 and 2 denote the cross-range and range coordinates, respectively; see Figure 1. Note that we assume a narrow-angle aperture, that is, $\varphi_T \ll 1$, to be able to ignore the dependence of v on the antenna position; this is a typical part of a SAR setup. We will use the following definition of the sinc function: $\operatorname{sinc} \xi \stackrel{\text{def}}{=} \frac{\sin \xi}{\xi}$. Additionally, we define two parameters:

$$k_{0\theta} \stackrel{\text{def}}{=} \frac{\omega_0}{c} \sin\theta, \;\; T^0 = \frac{y_2 - z_2}{c} \sin\theta + \frac{t_{\mathbf{y}} - t_{\mathbf{z}}}{2}. \tag{2}$$

Physically, $k_{0\theta}$ is the projection of the incident wave vector on the horizontal plane, and T^0 can be understood as the difference of one-way travel times for the following pair of coordinate-delay point scatterers: $v(t'_{\mathbf{y}}, \mathbf{y}') = \delta(t'_{\mathbf{y}} - t_{\mathbf{y}})\delta(\mathbf{y}' - \mathbf{y})$ and $v(t'_{\mathbf{y}}, \mathbf{y}') = \delta(t'_{\mathbf{y}} - t_{\mathbf{z}})\delta(\mathbf{y}' - \mathbf{z})$, where the antenna is at the center of the synthetic aperture. Using a function of two arguments,

$$\Phi(v_1, v_2) \stackrel{\text{def}}{=} \int_{-1/2}^{1/2} e^{2iv_1 s} e^{iv_2 s^2} ds, \tag{3}$$

we obtain, accurate to an inessential constant factor, the imaging kernel as a product of one fast phase factor and two slow amplitude factors:

$$W(t_{\mathbf{y}}, \mathbf{y}; t_{\mathbf{z}}, \mathbf{z}) = e^{-2i\omega_0 T^0} \cdot \operatorname{sinc}(BT^0) \cdot \Phi\left(k_{0\theta}\varphi_T(y_1 - z_1), k_{0\theta}\varphi_T^2(y_2 - z_2)\right). \tag{4}$$

Transition from (1)–(4) to standard SAR can be made by taking $t_{\mathbf{y}} = 0$ and assuming that the reflectivity is instantaneous, that is, $v(t_{\mathbf{z}}, \mathbf{z}) = v_{\text{inst}}(\mathbf{z})\delta(t_{\mathbf{z}})$. If we additionally assume that φ_T is small so that the second argument of Φ in (4) vanishes, then we can replace $\Phi(v_1, 0)$ with $\operatorname{sinc} v_1$ and reduce (1) and (4) to the well-known form (Cheney & Borden, 2009; Cumming & Wong, 2005; Gilman et al., 2017):

$$\begin{aligned} I_{\text{std}}(\mathbf{y}) \stackrel{\text{def}}{=} I(0, \mathbf{y}) &= \int d\mathbf{z}\, v_{\text{inst}}(\mathbf{z}) W_{\text{std}}(\mathbf{y}, \mathbf{z}), \;\; \text{where} \\ W_{\text{std}}(\mathbf{y}, \mathbf{z}) &= e^{-2ik_{0\theta}(y_2 - z_2)} \cdot \operatorname{sinc}\left(\frac{B}{\omega_0} k_{0\theta}(y_2 - z_2)\right) \cdot \operatorname{sinc}\left(k_{0\theta}\varphi_T(y_1 - z_1)\right). \end{aligned} \tag{5}$$

In particular, the sinc terms in (5) provide standard expressions for the range and cross-range resolution of a SAR system:

$$\Delta_{\text{R}} = \pi \frac{c}{B \sin\theta}, \;\; \Delta_{\text{A}} = \pi \frac{1}{k_{0\theta}\varphi_T}. \tag{6}$$

The behavior of W along the direction defined by

$$T^0 = \text{const and } y_1 - z_1 = \text{const} \tag{7}$$

reveals what is called the range-delay ambiguity. If we assume a (infinitesimally) small aperture as in (5), that is, replace the term $\Phi(\cdot, \cdot)$ in (4) with $\operatorname{sinc}\left(k_{0\theta}\varphi_T(y_1 - z_1)\right)$, then the dependence of W on the range coordinates y_2 and z_2 is solely via T^0 of (2):

$$W_{\text{ambig}}(t_{\mathbf{y}}, \mathbf{y}; t_{\mathbf{z}}, \mathbf{z}) = e^{-2i\omega_0 T^0} \cdot \operatorname{sinc}(BT^0) \cdot \operatorname{sinc}\left(k_{0\theta}\varphi_T(y_1 - z_1)\right). \tag{8}$$

If W_{ambig} is used instead of W in (1), then the effects of scattering delay and range shift in the resulting image I_{ambig} become equivalent, that is, for any $v(t_{\mathbf{z}}, \mathbf{z})$ and any $t_{\mathbf{y}}$ the following relation holds:

$$I_{\text{ambig}}(t_{\mathbf{y}}, \mathbf{y}) = I_{\text{ambig}}(0, \tilde{\mathbf{y}}) \text{ for } \tilde{y}_1 = y_1, \;\; \tilde{y}_2 = y_2 + \frac{1}{\sin\theta} \frac{ct_{\mathbf{y}}}{2}. \tag{9}$$

Hence, when the imaging kernel is given by (8), the discrimination between the delayed and instantaneous return should be considered impossible. In other words, the use of the cdSAR signal processing algorithm that yields $I(t_{\mathbf{y}}, \mathbf{y})$ given by (1) provides no new information as compared to the standard SAR processing that results in $I_{\text{std}}(\mathbf{y})$ of (5).

For finite φ_T, equality (9) no longer holds; however, the "resolution" along the direction (7) is impaired by a slow decay of $|\Phi(\nu_1, \nu_2)|$ with $|\nu_2|$, for example, $|\Phi(0, \nu_2)| \sim (|\nu_2|/\pi)^{-1/2}$ as $|\nu_2| \to \infty$. Hence, the separation between the values of $I(t_{\mathbf{y}}, \mathbf{y})$ and $I(0, \tilde{\mathbf{y}})$ (cf. (9)) may be small, and the effects of noise and speckle on our ability to make the above discrimination cannot be ignored.

Note that the misclassification of instantaneous and delayed targets due to the range-delay ambiguity is not directly related to the simplified EM model of scattering employed by SAR. As stated in the beginning of section 2, the standard SAR is built upon the model of a target as a configuration of point scatterers located on a horizontal plane and treated using the first Born approximation. At the same time, nonlinear EM models, for example, the second-order Born approximation presented in Pierri et al. (2002), as well as more advanced methods, such as the linear sampling method (Cakoni et al., 2011) and the factorization method (Kirsch & Grinberg, 2008), can be used for reconstruction of the shape and/or dielectric coefficient of the target. However, the limitations of resolution in SAR imaging are fundamental and related to the system parameters, such as the bandwidth and angular width of the synthetic aperture. If the small details of the scatterer shape responsible for the scattering dispersion (as schematically illustrated in Figure 1) cannot be resolved, then the simple phenomenological model of reflectivity as a function of coordinates and delay, $\nu(t_{\mathbf{z}}, \mathbf{z})$, as well as the treatment of problem as linear, are appropriate.

2.2. Simple Statistical Models of Instantaneous and Dispersive Scatterers

Once the support of ν becomes nonsingular in any argument, a statistical approach appears to be the only feasible way of describing the resulting radar images. Arguably, no reasonable deterministic expression for $\nu(t_{\mathbf{z}}, \mathbf{z})$ or $\nu_{\text{inst}}(\mathbf{z})$ in the convolution formulae (1) and (5) can explain the abrupt variations of the intensity and phase of the observed images, the effect called speckle. Imaging systems suffer from speckle when they employ coherent illumination sources (Gilman & Tsynkov, 2019; Goodman, 1984; Oliver & Quegan, 1998), which is the case for SAR.

In particular, the following is a model of reflectivity of a homogeneous instantaneous scatterer, also called background or clutter, with statistically averaged reflectivity σ_{b}^2 (hence, σ_{b}^2 is a non-negative deterministic constant):

$$\nu_{\text{b}}(t_{\mathbf{z}}, \mathbf{z}) = \delta(t_{\mathbf{z}})\mu_{\text{b}}(\mathbf{z}), \tag{10}$$

(see Oliver & Quegan, 1998) where $\mu_{\text{b}}(\mathbf{z})$ is a two-dimensional circular Gaussian white random field (Gallager, 2008a, 2008b; Lapidoth, 2017):

$$\langle \mu_{\text{b}}(\mathbf{z}) \rangle = 0, \ \left\langle \mu_{\text{b}}(\mathbf{z})\mu_{\text{b}}(\mathbf{z}') \right\rangle = 0, \ \left\langle \overline{\mu_{\text{b}}(\mathbf{z})}\mu_{\text{b}}(\mathbf{z}') \right\rangle = \sigma_{\text{b}}^2 \delta(\mathbf{z} - \mathbf{z}'). \tag{11}$$

In (11), $\langle \dots \rangle$ denotes statistical averaging. This is a model for the so-called fully developed speckle where no additive deterministic component is present (see, e.g., Pedersen, 1974).

