Benchmarking Framework for Multitemporal SAR Despeckling

Gerardo Di Martino, *Senior Member, IEEE*, Alessio Di Simone, *Member, IEEE*,
Antonio Iodice, *Senior Member, IEEE*, and Daniele Riccio, *Fellow, IEEE*

Abstract—In this article, we propose a novel benchmarking framework for a quantitative assessment of the performance of despeckling algorithms for multitemporal synthetic aperture radar (SAR) imagery. A number of canonical scenes and data sets are analyzed so as to investigate both speckle reduction and feature preservation capabilities of the filters. Despeckling performance is evaluated by proper quality measures that are defined according to the scene. Due to the lack of real-world speckle-free SAR images, the proposed benchmarking tool relies on an accurate and well-assessed SAR simulator which allows for generating realistic SAR images accounting for electromagnetic (EM) and geometrical parameters of the sensed surface. Accuracy and convergence properties of filters are first measured on scenes with stationary reflectivity. Then, for a more realistic performance prediction in practical situations, the effects of temporal changes of the scene reflectivity on the despeckled images are measured on time series with time-varying reflectivity. In the latter case, performance parameters are intended to measure the capability of the filter to preserve both the perturbation and its impact on the other bands. The whole benchmarking framework is applied to a representative set of state-of-the-art multitemporal filters. Interestingly, their performance as evaluated by means of our framework is well in agreement with (qualitative) visual inspections by SAR specialists. Proposed quality metrics are measured under the hypothesis of uncorrelated bands, which defines the best case for most multitemporal filters. A numerical sensitivity analysis of the performance of filters against correlation coefficient is carried out to investigate the temporal correlation effects on the despeckled time series.

Index Terms—Multitemporal filtering, quality assessment, synthetic aperture radar (SAR), speckle, synthetic aperture radar (SAR) simulation.

I. INTRODUCTION

REMOTE sensing instruments and data offer an invaluable support to many fields and applications related to Earth observation and monitoring, such as agriculture, urban planning, disaster management, security and safety, resource control, climate change. Among others, space-based synthetic aperture radar (SAR) enables remote sensing at a global scale regardless of daylight and weather conditions, in contrast

with multispectral and hyper-spectral sensors whose imaging capabilities are impaired by cloudiness and scarce sunlight illumination conditions. Additionally, SAR systems can provide polarization diversity which is useful for classification purposes [1]. Actually, SAR end-users can now benefit from a huge amount of data which are acquired at different resolution (both in space and time), frequency, look angle, and polarization and that can be fruitfully exploited in advanced image processing techniques, e.g., machine learning and tomography. On the one hand, the recent class of meter-resolution spaceborne SAR systems, e.g., TerraSAR-X, COSMO-SkyMed, and RADARSAT-2, along with the new ultrahigh-resolution (submetric) sensors, e.g., COSMO-SkyMed 2, allows for information retrieval with unprecedented spatial resolution at a cost of reduced coverage area. On the other hand, medium-resolution systems, such as Sentinel-1, are experiencing a renewed interest as they offer imaging capabilities complementary to very-high-resolution and ultrahigh-resolution sensors, thanks to the larger coverage area which allows for a reduced revisit time.

Among past and current SAR missions, Sentinel-1 is particularly attractive in the remote sensing field, thanks to the low revisit time (down to three days at the Equator), polar orbit (ensuring high-latitude coverage), and, last but not least, the free availability of a long-term data archive, which also promoted the development of algorithms for the analysis of long multitemporal data. Multitemporal SAR images refer to a series of images that are acquired over the same scene at different time intervals and collected in a 3-dimensional (3-D) stack (range, azimuth, and time). They are typically acquired by the same sensor and, therefore, share the same sensor features, e.g., impulse response, bandwidth, and operating frequency. Additionally, the images of the stack are typically acquired under the same acquisition mode (and therefore exhibit the same spatial resolution), polarization, viewing angle, and over multiple passes on the same orbit that make the whole image stack co-registered (with accuracy depending on the orbit stability). Accordingly, a multitemporal SAR image can be regarded as a collection of multiple snapshots of the same scene under the same acquisition and illumination conditions. Therein, variations of the backscattered energy along the time coordinate are solely related to temporal changes in the scene reflectivity. Such changes can be due to numerous factors, e.g., soil moisture content variation, scene dynamics and movements (e.g., sand dunes, sea, and mobile objects), vegetation growing, and human activity.

As a matter of fact, long time series spanning over years are available and allows for collecting accurate and comprehensive information in both spatial and temporal domains. Indeed, numerous applications rely on the processing of multitemporal SAR images, e.g., surface deformation analysis [2], information retrieval [3], [4], change detection [5], [6], agricultural areas and grasslands monitoring [7], [8], waterbodies monitoring [9], [10], RGB data formation [11], and land-cover classification [12]–[14].

Notwithstanding, SAR imagery interpretation by human users and processing by automatic algorithms are affected by speckle noise which is due to the coherent combination of echoes from sub-pixel scatterers [15]. SAR image denoising (also referred to as despeckling) is, therefore, a key enabling factor that has driven intense research in the recent past [16]. However, most efforts have been dedicated to reduce speckle effects in single-band (i.e., 2-D range-azimuth) SAR imagery, for which several approaches have been developed so far, including classical local adaptive filtering [17], [18], homomorphic approaches [19], [20], Bayesian estimators [21], [22], and the more effective resolution-preserving wavelet-based methods [23], [24], and nonlocal mean algorithms [25]–[27].

Conversely, multitemporal SAR despeckling is gaining an increasing interest only in the recent past, likely because of the (much) larger computational complexity required by multitemporal SAR image processing. As a result, the literature dealing with multitemporal SAR despeckling is increasing [28]–[33]. Such approaches will be discussed in detail in the following of this work.

Despite the key role played by despeckling algorithms in improving SAR imagery readability and information retrieval accuracy, few efforts have been focused on the investigation of the filtering capabilities in a wide variety of contexts and on the assessment of despeckling performance on an objective and standardized basis. Actually, a benchmarking framework has been presented in [34], where five canonical scenes (homogeneous, topography, squares, corner reflector, and building) have been defined and made publicly available along with ad hoc quality measures in order to provide a quantitative and replicable evaluation of the despeckling capabilities in several respects, such as edge smoothing, speckle reduction, texture, and target feature preservation.

Taking cues from the work in [34], which is designed for single SAR images, we here extend that analysis to multitemporal SAR despeckling, thus defining new canonical scenes and performance parameters for the evaluation of filtering quality under both stationary and time-varying reflectivity.[1]

The remainder of this work is as follows: Section II points out the motivations of our work and the need for a simulation approach for an objective assessment of the performance of multitemporal despeckling filters. Section III presents a brief overview of despeckling benchmarking for single-channel filters and the main despeckling approaches and algorithms

for multitemporal SAR images. The canonical scenes and the corresponding performance measures are described in Section IV, whereas Section V presents and discusses the experimental results obtained by applying the proposed benchmarking framework to state-of-the-art multitemporal SAR despeckling algorithms. Finally, main conclusions and guidelines for future research activities are briefly outlined in Section VI, where Table VII summarizes all the introduced quality metrics and their target values for each test case.

II. MOTIVATIONS

In classical SAR signal processing and most SAR applications, speckle is treated as noise, i.e., as a stochastic process distorting an underlying clean signal conveying the actual information to be retrieved. Accordingly, the objective of despeckling techniques is to remove the speckle component from the noisy signal by designing proper estimators of the clean signal. However, from a physical viewpoint, there is no actual separation between speckle and the clean signal as they combine in the noisy signal as a result of the electromagnetic (EM) interaction between the radar signal and the illuminated scene. Accordingly, a true (i.e., physically meaningful) speckle-free SAR image can only be obtained with the temporal multilook of an infinite number of independent and stationary over time single-look images. This is certainly not viable with real-world SAR imagery. Therefore, it is not possible to obtain reliable (for despeckling performance assessment purposes) actual speckle-free images.

Additionally, a comprehensive and quantitative assessment of the performance of a despeckling algorithm relies on the evaluation of proper measures of similarity between the filtered time series and the speckle-free (reference) one. Such quality indicators, referred to as full-reference measures in the despeckling community, provide quantitative information about the capabilities of the filter to reject speckle noise while retaining the specific features of the clean image.

As a result, for a meaningful, fair, and objective comparison of available methods for multitemporal SAR despeckling, the availability of speckle-free images is of key relevance. Conversely, using real data, only no reference measures could be adopted and, therefore, limited insights into the despeckling chain might be gathered. Indeed, a common practice in the SAR despeckling community for the evaluation of full-reference measures using real-world images relies on the generation of noisy images by superimposing the speckle component on real optical images. This approach has two severe drawbacks. First, SAR and optical images are very different under many respects including, acquisition geometry, signal processing, data distribution, and, last but not least, EM scattering. Indeed, scene response and EM scattering phenomena at microwaves, where SAR systems typically operate, are different from those in the visible spectrum (e.g., multiple-bounce scattering and the presence of strong and persistent scatterers in urban areas); second, in SAR imagery the overall noisy signal, including speckle, undergoes the SAR processing chain, and, therefore, both the clean signal and the speckle component are filtered with the SAR impulse response and then focused. Actually, the superposition of

---

[1]To ensure the full reproducibility of the experiments as well as the applicability of the proposed benchmarking tool to any multitemporal SAR despeckling algorithm, all data and scripts are publicly available for download at http://wpage.unina.it/alessio.disimone/download/download.htm.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

DI MARTINO *et al.*: BENCHMARKING FRAMEWORK FOR MULTITEMPORAL SAR DESPECKLING

3

speckle over an already focused image prevents obtaining appropriate spatial statistics. The free availability of some SAR data, e.g., Sentinel-1, is encouraging the generation of reference speckle-free images by temporal multilook of long time series (as in [33]), which requires stability of the scene features along the whole observation time period and is therefore poorly reliable for despeckling performance assessment purposes.

For the purpose of despeckling performance assessment, the simulation approach, despite its limits in reproducing real SAR measurements, offers some major advantages. First of all, a reliable reference image can be obtained by averaging an arbitrary large number of independent images, thus approximating speckle-free data with arbitrary accuracy. Additionally, a SAR simulator allows for generating realistic SAR signals in controlled environments, thus avoiding spurious unwanted perturbations taking place over the scene that might affect an accurate evaluation of the quality metrics. This is of crucial importance in the multitemporal case, where scene characteristics must be controlled over multiple acquisitions. Last but not least, the simulation approach enables the investigation of different despeckling features (e.g., speckle suppression in homogeneous areas, edge, and point target preservation) separately from each other by defining proper canonical scenes, something that is hardly achievable in real imagery. Just as an example, a reliable evaluation of the capability of the filter of retaining sharp edges, typically measured through the edge smearing (ES) parameter, should be kept separated by residual speckle effects [34]. Accordingly, for the evaluation of the ES parameter, several image profiles of the edge are averaged altogether in order to reduce intensity variations due to residual speckle that, otherwise, might lead to spurious variations of the quality metric and, then, to an incorrect interpretation of the filter behavior. The averaged profile can then be compared with the clean one. Such an averaging procedure requires the presence of a well-contrasted, sufficiently long edge which must be also aligned with one of the image axes. This is easily viable through the simulation approach. Another example is the evaluation of the speckle rejection in homogeneous areas. Hence, an accurate evaluation of the equivalent number of looks (ENL) requires the presence of a sufficiently wide area with homogeneous scattering properties so that amplitude image samples are Rayleigh-distributed. While this is straightforward to accomplish in simulated data with high reliability, it is much more complicated on real data.