In Gilman and Tsynkov (2019), the models (10) and (11) are generalized to design two reflectivity models for inhomogeneous scatterers: ν_t represents a delayed point scatterer, while ν_s is an extended instantaneous scatterer (a linear inhomogeneity of the background). These additional models are characterized by the averaged reflectivities $\sigma_t^2 \geq 0$ and $\sigma_s^2 \geq 0$, respectively. Besides, the reference location $\mathbf{z}_{\text{d}} = (z_{\text{d}1}, z_{\text{d}2})$ and the reflectivity profile $F(\zeta)$, which is a dimensionless non-negative function of a dimensionless argument, are shared by the two models. The expressions for the reflectivity functions are

$$\nu_t(t_{\mathbf{z}}, \mathbf{z}) \equiv \nu_t(t_{\mathbf{z}}, \mathbf{z}; \mathbf{z}_{\text{d}}) = \mu_t \left(\frac{Bt_{\mathbf{z}}}{2} \right) \delta(\mathbf{z} - \mathbf{z}_{\text{d}}) \tag{12a}$$

and

$$\nu_s(t_{\mathbf{z}}, \mathbf{z}) \equiv \nu_s(t_{\mathbf{z}}, \mathbf{z}; \mathbf{z}_{\text{d}}) = \mu_s \left(\frac{Bk_{0\theta}(z_2 - z_{\text{d}2})}{\omega_0} \right) \delta(t_{\mathbf{z}})\delta(z_1 - z_{\text{d}1}), \tag{12b}$$

where $\mu_s(\zeta)$ and $\mu_t(\zeta)$ are inhomogeneous one-dimensional circular Gaussian white random processes described by

$$\left\langle \overline{\mu_s(\zeta)}\mu_s(\zeta') \right\rangle = \sigma_s^2 F(\zeta)\delta(\zeta - \zeta'),$$
$$\left\langle \overline{\mu_t(\zeta)}\mu_t(\zeta') \right\rangle = \sigma_t^2 F(\zeta)\delta(\zeta - \zeta').$$

The models (12a) and (12b) are designed specifically to set up the problem of discrimination between the instantaneous and delayed scatterer in the simplest yet physically meaningful form. In particular, the inessential azimuthal dimension is eliminated from both models in (12) with the help of the respective delta functions (this means that the corresponding scales are not resolved). The remaining two arguments of the reflectivity function, that is, one spatial coordinate (range) and time delay, are subject to range-delay ambiguity, as indicated by (9). Each of the two models in (12) is localized in one of these arguments, but not the other. As a result, when the range-delay ambiguity is not resolved, as in (9), it is impossible to tell apart the additional propagation delay due to model (12b) and the "internal" delay due to model (12a), so the images due to the scatterers ν_t and ν_s are identical provided $\sigma_t^2 = \sigma_s^2$. For example, in the "instantaneous slice" of a cdSAR image, that is, $I(0, \mathbf{y})$, both scatterers yield a linear inhomogeneity, also called a streak, extending from $\mathbf{y} = \mathbf{z}_\mathrm{d}$ to $\mathbf{y} = \left(\mathbf{z}_\mathrm{d} + \zeta_\mathrm{max}\frac{\omega_0}{Bk_{0\theta}}\mathbf{e}_2\right)$, where $\mathbf{e}_2$ is a unit vector in the downrange direction and ζ_max characterizes either the maximum delay or the extent of inhomogeneity in the range direction, depending on the type of the scatterer in (12). Such streaks have been observed in inverse SAR, standard SAR, and simulated SAR images; see, for example, Borden (1998), Ferrara et al. (2017), and Trintinalia and Ling (1997), respectively. Note that the interpretation of the reference position $\mathbf{z}_\mathrm{d}$ depends on the assumed model: For (12a), it means the location of a delayed point scatterer, whereas for (12b), it is the nearest end of a linearly shaped inhomogeneity of instantaneous reflectivity.

The form of $F(\zeta)$ will characterize the scatterer: We require $0 \leqslant F \leqslant 1$, and $F(\zeta) = 0$ for $\zeta < 0$ to account for the causality of scattering (see also (1)). In order to simplify the setup, a constant reflectivity profile is taken:

$$F(\zeta) = \mathbf{1}_{[0,\zeta_\mathrm{max}]}. \tag{13}$$

2.3. Second Moments of the cdSAR Image

When a random reflectivity function, such as (10), (12a), or (12b), is substituted into the imaging operator (1), the resulting image is a circular Gaussian random field. For the lower moments of such field, we have for any of these models:

$$\left\langle I(t_\mathbf{y}, \mathbf{y}) \right\rangle = 0, \quad \left\langle I^2(t_\mathbf{y}, \mathbf{y}) \right\rangle = 0, \quad \mathrm{Var}\left(|I(t_\mathbf{y}, \mathbf{y})|^2\right) = \left\langle |I(t_\mathbf{y}, \mathbf{y})|^2 \right\rangle^2. \tag{14}$$

The large variations of intensity described by (14) are responsible for the visual roughness of the resulting speckled images.

The models of scatterers described in section 2.2 are combined to create two different scenarios called the t-model and s-model:

$$\nu(t_\mathbf{z}, \mathbf{z}) = \nu_\text{s-model}(t_\mathbf{z}, \mathbf{z}; \mathbf{z}_\mathrm{d}) = \nu_\mathrm{b}(t_\mathbf{z}, \mathbf{z}) + \nu_s(t_\mathbf{z}, \mathbf{z}; \mathbf{z}_\mathrm{d}) \tag{15a}$$

and

$$\nu(t_\mathbf{z}, \mathbf{z}) = \nu_\text{t-model}(t_\mathbf{z}, \mathbf{z}; \mathbf{z}_\mathrm{d}) = \nu_\mathrm{b}(t_\mathbf{z}, \mathbf{z}) + \nu_t(t_\mathbf{z}, \mathbf{z}; \mathbf{z}_\mathrm{d}), \tag{15b}$$

where ν_b, ν_t, and ν_s are defined in (10), (12a), and (12b), respectively, and $\mathbf{z}_\mathrm{d}$ in (15a) and (15b) is the same. The names "s-model" and "t-model" are intended to match the notations ν_s and ν_t, respectively. The coordinate-delay SAR images resulting from substitution of (15) into (1) are then given by either

$$I_\text{s-model}(t_\mathbf{y}, \mathbf{y}; \mathbf{z}_\mathrm{d}) = I_\mathrm{b}(t_\mathbf{y}, \mathbf{y}) + I_\mathrm{n}(t_\mathbf{y}, \mathbf{y}) + I_s(t_\mathbf{y}, \mathbf{y}; \mathbf{z}_\mathrm{d}) \tag{16a}$$

or

$$I_\text{t-model}(t_\mathbf{y}, \mathbf{y}; \mathbf{z}_\mathrm{d}) = I_\mathrm{b}(t_\mathbf{y}, \mathbf{y}) + I_\mathrm{n}(t_\mathbf{y}, \mathbf{y}) + I_t(t_\mathbf{y}, \mathbf{y}; \mathbf{z}_\mathrm{d}). \tag{16b}$$

In (16), the terms I_{n} are introduced to represent the receiver noise and processing errors.

Taking into account the range-delay ambiguity described in section 2.1 (see (9)), we express the autocorrelations of cdSAR images in the following coordinates:

$$\begin{aligned}\eta &= k_{0\theta}\varphi_T(y_1 - z_{\mathrm{d1}}),\\ \zeta &= \frac{B}{\omega_0}k_{0\theta}(y_2 - z_{\mathrm{d2}}) + \frac{Bt_{\mathrm{y}}}{2},\\ \psi &= \frac{B}{\omega_0}k_{0\theta}(y_2 - z_{\mathrm{d2}}) - \frac{Bt_{\mathrm{y}}}{2}.\end{aligned} \tag{17}$$

The coordinate ψ is aligned with the ambiguity line (7). We are interested in the correlations $\left\langle I(\eta,\zeta,\psi)\overline{I(\eta',\zeta',\psi')}\right\rangle$ between the points on the ambiguity lines that intersect the support of ν_t and ν_s, that is, $\eta = \eta' = 0$, $0 \leqslant \zeta = \zeta' \leqslant \zeta_{\max}$. This way, we select the area of the image affected by the inhomogeneous scatterer, while avoiding the fast decorrelation due to the term $\mathrm{sinc}(BT^0)$ in (4) (see details in Gilman & Tsynkov, 2019).