The SARAS simulator [35] adopted in our approach is based on the evaluation of the EM scattering from the illuminated surface and accounts for both geometric (small- and large-scale surface roughness) and EM (complex dielectric constant) parameters of the illuminated scene. Complex scattering phenomena, such as multiple bounces from the building, are modeled as well. Accordingly, it is able to generate SAR images of arbitrary scenes, within the validity limits of the adopted scattering models. The reliability and effectiveness of the SARAS simulator have been assessed both quantitatively and qualitatively via comparisons with actual SAR imagery in a wide set of scenarios, including built-up areas [36], natural surfaces [37], [38], and ocean [39].

III. RELATED WORKS

A. Benchmarking Framework for Single-Image SAR Despeckling

The huge number of scientific contributions published in the last two decades on the topic of SAR speckle filtering has not been always followed by the simultaneous development of appropriate performance evaluation frameworks. Actually, in the last years only a few works focused on quality assessment of SAR despeckling filters [34], [40], [41]. In particular, in [34] a complete benchmarking framework for SAR despeckling was proposed. It is based on the definition of five meaningful canonical cases, which constitute the minimum set of cases up to the task of assessing the despeckling performance with reference to some main features, namely speckle reduction power, ability to tell apart speckle from texture, contour preservation, radiometric preservation, and preservation of features related to man-made structures. To obtain appropriate figures about these aspects the following test images were selected: 1) a flat region with constant EM parameters (Homogeneous), 2) a region with constant EM parameters but nonflat orography [digital elevation model (DEM)], 3) four flat regions with different EM parameters separated by straight lines (Squares), 4) a corner reflector placed on a homogeneous background (Corner), and 5) an isolated building placed on a homogeneous background (Building). For each of these five canonical cases 512 speckled images were simulated, using the physical-based SAR simulator described in [35]. Indeed, these 512 images can be averaged in order to obtain an image with a very high number of looks, to be used as a reference for performance assessment. In Fig. 1, one of the single-look realizations is shown for each canonical case, along with the respective reference image.

Appropriate full-reference and no-reference quality measures were also introduced in [34] for each of the five selected canonical scenes. They are defined and briefly discussed in the following of this section, where z stands for the noisy image, x is the speckle-free image, and $\hat{x}$ is the filtered one, all in intensity format. The interested reader is referred to [34] for more details regarding all the mentioned measures and test cases.

1) Homogeneous: The speckle reduction power is evaluated on the Homogeneous test case using the ENL and the despeckling gain (DG), the latter providing information about mean-square error (MSE) reduction offered by the despeckling filter. They are defined as follows:

$$\text{ENL} = \frac{\mu^2}{\sigma^2} \tag{1}$$

$$\text{DG} = 10 \log_{10}\big(\text{MSE}(x, z)/\text{MSE}(x, \hat{x})\big) \tag{2}$$

where μ and σ stand for the estimated mean and the standard deviation of the intensity image, respectively. In addition, the Homogeneous test case is used to evaluate bias indicators, i.e., the mean of image (MoI), the mean of the ratio image (MoR), and its variance (VoR), the latter conveying information on whether the speckle is insufficiently filtered. In MoR and VoR, the ratio image is defined as the ratio of the noisy and filtered image. This is a well-established way to assess
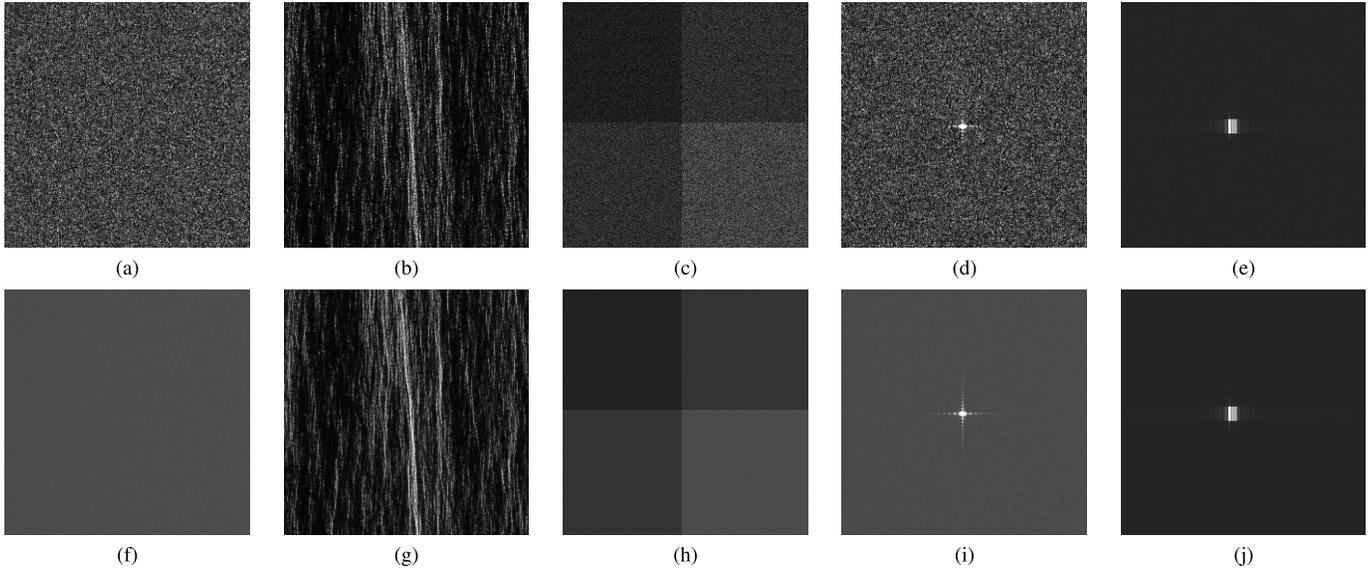
Fig. 1. Test cases for single-image despeckling algorithm benchmarking. (a)–(e) Single-look image. (f)–(j) 512-look reference image. (a) and (f) Homogeneous (256 × 256). (b) and (g) DEM (512 × 512). (c) and (h) Squares (512 × 512). (d) and (i) Corner (256 × 256). (e) and (j) Building (256 × 256).

despeckling performance, also adopted in [42] and [43] for the evaluation of the performance of single-channel despeckling filters. As a matter of fact, a good speckle filter is expected to offer large ENL and DG values as well as to preserve the MoI and to provide unitary MoR and VoR for single-look SAR image filtering, whereas for multilook SAR images the VoR should be equal to the inverse of the number of looks.

2) Digital Elevation Model (DEM): The ability to tell apart speckle from texture is evaluated on the DEM case through the joint use of DG and coefficient of variation, $C_x = \sigma/\mu$, which considers the region heterogeneity. Good texture preservation is achieved if C_x estimated on the filtered image is close to the reference value. Bias indicators MoI, MoR, and VoR are evaluated on the DEM test case as well.

3) Squares: Contour preservation is evaluated on the Squares image introducing appropriate ES and figure of merit (FOM) indexes. The ES indicator is evaluated as the weighted square error between the filtered and the clean edge profiles (EPs), i.e., as

$$\text{ES} = \int g(t - t_0)(\text{EP}_{\hat{x}}(t) - \text{EP}_x(t))^2 \, dt \tag{3}$$

where $g(\cdot)$ is a Gaussian kernel centered at the edge location t_0, and $\text{EP}_{\hat{x}}(\cdot)$ and $\text{EP}_x(\cdot)$ stand for the filtered and clean EPs, respectively. Although ES is more related to the radiometric preservation of the edge, FOM quantifies the reliability of the filtered image for subsequent applications, such as edge detection and classification. Indeed, it is a full-reference measure of the similarity between the outputs of an edge detector applied to both the despeckled image and the reference one. More specifically, it is defined as

$$\text{FOM} = \frac{1}{\max(n_d, n_r)} \sum_{i=1}^{n_d} \frac{1}{1 + \gamma \, d_i^2} \tag{4}$$

where n_d and n_r are the number of detected edge pixels on the despeckled and reference images, respectively; d_i is the Euclidean distance between the ith pixel of the detected edge and the closest reference one; γ weights the edge displacement. Good preservation of edges requires an ES value as small as possible and FOM close to one.

4) Corner: Radiometric preservation is evaluated on the Corner image defining meaningful intensity contrast measures, namely C_{NN} and C_{BG}. They are defined as

$$C_{\text{NN}} = 10 \log_{10} \frac{x_{\text{CF}}}{x_{\text{NN}}}, \tag{5}$$

$$C_{\text{BG}} = 10 \log_{10} \frac{x_{\text{CF}}}{x_{\text{BG}}} \tag{6}$$

where x_{CF} is the intensity measured in the corner reflector site, x_{NN} is the average intensity computed in the eight-connected nearest neighbors, and x_{BG} is the background average intensity. An accurate filtering of the point target is characterized by performance values as close as possible to the values measured on the reference image.

5) Building: Finally, man-made structure preservation is assessed on the Building test case introducing both an intensity contrast measure C_{DR} and a building smearing figure (BS). The C_{DR} parameter measures the radiometric precision of the filter by evaluating the contrast between the average intensity of the double-scattering line x_{DR} and the average background intensity x_{BG}. It is, therefore, defined as

$$C_{\text{DR}} = 10 \log_{10} \frac{x_{\text{DR}}}{x_{\text{BG}}}. \tag{7}$$

The BS parameter is defined as

$$\text{BS} = \int \Pi\left(\frac{t - t_0}{T}\right) |\log_{10}(\text{BP}_{\hat{x}}(t) + \epsilon) - \log_{10}(\text{BP}_x(t) + \epsilon)| \, dt \tag{8}$$

where ϵ is a small number and $\Pi(\cdot)$ denotes the normalized boxcar function which selects a limited range profile of spatial

length T and centered in the double-reflection range t_0. It measures the preservation of the radiometric building range profile by computing the log-difference between the intensity building range profiles of the filtered $BP_{\widehat{x}}$ and the reference image BP_x, where both profiles are averaged along the azimuth direction. For a reliable preservation of the double-scattering mechanism, it is required that C_{DR} is as close as possible to the reference value and BS as small as possible.