For the image component I_α that corresponds to a specific scatterer type (i.e., $\alpha \in \{\mathrm{b}, s, t, \mathrm{n}\}$, see (16)), the second order statistics are presented in the following form:

$$\left\langle I_\alpha(0,\zeta,\psi)\overline{I_\alpha(0,\zeta,\psi')}\right\rangle = \sigma_\alpha^2 K_\alpha H_\alpha(\zeta,\psi,\psi'), \tag{18}$$

where the normalizing coefficients K_α and covariance profiles $H_\alpha(\zeta,\psi,\psi')$ are defined as follows.

1. For $\alpha \in \{\mathrm{b}, s, t\}$ denoting the scatterer types given by (10), (12b), and (12a), respectively, formula (18) is obtained directly by substituting the expressions for $\nu_\alpha(t_{\mathbf{z}}, \mathbf{z})$ into (1) and (4). We can choose H_α to be dimensionless with $\sup|H_\alpha| \sim 1$:

$$\begin{aligned}H_{\mathrm{b}}(\zeta,\psi,\psi') &= \Phi\left(0, \kappa\frac{\psi-\psi'}{2}\right),\\ H_t(\zeta,\psi,\psi') &= \Phi\left(0, \kappa\frac{\zeta+\psi}{2}\right)\overline{\Phi\left(0, \kappa\frac{\zeta+\psi'}{2}\right)}\\ &\quad\cdot\frac{1}{\pi}\int_0^\infty F_t^2(\xi)\,\mathrm{sinc}^2(\zeta-\xi)\,d\xi,\\ H_s(\zeta,\psi,\psi') &= \frac{1}{\pi}\int_0^\infty F_s^2(\xi)\,\mathrm{sinc}^2(\zeta-\xi)\\ &\quad\cdot\Phi\left(0, \kappa\frac{\zeta+\psi}{2} - \kappa\xi\right)\overline{\Phi\left(0, \kappa\frac{\zeta+\psi'}{2} - \kappa\xi\right)}\,d\xi,\end{aligned} \tag{19}$$

   where

$$\kappa \overset{\mathrm{def}}{=} \varphi_T^2\frac{\omega_0}{B} \tag{20}$$

   and the normalizing coefficients are given by

$$K_{\mathrm{b}} = N^2\tau^2\frac{\omega_0}{Bk_{0\theta}}\frac{1}{k_{0\theta}\varphi_T}\cdot\pi^2,\ K_t = N^2\tau^2\left(\frac{2}{B}\right)^2\pi,\ K_s = N^2\tau^2\left(\frac{\omega_0}{Bk_{0\theta}}\right)^2\pi.$$

2. For $\alpha = \mathrm{n}$ in (18) representing the noise terms I_{n} introduced in (16), we formally set

$$H_{\mathrm{n}}(\zeta,\psi,\psi') = 0 \text{ if } \psi \neq \psi',\ \text{and } H_{\mathrm{n}}(\zeta,\psi,\psi') = 1 \text{ if } \psi = \psi', \tag{21}$$

   so that all noise terms are uncorrelated. At the same time, the value of $\sigma_{\mathrm{n}}^2 K_{\mathrm{n}}$ is a separate problem parameter to be determined below; see (22).

The terms $I_{\mathrm{b}}, I_{\mathrm{n}}, I_s$ in (16a) and $I_{\mathrm{b}}, I_{\mathrm{n}}, I_t$ in (16b) are considered independent. Because they are also circular Gaussian random variables, formulae (18) and (19) provide a complete description of the statistics of the

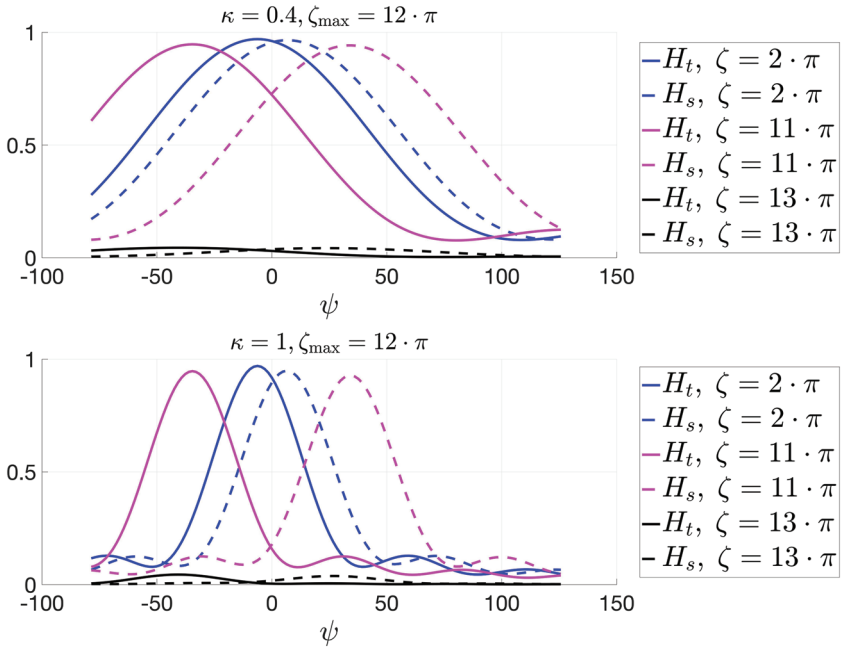

Figure 2. Plots of $H_{s,t} \equiv H_{s,t}(\zeta, \psi, \psi)$, see (19), as a function of ψ for different values of ζ and κ. On a given ambiguity line (i.e., for a certain value of ζ), the distinction between the two types of inhomogeneous scatterers given by (12a) and (12b) is characterized by the separation of solid and dashed lines of the same color.

total image for the arguments corresponding to one and the same ambiguity line if the values of $\sigma_\alpha^2 K_\alpha$ are known. In what follows, we will often use the relative scatterer intensities, or contrasts, defined as follows:

$$q_s = \frac{\sigma_s^2 K_s}{\sigma_s^2 K_s + \sigma_{\rm b}^2 K_{\rm b} + \sigma_{\rm n}^2 K_{\rm n}}, \quad q_t = \frac{\sigma_t^2 K_t}{\sigma_t^2 K_t + \sigma_{\rm b}^2 K_{\rm b} + \sigma_{\rm n}^2 K_{\rm n}}, \quad p_{\rm n} = \frac{\sigma_{\rm n}^2 K_{\rm n}}{\sigma_{\rm b}^2 K_{\rm b}}. \tag{22}$$

As seen from (17)–(21), the quantities (22) are dimensionless.

In Figure 2, we show the expectations of intensities of the image components, $\langle |I_t|^2 \rangle$ and $\langle |I_s|^2 \rangle$ (given by $H_t(\zeta, \psi, \psi)$ and $H_s(\zeta, \psi, \psi)$, respectively, with the help of (17)), for a certain value of $\zeta_{\max}$; see (13). We can see that the separation between the two image components can be made larger by increasing ζ (but not to exceed $\zeta_{\max}$) and/or κ defined in (20). From the analysis of expressions (19), it can also be seen (Gilman & Tsynkov, 2019) that the expectation of image intensity restricted to the ambiguity line peaks approximately at the intersection of this line with the support of the corresponding reflectivity function. In particular,

$$\begin{aligned} \langle |I_{\text{t-model}}(0, \zeta, -\zeta)|^2 \rangle &> \langle |I_{\text{t-model}}(0, \zeta, \zeta)|^2 \rangle, \\ \langle |I_{\text{s-model}}(0, \zeta, -\zeta)|^2 \rangle &< \langle |I_{\text{s-model}}(0, \zeta, \zeta)|^2 \rangle. \end{aligned} \tag{23}$$

It can also be shown (Gilman & Tsynkov, 2019) that the peaks of H_s and H_t in Figure 2are separated by at least one peak half-width provided that

$$\kappa\zeta \gtrsim 20. \tag{24}$$

2.4. Discrimination Between Instantaneous and Dispersive Targets in Individual Images

If statistical averages were available in observations, then inequalities (23) (cf. (9)) could be used to discriminate between the scatterers (15a) and (15b), with the criterion of detectability of scattering delay given by $\kappa\zeta_{\max} \gtrsim 20$; see (24) and (13). Unfortunately, even if this condition is satisfied so that the curves of the same color in Figure 2 are well separated, we may not be able to discriminate between the scatterer types in *individual*images. The reason is that relations (23) and (18) are formulated for expectations, whereas in a typical situation there is only one image, that is, a single realization of the stochastic process $I_{\text{s-model}}$ or $I_{\text{t-model}}$, for which these relations do not necessarily hold.