Since the perfect filter is hard to obtain, it is clear that all the mentioned features of the ideal filter are subject to more or less severe trade-offs in actual SAR despeckling filters. In this context, thanks to the framework of [34], it is possible to capture the different capabilities of the filters in order to identify the best filter for each specific application.

B. Overview on Multitemporal SAR Despeckling

Speckle reduction in single SAR images has been the subject of intensive research in the last four decades, during which diversified approaches have been investigated and developed, e.g., spatial, wavelet-based, and nonlocal filters [16]. Conversely, far less efforts have been dedicated to the development of despeckling algorithms for multitemporal SAR data.

The simplest choice is the pure temporal multilook where the multiple intensity bands are averaged pixel-wise to produce a single filtered image. Speckle noise variance is reduced by a factor not larger than the number of images and depending on the correlation among the bands. Despite its simplicity, the multilook filter is the optimum (unbiased minimum-variance) estimator for stationary SAR data and temporally decorrelated speckle (image values at any location are independent and identically distributed) [15]. Additionally, it preserves the spatial resolution of the data set as filtering is carried out only along the time coordinate. The development of more advanced despeckling algorithms for SAR time series dates back to the late 1990s and was driven by particular thematic applications, e.g., tropical forest monitoring, wherein forest/nonforest classification required accurate estimation of the backscattering coefficient value [28], [32].

De Grandi *et al.* [28] proposed a simple, yet effective, algorithm for reflectivity estimation using multitemporal data. The filtering of multiple images acquired on the same location at different time instants allowed to overcome the spatial resolution degradation caused by spatial filters. Filtering is performed through linear minimum-variance unbiased estimators, which, however, ensure optimum weighting only with stationary SAR data. Accordingly, a series of detectors and adaptive progressively smaller windows are used in [28] to locate inhomogeneities and, then, improve estimation of local statistics in the presence of nonhomogeneous areas (edges, fine structures, and man-made objects). Spatial inhomogeneities are discriminated with the normalized second moment. Temporal changes in reflectivity are retained by applying the adaptive windows algorithm separately to each band.

A simplified form of the minimum-variance unbiased estimator tailored to ERS data is also adopted in the UTA filter [32], where simplification comes out from the assumption of uncorrelated ERS bands. This is justified by the low (35 days) repeat-pass cycle of ERS satellites. The filter proposed in [32] is based on a two-step procedure: the first step is temporal pixel-based filtering consisting in a weighted averaging along the time coordinate only. Within this step, the local average intensity is estimated in order to produce unbiased results. Then, in order to achieve the ENL required for accurate forest mapping, an additional averaging in the spatial domain is applied in a 11×11 window. This brings to some unavoidable loss of spatial resolution of the filtered products. The performance in terms of ENL achievable with the UTA filter is discussed in [29].

The anisotropic nonlinear diffusion (ANLD) filter proposed in [44] was conceived primarily to improve segmentation tasks where it is mandatory to reduce speckle in order to extract meaningful segments. The filter takes advantage of the redundant information content available in multitemporal series and embeds the original data set in a family of derived images obtained by convolving the original image with the Gaussian kernel of varying variance. This approach can be equivalently formulated through the heat conduction or diffusion equation [44]. Anisotropic diffusion is enforced by letting the diffusion coefficient no longer constant in space and time. Such a strategy allows a good balance between speckle reduction in homogeneous areas and edge sharpness preservation. The ANLD filter is sensor-independent as it does not require any statistical description of SAR data and can be therefore applied to images acquired from different sensors.

Those denoising approaches of multitemporal images are local methods. Conversely, more recent filters have been developed within the nonlocal paradigm which, by taking advantage of image self-similarity and by introducing appropriate patch-based similarity measures, ensures effective speckle reduction with highly heterogeneous images. The block-matching 4-D (BM4D) filter presented in [45] is an extension of the BM3D filter to volumetric data. It implements the grouping and collaborative filtering procedures and applies them to cubes of voxels (i.e., group of pixels extracted from a 3-D image stack) instead of the traditional patches adopted in BM3D algorithms. As in BM3D, the spectrum of the grouped voxels is highly sparse and an effective separation of signal and noise is achieved by either thresholding or Wiener filtering. Finally, aggregation is performed after inverse transformation in order to place the voxels at their original location with a proper weight.

In [30], the authors propose an adaptation of the iterative weighted probabilistic patch-based (PPB) algorithm to multitemporal images. Filtering is carried out in a two-step framework (accordingly, the new algorithm was named 2S-PPB): the first step is temporal averaging which consists of a temporal multilook applied to stable pixels, i.e., those pixels not affected by temporal changes and well registered. The Kullback–Leibler divergence and a generalized likelihood ratio are used as similarity measures to discriminate stable pixels. The second step is a spatial filtering where a patch-based approach similar to PPB is applied in order to derive the final estimate of the reflectivity maps.

Temporal and spatial filtering are also exploited in [31] where the authors propose a multitemporal oriented version of the single-image SAR-BM3D algorithm [23]. The temporal averaging is performed through a minimum-variance unbiased estimator as proposed in [32]. However, the estimation of statistics is carried out in the framework of the nonlocal paradigm which is preferred to the sliding-window approach adopted in the UTA filter. Additionally, the correlation among bands is not neglected as done instead in [32] and affects the weighted average. This first step is called nonlocal temporal filter (NLTF) and operates only along the temporal axis, therefore not reducing the spatial resolution.

The NLTF algorithm is then exploited in a two-step procedure, where each step comprises the grouping, collaborative filtering, and aggregation typical of block-matching algorithms. NLTF is performed in the first step as a prefiltering of the grouped blocks in order to enhance the speckle rejection in the subsequent wavelet-shrinkage-based denoiser. Temporal denoising is instead omitted in the second step in order to preserve the information carried by noisy data. In this step, a Wiener filtering in the transform domain is carried out in order to produce the final filtered blocks.

Whereas most multitemporal denoising methods process the whole time series to produce the filtered images, the ratio-based filter (RABASAR) presented in [33] first forms a summary of the multitemporal image, the so-called superimage, which is obtained via an approach analogous to that in [30]. Then, any filtered band is obtained by using only the corresponding noisy image (rather than the whole time series) and the corresponding superimage. More specifically, a ratio image between the noisy image and the superimage is formed and then denoised through a nonlocal denoiser (called RuLoG) adapted to the ratio image Fisher distribution. The final despeckled band is then obtained by multiplying the denoised ratio image with the superimage. Temporal correlation of speckle or a limited data set would reduce speckle reduction in the superimage. To overcome this issue, a spatial filtering step is introduced to improve the quality of the superimage.

More recently, the deep learning paradigm has been experiencing an increasing interest in the SAR despeckling community boosted by the huge availability of images at no cost which allows for accurate training and learning of sophisticated image models. A comprehensive review of current trends is highlighted in [46]. Traditional deep learning approaches require the exploitation of noisy-clean pairs to train the neural network. However, as already discussed in Section II, this idea is hardly viable for SAR despeckling purposes due to the lack of real-world speckle-free data. To overcome this issue, self-learning methods are gaining much interest. An example is the recent NR-SAR-DL multitemporal filter [47], which is based on the exploitation of noisy reference images for the training of the network. More specifically, two time series of SAR images acquired on the same scene by the same sensor are used as image pairs, thus avoiding the requirement of clean data. Such an approach is based on the assumption that speckle distribution is the same on the noisy and reference data. Once the network is trained, both time series can be despeckled.

IV. PROPOSED BENCHMARKING FRAMEWORK

When moving from single SAR images to multitemporal data, time coordinate comes into play. Accordingly, proper procedures must be set up in order to analyze the despeckling capabilities under different temporal behaviors of the illuminated scene. In this work, we define canonical scenes with both stationary (i.e., constant in time) and time-varying reflectivity.

The stationary test cases are aimed at evaluating filter's accuracy in the five canonical scenes described in Section III-A as well as at analyzing convergence properties of the filter, i.e., the image quality behavior as a function of the number of bands. Conversely, the aim of the test cases with temporal changes is twofold: first, to investigate the sensitivity of the filter against such variations, i.e., to give an accurate and quantitative idea of the distortions and artifacts introduced in the despeckled data due to the reflectivity changes; second, its capability to keep as unaltered as possible the reflectivity perturbation and its temporal dynamic since this is crucial in specific monitoring activities and applications, such as in change detection algorithms.

All test cases with the exception of the Convergence one are defined assuming an arbitrary number of bands M. Indeed, as further discussed in Section V-C, the proposed benchmarking methodology offers some flexibility and can be easily adapted to specific SAR user needs, e.g., changing the total number of bands, the number of perturbed bands, or their location within the time series.

For any test case, synthetic and quantitative performance metrics are defined so as to provide an objective similarity measure between the actual despeckling algorithm and an ideal filter, whose behavior is defined for any selected scene.

In the following, $\mathbf{z}$ will denote the multitemporal image consisting of M single-look SAR images z_i, $i = 1, \ldots, M$; x_i is the ith time component of the reference image stack $\mathbf{x}$ which coincides with the noise-free reflectivity. Finally, $\hat{x}_i$ will denote the ith band of the despeckled time series $\hat{\mathbf{x}}$. All such quantities denote SAR images in intensity format.

It is also worth mentioning here that the whole benchmarking framework is based on the fundamental assumption of uncorrelated bands, i.e., it is assumed that $\rho_{ij} = 0$ $\forall i, j = 1, \ldots, M$ with $i \neq j$, where ρ_{ij} is the correlation coefficient between z_i and z_j. Even though this assumption is quite unrealistic in practical situations, it allows us to investigate filter performance under most favorable, despite ideal, conditions. Indeed, as it will be shown further in this work, for most filters, temporal correlation weakens speckle suppression capabilities along the time coordinate and, consequently, reduces the overall despeckling performance of the filter. Therefore, under the uncorrelation hypothesis, it is possible to achieve the upper bound in filter performance: for most filters, no better performance is reasonably expected by relaxing the correlation hypothesis. To better support our approach, a preliminary analysis of the time correlation effects on filtered image quality is carried out in Section V-D.

A. Stationary Reflectivity

In this section, we describe the test cases conceived for the evaluation of despeckling performance in presence of a time series of SAR images of scenes exhibiting stationary

reflectivity. This assumption, along with the uncorrelated band hypothesis, represents the most favorable condition for despeckling purposes as the time series consists of M uncorrelated images of the same scene, whose reflectivity is here denoted with x and is independent on the time coordinate. Accordingly, the collection of the M noisy image samples at site s, $\mathbf{z}(s) = [z_1(s), \ldots, z_M(s)]$, is a vector of independent and identically distributed (i.i.d.) random variables with mean $x(s)$ and variance $x^2(s)$ and, therefore, minimizes the variance of the reflectivity estimator. Under these assumptions, the temporal multilook represents the minimum-variance unbiased estimator and the output image variance will be reduced by M. Conversely, correlation and temporal changes of scene reflectivity will reduce the overall despeckling capabilities of the filter as the bands are no longer i.i.d. However, their effects might significantly differ. Indeed, temporal correlation is expected to impact the suppression of the speckle noise, whose variance will be reduced by a factor smaller than M and depending on the temporal correlation coefficient. Time-varying reflectivity, if not properly faced, might lead to artifacts in the filtered image stack resulting from averaging along the time coordinate. A preliminary analysis of the temporal correlation effects is carried out in Section V-D, leaving to future works the definition of ad hoc test cases including correlation, whereas test cases for time-varying reflectivity are presented and described in Section IV-B.