While statistical averages are unavailable, it is still possible to increase the robustness of the discrimination by analyzing a single coordinate-delay SAR image at multiple locations and delays. We assume that external methods, such as edge detection (Basu, 2002; Canny, 1986; Marr & Hildreth, 1980; Ziou & Tabbone, 1998), can find candidate reference locations of the inhomogeneous scatterer, that is, $\mathbf{z}_\mathrm{d}$. A neighborhood of each $\mathbf{z}_\mathrm{d}$ goes through the discrimination procedure described below. This procedure attributes the target to one of the two classes in (15). For further analysis, we will be selecting a finite set of ζ and ψ using the given $\mathbf{z}_\mathrm{d}$ as a reference; see (17).

Let us take M ambiguity lines for a certain $\mathbf{z}_\mathrm{d}$; these lines will be defined using a set of values $\{\zeta_m\}, 1 \leqslant m \leqslant M$. Next, for each m, we choose N_m values of ψ_{mj}, $1 \leqslant j \leqslant N_m$; these values will play the role of ψ and ψ' for a given $\zeta = \zeta_m$ in (18). Then, I_{mj} will denote the coordinate-delay SAR image sampled in a neighborhood of $(0, \mathbf{z}_\mathrm{d})$. Under the assumption of independence between different image components I_α, we can represent the second order statistics for expressions (16a) and (16b) with the help of (18) as follows:

$$\begin{aligned}\left\langle I_{mj,\text{s-model}}\,\overline{I_{mj',\text{s-model}}}\right\rangle &= \sum_{\alpha\in S}\sigma_\alpha^2 K_\alpha H_\alpha(\zeta_m,\psi_{mj},\psi_{mj'}), \quad 1\leqslant j,j'\leqslant N_m,\\ \left\langle I_{mj,\text{t-model}}\,\overline{I_{mj',\text{t-model}}}\right\rangle &= \sum_{\alpha\in \mathcal{T}}\sigma_\alpha^2 K_\alpha H_\alpha(\zeta_m,\psi_{mj},\psi_{mj'}), \quad 1\leqslant j,j'\leqslant N_m,\end{aligned} \tag{25}$$

where

$$\mathcal{S} = \{\mathrm{b},\mathrm{n},s\}, \quad \mathcal{T} = \{\mathrm{b},\mathrm{n},t\}. \tag{26}$$

For each of the two scenarios in (16), that is, for $\mathcal{S}$ and $\mathcal{T}$ in (26), the discrimination algorithm seeks a set of values for unknowns σ_α^2 that maximizes the probability density of the image with the statistics described by (25). Then, the model that yields the larger of the two maxima is chosen. Essentially, this is a maximum likelihood (ML) based procedure (Mendenhall & Scheaffer, 1973; Oliver & Quegan, 1998).

The probability density of the sampled image $\{I_{mj}\}$ for either of the two models is calculated as follows. For each m, a real-valued vector $\mathbf{r}_m$ of length $2N_m$ is defined as follows:

$$\mathbf{r}_m = \left(\mathrm{Re}\,I_{m1}, \mathrm{Im}\,I_{m1}, \mathrm{Re}\,I_{m2}, \mathrm{Im}\,I_{m2}, \ldots \mathrm{Re}\,I_{mN_m}, \mathrm{Im}\,I_{mN_m}\right)^\mathbf{T}. \tag{27}$$

Then, due to the circular Gaussianity and independence of all I_α, each of the two lines in (25) can be recast as

$$\left\langle \mathbf{r}_m\mathbf{r}_m^\mathbf{T}\right\rangle \overset{\text{def}}{=} \mathbf{M}^{(m)} = \begin{pmatrix} \mathbf{M}_{11}^{(m)} & \mathbf{M}_{12}^{(m)} & \cdots & \mathbf{M}_{1N_m}^{(m)} \\ \mathbf{M}_{21}^{(m)} & \mathbf{M}_{22}^{(m)} & \cdots & \mathbf{M}_{2N_m}^{(m)} \\ \vdots & & \ddots & \\ \mathbf{M}_{N_m1}^{(m)} & \mathbf{M}_{N_m2}^{(m)} & \cdots & \mathbf{M}_{N_mN_m}^{(m)} \end{pmatrix}, \tag{28}$$

where the individual 2×2 blocks on the right-hand side of (28) are given by

$$\mathbf{M}_{jj'}^{(m)} = \frac{1}{2}\sum_{\alpha\in\mathcal{A}}\sigma_\alpha^2 K_\alpha \cdot \begin{pmatrix} \mathrm{Re}\,H_\alpha(\zeta_m,\psi_{mj},\psi_{mj'}) & -\mathrm{Im}\,H_\alpha(\zeta_m,\psi_{mj},\psi_{mj'}) \\ \mathrm{Im}\,H_\alpha(\zeta_m,\psi_{mj},\psi_{mj'}) & \mathrm{Re}\,H_\alpha(\zeta_m,\psi_{mj},\psi_{mj'}) \end{pmatrix}. \tag{29}$$

Choosing $\mathcal{A} = \mathcal{S}$ or $\mathcal{A} = \mathcal{T}$ in (29) yields two expressions for the matrices $\mathbf{M}^{(m)}$ in (28), henceforth called $\mathbf{M}_{\text{s-model}}^{(m)}$ and $\mathbf{M}_{\text{t-model}}^{(m)}$. These matrices give rise to two multivariate Gaussian probability distribution functions (pdfs):

$$\begin{aligned} p_{\text{s-model}}(\mathbf{r}_m) &= \frac{1}{\sqrt{\det\left(2\pi\mathbf{M}_{\text{s-model}}^{(m)}\right)}}\exp\left(-\frac{1}{2}\mathbf{r}_m^\mathbf{T}(\mathbf{M}_{\text{s-model}}^{(m)})^{-1}\mathbf{r}_m\right),\\ p_{\text{t-model}}(\mathbf{r}_m) &= \frac{1}{\sqrt{\det\left(2\pi\mathbf{M}_{\text{t-model}}^{(m)}\right)}}\exp\left(-\frac{1}{2}\mathbf{r}_m^\mathbf{T}(\mathbf{M}_{\text{t-model}}^{(m)})^{-1}\mathbf{r}_m\right).\end{aligned} \tag{30}$$

Then, formula (30) is extended by including the data from multiple ambiguity lines given by a set of ζ_m. When the sampling step in ζ is equal to or exceeds π, the data for different ζ_m can be considered independent

(see Gilman & Tsynkov, 2019, and Oliver & Quegan, 1998, Section 4.10). Then, for the full data set vector $\mathbf{R}$ that combines all $\mathbf{r}_m$-vectors (27):

$$\mathbf{R} = \left(\mathbf{r}_1^{\mathrm{T}}, \mathbf{r}_2^{\mathrm{T}}, \ldots, \mathbf{r}_m^{\mathrm{T}}, \ldots, \mathbf{r}_M^{\mathrm{T}}\right)^{\mathrm{T}} \tag{31a}$$

we have

$$p(\mathbf{R}) = \prod_{m=1}^{M} p(\mathbf{r}_m). \tag{31b}$$

This yields

$$\begin{aligned} p_{\text{s-model}}(\mathbf{R}) &= \frac{1}{\sqrt{\det\left(2\pi\mathbf{M}_{\text{s-model}}\right)}} \exp\left(-\frac{1}{2}\mathbf{R}^{\mathrm{T}}(\mathbf{M}_{\text{s-model}})^{-1}\mathbf{R}\right), \\ p_{\text{t-model}}(\mathbf{R}) &= \frac{1}{\sqrt{\det\left(2\pi\mathbf{M}_{\text{t-model}}\right)}} \exp\left(-\frac{1}{2}\mathbf{R}^{\mathrm{T}}(\mathbf{M}_{\text{t-model}})^{-1}\mathbf{R}\right), \end{aligned} \tag{32}$$

where the matrices $\mathbf{M}_{\text{s-model}}$ and $\mathbf{M}_{\text{t-model}}$ are block diagonal with blocks $\mathbf{M}^{(m)}_{\text{s-model}}$ and $\mathbf{M}^{(m)}_{\text{t-model}}$, respectively:

$$\mathbf{M}_{\text{s-model}} \overset{\text{def}}{=} \operatorname{diag}_{1\leqslant m\leqslant M}\left(\mathbf{M}^{(m)}_{\text{s-model}}\right), \quad \mathbf{M}_{\text{t-model}} \overset{\text{def}}{=} \operatorname{diag}_{1\leqslant m\leqslant M}\left(\mathbf{M}^{(m)}_{\text{t-model}}\right). \tag{33}$$

The actual data set vector $\mathbf{Q}$ representing a sampled image has the same structure as $\mathbf{R}$ of (31):

$$\mathbf{Q} = \left(\mathbf{q}_1^{\mathrm{T}}, \mathbf{q}_2^{\mathrm{T}}, \ldots, \mathbf{q}_m^{\mathrm{T}}, \ldots\right)^{\mathrm{T}}, \tag{34}$$

where each vector $\mathbf{q}_m$ corresponds to image values taken at a certain ambiguity line. The likelihood functions (Mendenhall & Scheaffer, 1973) are obtained by considering the expressions $p_{\text{s-model}}(\mathbf{Q})$ and $p_{\text{t-model}}(\mathbf{Q})$ of (32) as functions of the unknown scatterer intensities $\{\sigma_\alpha^2\}$ that enter $\mathbf{M}^{(m)}$ via (29) for each of the models in (15). Then, two optimization problems are formulated as follows:

$$\check{p}_s = \max_{\sigma_{\mathrm{b}}^2,\sigma_{\mathrm{n}}^2,\sigma_s^2} p_{\text{s-model}}(\mathbf{Q}), \quad \check{p}_t = \max_{\sigma_{\mathrm{b}}^2,\sigma_{\mathrm{n}}^2,\sigma_t^2} p_{\text{t-model}}(\mathbf{Q}), \tag{35}$$

subject to $\sigma_{\mathrm{b}}^2, \sigma_{\mathrm{n}}^2, \sigma_s^2, \sigma_t^2 \geq 0$. The resulting solutions $\check{p}_s$ and $\check{p}_t$ yield the ML values for the corresponding scatterer models.

The discrimination procedure suggested in Gilman and Tsynkov (2019) relies on comparing the two maxima, $\check{p}_s$ and $\check{p}_t$, defined by (35) and (32). It is common to consider the logarithm of the likelihood rather than the likelihood itself. Accordingly, the discrimination algorithm operates with the parameter

$$l \overset{\text{def}}{=} \log\check{p}_t - \log\check{p}_s \tag{36}$$

and implements the following classification rule:

```
if l > 0
then
   the target is classified as a delayed (dispersive) scatterer (15b), (16b)
else
   the target is classified as an instantaneous scatterer (15a), (16a).
```
(37)

2.5. Average Misclassification Frequency

The discrimination method of section 2.4 works with individual images because a typical acquisition produces only a single image. However, as the data (i.e., cdSAR images) are of statistical nature, the characterization of the quality of classification should involve statistical ensembles.

In Gilman and Tsynkov (2019), Monte Carlo simulations have been used to generate pairs of ensembles of sampled cdSAR images according to the statistics (32) with the same values of target contrast

$$q_s = q_t = q, \tag{38}$$

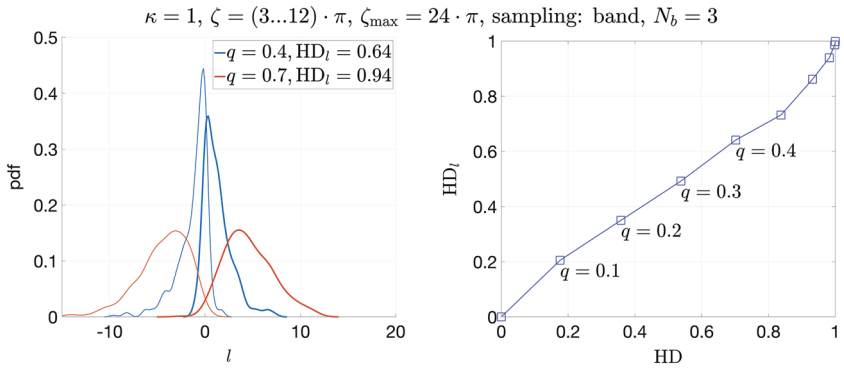

Figure 3. (left) Empirical probability density functions for the quantity l of (36) for two values of target contrast defined by (38). Thick and thin curves correspond to the ensembles generated from t-model and s-model, respectively; see (16). HD_l denotes the Hellinger distance between the empirical pdfs of l obtained from ensembles of sampled cdSAR images. (right) Relation between HD_l and $\mathrm{HD} \equiv \mathrm{HD}[p_{\text{s-model}}, p_{\text{t-model}}]$ given by formula (41). The parametric form of this plot is $\big(\mathrm{HD}(q), \mathrm{HD}_l(q)\big)$ as the target contrast q varies from 0 to 0.9. The sampling pattern is illustrated in Figure 4.

see (22), for each of the two models in (15). The average misclassification frequency due to algorithm (37) was defined as

$$r_{st} = (r_s + r_t)/2, \tag{39}$$

where r_s and r_t are the fractions of misclassified images originating from the s-model and t-model, respectively. The dependence of r_{st} on observation and target parameters, such as κ, $\zeta_{\max}$, M, and q, has been studied in Gilman and Tsynkov (2019) as well.

3. Misclassification Frequency and the Hellinger Distance

3.1. Hellinger Distance: An Example Using Empirical Probability Distributions

Let us note that algorithm (37) represents one of many ways of interpreting the value of l. If, for example, false positives in certain applications are much less harmful than false negatives, then a threshold (or bias) can be introduced in the comparison in (37). Alternatively, a threshold can be used to introduce a confidence level into the classification; see Gilman and Tsynkov (2020).

On the other hand, there is a way to characterize our ability to perform the discrimination based on the value of l without referring to any particular classification rule. This approach involves calculation of statistical divergence measures (Pardo, 2006; Pollard, 2002) between the pdfs. We will illustrate this concept using the pdfs of the quantity l given by (36). Since obtaining pdf(l) analytically is not straightforward, we will use the empirical pdfs (see, e.g., van der Vaart, 1998, Chapter 19). Figure 3 (left panel) presents two pairs of empirical pdfs of l; each pair of pdfs (shown by the curves sharing the same color) is generated from a pair of ensembles of sampled cdSAR images with the same target contrast; see section 2.5, by applying formula (36) to every image.

One particular example of statistical divergence measures is the Hellinger distance $\mathrm{HD}[f, g]$ defined as follows:

$$\mathrm{HD}^2[f, g] = \frac{1}{2}\int \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx, \tag{40}$$

(see Pardo, 2006; Pollard, 2002), where $f(x)$ and $g(x)$ are two pdfs. The convenience of using the Hellinger distance can be seen in that it is a metric (in particular, $\mathrm{HD}[f, g] = \mathrm{HD}[g, f] \leqslant \mathrm{HD}[f, h] + \mathrm{HD}[h, g]$) and that in the limiting cases, $\mathrm{HD}[f, g] = 1$ when f and g are completely disjoint, whereas $\mathrm{HD}[f, f] = 0$. If the two pdfs coincide ($\mathrm{HD} = 0$), it is easy to show that any binomial classifier yields $r_{st} = 0.5$. If the supports of f and g are disjoint ($\mathrm{HD} = 1$) and their boundaries known, then we can use these boundaries to build a classifier that will never make a mistake, that is, $r_{st} = 0$. For the intermediate cases, it is reasonable to associate low misclassification frequencies with high values of the Hellinger distance between the two pdfs, and vice versa.

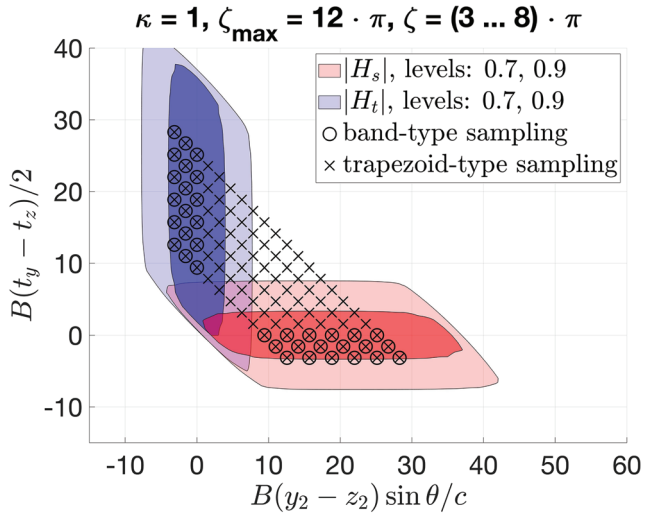

Figure 4. Two sampling patterns for $N_b = 3$. The background displays the expectations of intensities of the inhomogeneous image components according to (18) and (19) for scatterers described by (12) and (13); these expectations are obtained by setting $\psi = \psi'$ in (18).