1) Single-Image Benchmarking Framework: The first part of the proposed benchmarking framework aims at providing a comprehensive set of well-assessed performance measures in presence of stationary multitemporal images of canonical scenes. To this end, we use the benchmarking framework for single-image despeckling filters presented in [34] and extend it to deal with multitemporal data. In particular, for each of the five canonical scenes (Homogeneous, DEM, Squares, Corner, and Building), the noisy time series is built by collecting M uncorrelated single-look images simulated according to the scene under study. The corresponding multitemporal reference image is then created by replicating the 512-look image along the M bands.

These five test cases allow for assessing despeckling capabilities of multitemporal filters under different respects, namely speckle reduction in homogeneous areas, radiometric distortions, and the preservation of textured areas, edges, and man-made structures. Indeed, they are analyzed in some works to support performance evaluation, see, e.g., [31].

For each scene, the corresponding quality indicators described in [34] and briefly recalled in Section III-A are evaluated in each band and the M measure values are averaged altogether in order to provide only one value for each measure and to reduce measurements variance. The only exception with respect to the performance measures introduced in [34] is the ES indicator which is here properly modified in order to better separate edge smoothing effects introduced by the filter from a potential bias, which is already captured by the MoI quality measure in the Homogeneous test case. The modified ES, called normalized edge smearing, ES*, is defined as

$$\text{ES*} = \int g(t - t_0)(\text{EP*}_{\hat{x}}(t) - \text{EP*}_x(t))^2 \, dt \qquad (9)$$

where $g(\cdot)$ is the Gaussian kernel, t_0 is the edge location, and $\text{EP*}_{\hat{x}}(\cdot)$ and $\text{EP*}_x(\cdot)$ stand for the filtered and clean edge profiles, respectively, which are normalized to their average intensity.

2) Convergence: Multitemporal data offer an additional degree of freedom for despeckling purposes, namely the number of bands (typically also referred to as components) forming the time series. As a matter of fact, the reduction of speckle noise affecting SAR imagery is basically an estimation problem, where the parameter to be estimated is the reflectivity of the illuminated scene. Accordingly, from basic estimation theory principles, it is reasonable to expect that, under the assumption of stationary reflectivity, i.e., reflectivity constant over time, the performance of the filter improves as the number of bands increases. Indeed, the output image stack of an ideal multitemporal despeckling algorithm should converge (under a specific convergence mode) to the reference image x (which is the same among all the bands, i.e., $x_i = x$, with $i = 1, \ldots, M$) as M approaches infinity. Accordingly, for an ideal filter:

$$\lim_{M \to \infty} d(\hat{\mathbf{x}}_M, x) = 0 \qquad (10)$$

where $d(\cdot, \cdot)$ denotes the selected convergence metric or distance. Condition (10) ensures that the filtering procedure leads to a consistent estimator of the reflectivity.

It is also reasonable to expect that better image quality is reached with a larger number of bands. However, computational complexity increases as well. Accordingly, a trade-off between speckle reduction and execution time must be considered in the choice of the number of bands. Notwithstanding, for a convergent filter, there exists a minimum number of bands over which no appreciable performance improvements are obtained. Accordingly, given a fixed threshold α_{TH}, we here define the lowest number of bands $M_{\alpha_{\text{TH}}}$ such that:

$$\frac{|d(\hat{\mathbf{x}}_M, x) - d(\hat{\mathbf{x}}_{M-1}, x)|}{d(\hat{\mathbf{x}}_{M-1}, x)} \le \alpha_{\text{TH}}. \qquad (11)$$

Condition (11) is a convergence criterion based on the relative distance difference, whereas the $M_{\alpha_{\text{TH}}}$ parameter measures the convergence rate of the filter. This parameter is of key relevance in practical situations as it provides useful guidelines for an efficient application of the filter. Indeed, the $M_{\alpha_{\text{TH}}}$ parameter tells us the minimum number of bands required by the filter to converge. This has some implications in the practical application of the filter. Hence, the computational complexity of the filter might increase significantly with the number of bands (typically at least linearly). The convergence analysis tells us that filtering more than $M_{\alpha_{\text{TH}}}$ bands altogether will not give appreciable improvements of the overall quality of the despeckled time series and will then lead to an efficient usage of the filter. In the case, a good compromise between computational burden and despeckling performance would be to split the time series in groups of $M_{\alpha_{\text{TH}}}$ bands.

In order to evaluate the $M_{\alpha_{\text{TH}}}$ parameter, the filter is applied to time series with increasing number of bands ranging from 2 up to the minimum between $M_{\alpha_{\text{TH}}}$ and 64, with a step of one image. Therefore, the filter is run $M_{\alpha_{\text{TH}}} - 1$ times and, however, not more than 63. After each run following the first, condition (11) is tested. If the condition is fulfilled,

convergence is reached and $M_{\alpha_{\text{TH}}}$ is evaluated; otherwise, the filter does not converge for the selected threshold α_{TH} and $M_{\alpha_{\text{TH}}}$ cannot be evaluated.