For illustration purposes, Figure 3 (left) presents the values of the Hellinger distance HD_l for the displayed pairs of experimental pdfs. As expected, the pair of ensembles of high-contrast targets yields a pair of pdfs with better separation and, accordingly, a larger value of the Hellinger distance between them, as compared to low-contrast targets.

3.2. Model-Based Hellinger Distance

While the Hellinger distance HD_l can help quantify the separation between the empirical pdfs of l (as in Figure 3, left), obtaining the empirical pdfs themselves is computationally costly. Instead, we would like to calculate the Hellinger distance directly from the pdfs defined in (30). This will allow us to bypass the time-consuming step of Monte Carlo simulations; see section 2.5. Note that the resulting $\mathrm{HD}[p_{\text{s-model}}, p_{\text{t-model}}]$ will characterize the pdfs of vectors $\mathbf{R}$ or $\mathbf{Q}$ defined in (31) and (34), respectively. Accordingly, the integral in (40) becomes multidimensional. Henceforth, we will denote $\mathrm{HD}[p_{\text{s-model}}, p_{\text{t-model}}]$ by merely HD, to distinguish it from the empirically obtained HD_l of section 3.1. An example of the relation between HD and HD_l is presented in Figure 3 (right).

For the Hellinger distance between two zero-mean Gaussian pdfs as in (30), we can use the following analytic expression (Pardo, 2006):

$$\mathrm{HD}^2[p_{\text{s-model}}, p_{\text{t-model}}] = 1 - \frac{\det\,(\mathbf{M}_{\text{s-model}})^{1/4}\det\,(\mathbf{M}_{\text{t-model}})^{1/4}}{\det\,\left((\mathbf{M}_{\text{s-model}} + \mathbf{M}_{\text{t-model}})/2\right)^{1/2}}. \quad (41)$$

In section 4, we study the properties of the Hellinger distance given by formula (41).

4. Hellinger Distance for Sampled cdSAR Images

Let us recall that the matrices $\mathbf{M}_{\text{s-model}}$ and $\mathbf{M}_{\text{t-model}}$ used in (41) depend on the sampling patterns, that is, the set of arguments of I for each I_{mj} in (25). We have constructed two sampling patterns, called band-type and trapezoid-type; see Figure 4. For any positive integer N_b these two patterns are defined as follows:

$$\begin{aligned} \text{band-type:} \quad & \zeta_m \leqslant |\psi_{mj}| \leqslant \zeta_m + (N_b - 1)\pi, \\ \text{trapezoid-type:} \quad & 0 \leqslant |\psi_{mj}| \leqslant \zeta_m + (N_b - 1)\pi, \end{aligned} \quad (42)$$

where all $|\psi_{mj}|$ are multiples of π,

$$\zeta_m = \zeta_{\min} + (m-1)\pi,\ m = 1, \ldots, M,\ M = 1 + [(\zeta_{\max} - \zeta_{\min})/\pi], \quad (43)$$

$[\ldots]$ denotes integer part, $\zeta_{\max}$ is defined in (13), and $\zeta_{\min} = 3\pi$ is introduced to cut off the transitional effects due to the behavior of $F(\zeta)$ of (13) in the vicinity of $\zeta = 0$ (besides, this excludes the ambiguity lines where the separation between $|H_s|$ and $|H_t|$ is small, see formulae (18) and (19) and Figures 2 and 4). These sampling patterns are designed to capture the difference between $\langle |I_t|^2 \rangle$ and $\langle |I_s|^2 \rangle$ in the plane $y_1 - z_{\text{d1}} = 0$. In each pattern, the sampling points are arranged into a set of ambiguity lines $\zeta = \zeta_m$; see (17), with increment π between the adjacent values of ζ_m (note that the increment in ψ is also chosen equal to π). For the band-type pattern, N_b of (42) is the number of samples per ambiguity line taken in the vicinity of each type of inhomogeneous scatterer. Figure 4 corresponds to $N_b = 3$, which yields a total of $2N_b = 6$ samples per ambiguity line in the band-type pattern. For simplicity, we use the same quantity N_b to define the corresponding trapezoid-type sampling. Note that the sampling pattern used in Gilman and Tsynkov (2019) can be called band type with $N_b = 1$. In general, using the trapezoid-type sampling or band-type with $N_b > 1$ allows one to explore the effect of a larger number samples per ambiguity line on the quality of discrimination; see Figure 6.

In Figure 5, we plot the average relative misclassification frequency (39) versus Hellinger distance (41) for a wide variety of parameters. The most interesting observation is that the data closely follow a simple linear pattern (also shown in the plots):

$$\mathrm{HD} + 2r_{st} \approx 1. \quad (44)$$

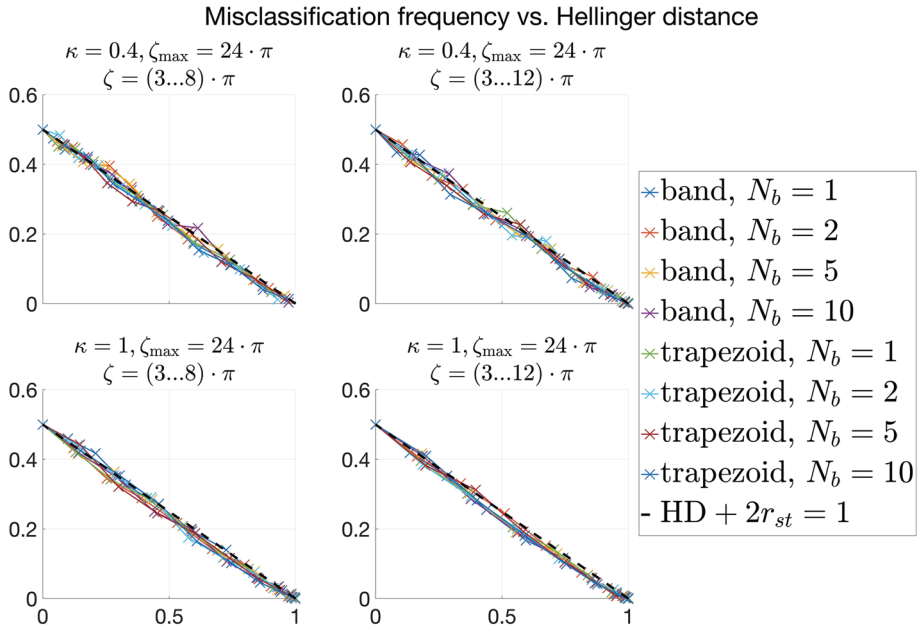

Figure 5. Average relative misclassification frequency due to algorithm (37) versus Hellinger distance (41). Different colored curves correspond to different sampling patterns and values of N_b (see Figure 4); each curve is a parametric plot $\big(\mathrm{HD}(q), r_{st}(q)\big)$, with the contrast q of (38) varying from 0 to 0.9 with increment 0.1. Different panels correspond to different values of κ and $\max(\zeta_m)$ covering a wide range of imaging conditions; for example, condition (24) is never satisfied for the top left panel, while for the bottom right panel, it is satisfied for approximately one half of the ambiguity lines in the sampling patterns. Notice a very good fit of all the data with the linear relation (44).

This result appears particularly intriguing if we realize that the matrices $\mathbf{M}_{\text{s-model}}$ and $\mathbf{M}_{\text{t-model}}$ in formula (41) are computed for certain values of the scatterer intensity (or contrast), whereas due to the stochasticity of the data, the optimization procedure (35) involves calculation of these matrices for rather different contrasts within a single ensemble. A theoretical justification of formula (44) remains a future task. Yet currently we see that the vertical deviation between the curves in all panels of Figure 5 is very small. This means that using the analytical formula (41), we can predict, with a fairly good accuracy, the performance of binary classification by the algorithm (37).

The foregoing observation gives us an analytic tool for gaining insight into a range of complex phenomena that involve the classification method of section 2.4. As an example, let us analyze whether there may be any advantage in sampling the homogeneous areas of a cdSAR image (in Gilman & Tsynkov, 2019, we used the term "homogeneous samples"). In particular, homogeneous samples can be taken for $\zeta \ll -\zeta_{\min}$ or $\zeta \gg \zeta_{\max}$ or $|y_1 - z_{\mathrm{d1}}| \gg (k_{0\theta}\varphi_T)^{-1}$ (the latter condition can be understood from the expression for the first argument of Φ in (4)). The homogeneous samples alone cannot deliver the discrimination between the scenarios in (15) because they are taken in the areas not affected by the inhomogeneous scatterers. It, however, can be argued that these samples may allow us to estimate the intensity of the background and noise with higher accuracy. In turn, these estimates effectively reduce the number of optimization parameters in (35) and thus may help improve the performance. Moreover, as the area affected by inhomogeneous scatterers is in the vicinity of the plane $y_1 = z_{\mathrm{d1}}$, we may be able to collect many more background samples than inhomogeneous samples and thus expect a high accuracy of estimation of σ_{b}^2 and σ_{n}^2. Nonetheless, in Gilman and Tsynkov (2019) we saw that the effect of background samples has always been insignificant. Then, given the relation (44), we should expect that adding the homogeneous samples to the data set does not change the Hellinger distance between the ensembles of section 2.5.

With the help formula (41), one can consider the effect of homogeneous samples on the Hellinger distance analytically. Denote by $\mathbf{M}^{\mathrm{h}}$ the covariance matrix of the homogeneous samples built similarly to (33). Note that there is no difference between the s-model and t-model for the homogeneous samples. Appending the

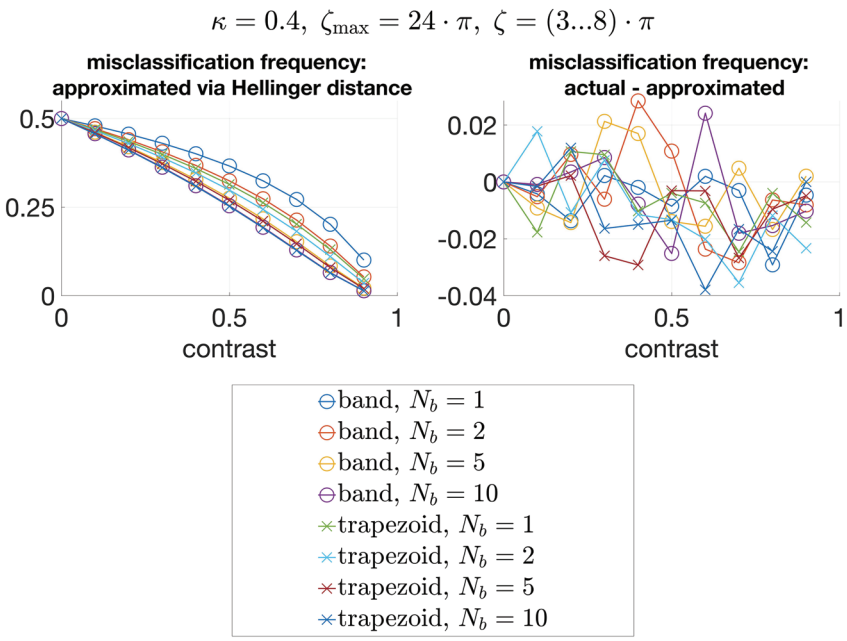

Figure 6. (left) Plots of misclassification frequency r_{st} obtained by formula (44) versus target contrast (see (22)). The values of κ and ζ correspond to the top left panel in Figure 5. As the number of samples per ambiguity line (see Figure 4) increases, its effect tends to saturate. (right) The difference between the actual and approximated misclassification frequency. Similarly to Figure 5, lines of different color correspond to different sampling patterns and values of N_b.

vector of homogeneous samples to $\mathbf{R}$ of (31), we introduce two covariance matrices, $\mathbf{M}'_{\text{s-model}}$ and $\mathbf{M}'_{\text{t-model}}$, associated with the new data set vector as follows:

$$\mathbf{M}'_{\text{s-model}} = \text{diag}(\mathbf{M}_{\text{s-model}}, \mathbf{M}^{\text{h}}),\ \mathbf{M}'_{\text{t-model}} = \text{diag}(\mathbf{M}_{\text{t-model}}, \mathbf{M}^{\text{h}}), \tag{45}$$

because we assume that the homogeneous samples are independent from the inhomogeneous samples. Obviously,

$$\det \mathbf{M}'_{\text{s-model}} = \det \mathbf{M}_{\text{s-model}} \cdot \det \mathbf{M}^{\text{h}},\ \det \mathbf{M}'_{\text{t-model}} = \det \mathbf{M}_{\text{t-model}} \cdot \det \mathbf{M}^{\text{h}}. \tag{46}$$

Consequently, the use of $\mathbf{M}'_{\text{s-model}}$ and $\mathbf{M}'_{\text{t-model}}$ in (41) instead of $\mathbf{M}_{\text{s-model}}$ and $\mathbf{M}_{\text{t-model}}$, respectively, does not change the resulting HD. Although the relation illustrated in Figure 5 is not exact, which makes any inference about the misclassification frequency approximate, we can still treat the invariance of HD with respect to homogeneous samples as an additional corroboration of the result obtained experimentally in Gilman and Tsynkov (2019), namely, that the homogeneous samples don't have a significant effect on the quality of classification.

In Figure 6, we demonstrate the prediction of misclassification frequency using Hellinger distance and relation (44). The prediction error (the right panel) is quite small. Note that the ensembles of computer-generated images obtained by Monte Carlo simulations contained 400 images for each of the two models (t-model and s-model). Accordingly, the stochastic error of determining the misclassification frequency was about 0.05 (Gilman & Tsynkov, 2019), which is of the same order of magnitude as the prediction error; see the right plot in Figure 6. Overall, the accuracy of predicting the misclassification rate with the help of formulae (44) and (41) appears quite acceptable.

5. Kullback-Leibler Divergence as an Alternative to Hellinger Distance

We chose the Hellinger distance as a statistical divergence measure in section 3 because of its convenience in implementation, both in the empirical (section 3.1) and model-based (section 3.2) cases. In this section, we reproduce some results of section 4 using an alternative characteristic that is popular in information

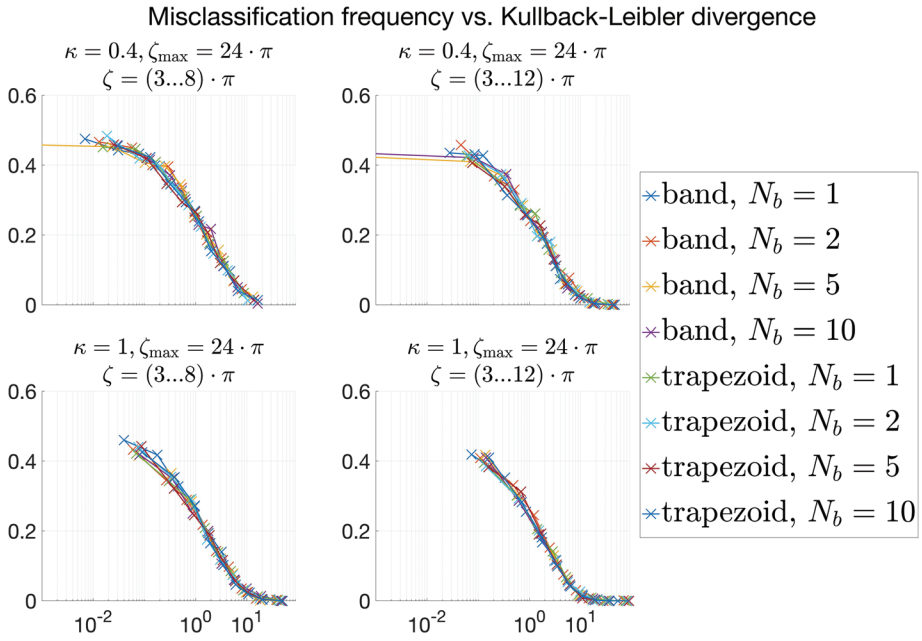

Figure 7. Same as Figure 5, but for the Kullback-Leibler divergence (48) instead of the Hellinger distance (41).

theory, namely, the Kullback-Leibler divergence (Goodfellow et al., 2016; Hastie et al., 2009). The definition is as follows (cf. (40)):

$$\mathrm{KL}[f, g] = \int f(x) \log \frac{f(x)}{g(x)} dx. \tag{47}$$

Unlike the Hellinger distance, the Kullback-Leibler divergence is not a metric; in particular, it is possible to have $\mathrm{KL}[f, g] \neq \mathrm{KL}[g, f]$. On the practical side, calculation of the integral in (47) from empirical data (e.g., as shown in Figure 3) is complicated because the integrand may become unbounded if $g(x) = 0$. Still, there is a closed form analytic expression for the Gaussian distributions (30) (Pardo, 2006):

$$\mathrm{KL}\left[p_{\text{s-model}}, p_{\text{t-model}}\right] = \frac{1}{2}\mathrm{trace}\left(\mathbf{M}^{-1}_{\text{t-model}}\mathbf{M}_{\text{s-model}} - \mathbf{I}\right) + \frac{1}{2}\log\frac{\det \mathbf{M}_{\text{t-model}}}{\det \mathbf{M}_{\text{s-model}}}, \tag{48}$$

where $\mathbf{I}$ is a unit matrix of the appropriate dimension.

Using formula (48), one can perform the same analysis as in section 4, but with a different divergence measure. In Figure 7, we plot the misclassification frequency (39) versus Kullback-Leibler divergence similarly to Figure 5 for the Hellinger distance. Although there is no simple analytic fit for these plots, we still observe monotonicity and a very tight grouping of all experimental data in every panel of Figure 7. This is a key similarity between these two divergence measures when it comes to predicting the misclassification frequency of the classifier (37).

Let us also note that, similarly to the Hellinger distance, the Kullback-Leibler divergence (47) is insensitive to the homogeneous samples, the procedure expressed mathematically by (45). There is no effect on the second term on the right-hand side of formula (48) as immediately follows from (46). In order to demonstrate the same for the first term, we additionally use the relation $(\mathbf{M}'_{\text{t-model}})^{-1} = \mathrm{diag}\left((\mathbf{M}_{\text{t-model}})^{-1}, (\mathbf{M}^{\text{h}})^{-1}\right)$, which follows from (45).

Altogether, we find that both the Hellinger distance and Kullback-Leibler divergence can be used to predict the misclassification frequency of the classifier (37). Moreover, the findings of section 4, with the exception of the linear fit (44), are likely not due to the particular form of the metric (40).

6. Conclusion

We have demonstrated that the statistical divergence measures can be used efficiently for predicting the performance of the classifier (37) that distinguishes between the instantaneous and delayed targets in

coordinate-delay SAR images. We have used two particular divergence measures: the Hellinger distance and the Kullback-Leibler divergence; see sections 3 and 5, respectively. Both divergence measures proved useful, although the Hellinger distance appears more convenient because of the empirical linear relation (44) between the measure and misclassification frequency.

The findings of this work can be used in the analysis of detection scenarios that involve instantaneous and delayed scatterers. In particular, we have demonstrated their functionality in the analysis of sensitivity of misclassification rate to background samples for both the Hellinger distance and Kullback-Leibler divergence. Possible extensions include the analysis of targets with a more general form of the reflectivity profile $F(\zeta)$ compared to (13) and/or those that are both delayed and extended in space. Another possibility is to generalize the analysis to non-Gaussian speckle that is characteristic for high-resolution radar images (Oliver & Quegan, 1998, Section 5.3). In the future, the current analysis of the reflectivity models, target classification algorithm, and the approach to predicting its error rate will be corroborated by additional numerical simulations.

Data Availability Statement

The data used in this article are archived by 4TU.Centre for Research Data and available at the following address (https://doi.org/10.4121/uuid:bb8c9484-531a-488e-8328-7db910547585.V2).

Acknowledgments
We are grateful to Profs. Alen Alexanderian, Arvind Saibaba, and Ralph Smith (NCSU) for fruitful discussions. We also thank the Associate Editor and the anonymous reviewers for their useful feedback. This material is based upon work supported by the US Air Force Office of Scientific Research (AFOSR) under award number FA9550-17-1-0230.

References

Basu, M. (2002). Gaussian-based edge-detection methods—A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *32*(3), 252–260.

Borden, B. (1998). Dispersive scattering for radar-based target classification and duct-induced image artifact mitigation, *NATO symposium on non-cooperative air target identification using radar* (pp. 14.1–14.7). Mannheim, Germany: North Atlantic Treaty Organization, Research and Technology Organization.

Cakoni, F., Colton, D., & Monk, P. (2011). *The linear sampling method in inverse electromagnetic scattering* (Vol. 80). Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). https://doi.org/10.1137/1.9780898719406

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, *PAMI-8*(6), 679–698.

Cheney, M., & Borden, B. (2009). *Fundamentals of radar imaging* (Vol. 79). Philadelphia, PA: SIAM.

Cumming, I. G., & Wong, F. H. (2005). *Digital processing of synthetic aperture radar data: Algorithms and implementation*. Boston: Artech House.

Ferrara, M., Homan, A., & Cheney, M. (2017). Hyperspectral SAR. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(3), 1–14.

Gallager, R. G. (2008a). Circularly-symmetric Gaussian random vectors. Online at http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf

Gallager, R. G. (2008b). *Principles of digital communication* (Vol. 1). UK: Cambridge University Press Cambridge.

Gilman, M., Smith, E., & Tsynkov, S. (2017). *Transionospheric synthetic aperture imaging*. Cham, Switzerland: Birkhäuser/Springer. https://doi.org/10.1007/978-3-319-52127-5

Gilman, M., & Tsynkov, S. (2019). Detection of delayed target response in SAR. *Inverse Problems*, *35*, 85005. https://doi.org/10.1088/1361-6420/ab1c80

Gilman, M., & Tsynkov, S. (2020). Statistical characterization of scattering delay in synthetic aperture radar imaging. *Inverse Problems & Imaging*, *14*(3), 511–533. https://doi.org/10.3934/ipi.2020024

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. http://www.deeplearningbook.org

Goodman, J. W. (1984). Statistical properties of laser speckle patterns. *Laser speckle and related phenomena* (pp. 9–75). New York/Heidelberg/Berlin/Tokyo: Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. note=Data mining, inference, and prediction, https://doi.org/10.1007/978-0-387-84858-7

Inglada, J., & Mercier, G. (2007). A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(5), 1432–1445. https://doi.org/10.1109/TGRS.2007.893568

Jakowatz, C. V. Jr., Wahl, D. E., Eichel, P. H., Ghiglia, D. C., & Thompson, P. A. (1996). *Spotlight-mode synthetic aperture radar: A signal processing approach*. Boston/London/Dordrecht: Springer.

Kirsch, A., & Grinberg, N. (2008). *The factorization method for inverse problems* (Vol. 36). Oxford: Oxford University Press.

Lapidoth, A. (2017). *A foundation in digital communication*. Cambridge, UK: Cambridge University Press.

Lee, J. S., & Pottier, E. (2009). *Polarimetric radar imaging from basics to applications*. Boca Raton, FL: CRC Press.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society B*, *207*(1167), 187–217.

Mendenhall, W., & Scheaffer, R. L. (1973). *Mathematical statistics with applications*. North Scituate, Mass: Duxbury Press.

Nascimento, A. D. C., Cintra, R. J., & Frery, A. C. (2010). Hypothesis testing in speckled data with stochastic distances. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(1), 373–385. https://doi.org/10.1109/TGRS.2009.2025498

Oliver, C., & Quegan, S. (1998). *Understanding synthetic aperture radar images*. Boston: Artech House.

Pardo, L. (2006). *Statistical inference based on divergence measures* (Vol. 185). Boca Raton, FL: Chapman & Hall/CRC.

Pedersen, H. (1974). The roughness dependence of partially developed, monochromatic speckle patterns. *Optics Communications*, *12*(2), 156–159. https://doi.org/10.1016/0030-4018(74)90380-0

Pierri, R., Soldovieri, F., Liseno, A., & De Blasio, F. (2002). Dielectric profiles reconstruction via the quadratic approach in 2-D geometry from multifrequency and multifrequency/multiview data. *IEEE Transactions on Geoscience and Remote Sensing*, *40*(12), 2709–2718.

Pollard, D. (2002). *A user's guide to measure theoretic probability* (Vol. 8). Cambridge: Cambridge University Press.

Torres, L., Sant'Anna, S. J., da Costa Freitas, C., & Frery, A. C. (2014). Speckle reduction in polarimetric SAR imagery with stochastic distances and nonlocal means. *Pattern Recognition*, *47*(1), 141–157. https://doi.org/10.1016/j.patcog.2013.04.001

Trintinalia, L. C., & Ling, H. (1997). Joint time-frequency ISAR using adaptive processing. *IEEE Transactions on Antennas and Propagation*, *45*(2), 221–227.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511802256

Ziou, D., & Tabbone, S. (1998). Edge detection techniques—An overview. *Pattern Recognition and Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii*, *8*, 537–559.