Different convergence modes and metrics may be selected also depending on the scene features to be analyzed. In this work, we chose the classical mean squared convergence as it allows for capturing the overall despeckling capabilities of the filter regardless of the peculiarities of the illuminated scene. Accordingly, in conditions (10) and (11), the distance $d(\widehat{\mathbf{x}}_M, x)$ between the filtered multitemporal image $\widehat{\mathbf{x}}_M$ and the reference image x is evaluated via the MSE averaged over the M bands (MSE_M):

$$ d(\widehat{\mathbf{x}}_M, x) = \text{MSE}_M(\widehat{\mathbf{x}}_M, x) = \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}\big[(\widehat{x}_i - x)^2\big] \quad (12) $$

where $\widehat{x}_i$ is the ith filtered band and $\mathbb{E}[\cdot]$ stands for the expectation operator.

The investigation of the filter convergence calls for additional comment. Indeed, condition (10) requires the evaluation of a limit operator, which is not possible in practice. Accordingly, in order to evaluate the despeckling capabilities in the limit of M approaching infinity, here we propose a simplified approach where the distance $\text{MSE}_M(\widehat{\mathbf{x}}_M, x)$ is evaluated for a fixed and large number of bands. Here, we choose $M = 64$ as it is much larger than typical values used in multitemporal SAR image processing.

Accordingly, condition (10) is replaced with the MSE_{64} parameter. It may happen that the filter does not fulfill (11) for the selected threshold α_{TH} after 64 bands. If this is the case, only the MSE_{64} parameter is computed.

Finally, convergence conditions are evaluated for the Homogeneous case [see Fig. 1(a) and (f)] as defined in [34].

B. Time-Varying Reflectivity

Whereas the test cases described in Section IV-A aim at evaluating the filter's performance with stationary images, the test cases introduced here are conceived to analyze the filter response in the presence of temporal variations of the reflectivity during the overall observation time.

As a matter of fact, a large number of factors influence the scene reflectivity and its representation on SAR data. These are related to the illuminated scene—complex dielectric constant, roughness (small-scale and large-scale), the sensor—operating frequency, bandwidth, polarization, and the acquisition geometry, i.e., viewing angle. Notwithstanding, in most applications dealing with multitemporal SAR data, it can be reasonably assumed that only scene parameters can lead to time variations of the reflectivity, see also the previous discussion in Section I. In passing, the impact of the atmosphere and the weather conditions can be put aside as it is mostly limited to the signal phase. Accordingly, it can be modeled as an additional speckle noise source and does not affect the surface reflectivity.

Reflectivity variations in time might be related to several factors, including vegetation growing, agricultural fields tillage, soil moisture variations, natural disasters, and human activity. Each factor might lead to specific features in the temporal trend of the scene reflectivity. Therefore, proper

modeling and simulation of all possible temporal variations observable in a multitemporal SAR image is meaningless and goes outside the scope of this work.

Following the rationale proposed in [34], here we define a limited set of scenes exhibiting canonical temporal variations of the reflectivity with the aim of providing meaningful test cases which are representative of real-world physical processes. As a matter of fact, most temporal changes observed in multitemporal SAR imagery could be grouped in a limited number of classes, such as the following:

1) Vegetation growing phenomena, which lead to a gradual variation of the backscattered signal strength over time.
2) Temporally- and spatially-localized backscattering intensity variations due to artificial targets appearing and disappearing during the whole observation time period, e.g., cars and ships, which cause strong variations (also identifiable as perturbations) in reflectivity localized in few pixels (depending on the size of the target and spatial resolution of the sensor) and limited to a single band.
3) Reflectivity variations that are localized in time but spread in space due to, for example, tillage activity, flooding, sudden variations of the soil moisture content.

As we are here interested in analyzing the effects due to temporal variations, we assume a homogeneous background reflectivity, apart from the latter case where edge preservation is important.

Accordingly, starting from the canonical scenes described in [34] for the assessment of single-image despeckling quality, we define the following canonical scenes:

1) Homogeneous varying (see Section IV-B1): a collection of SAR images with gradual intensity variation, each one with homogeneous EM parameters. Such a scenario is intended to analyze vegetation growing phenomena.
2) Homogeneous with Corner (see Section IV-B2): a corner reflector is injected in a single band of a time series of Homogeneous SAR images. The corner reflector is representative of strong point targets, e.g., cars.
3) Homogeneous with Building (see Section IV-B3): similar to the previous test case but with a building instead of the corner reflector. This test case is aimed at analyzing reflectivity perturbations characterized by strong double-scattering mechanisms, e.g., ships.
4) Squares perturbed (see Section IV-B4): a single band of a time series of stationary squares images is perturbed by changing the EM parameters of one flat region. Flooding and tillage activities fall within this test case.

For each test case, we define quality indicators for both unperturbed and perturbed bands, which are averaged on unperturbed and perturbed bands, respectively. Additionally, simulated images have size 256×256 and 512×512 for cases 1)–3) and for case 4), respectively. This is in accordance with the Homogeneous and Squares cases of [34].

The methodology used for simulating both single-look and reference multitemporal images along with corresponding quality indicators are detailed in Sections IV-B1 to IV-B4.

1) Homogeneous Varying: In this test case, the noisy time series is simulated as follows: M uncorrelated single-look SAR
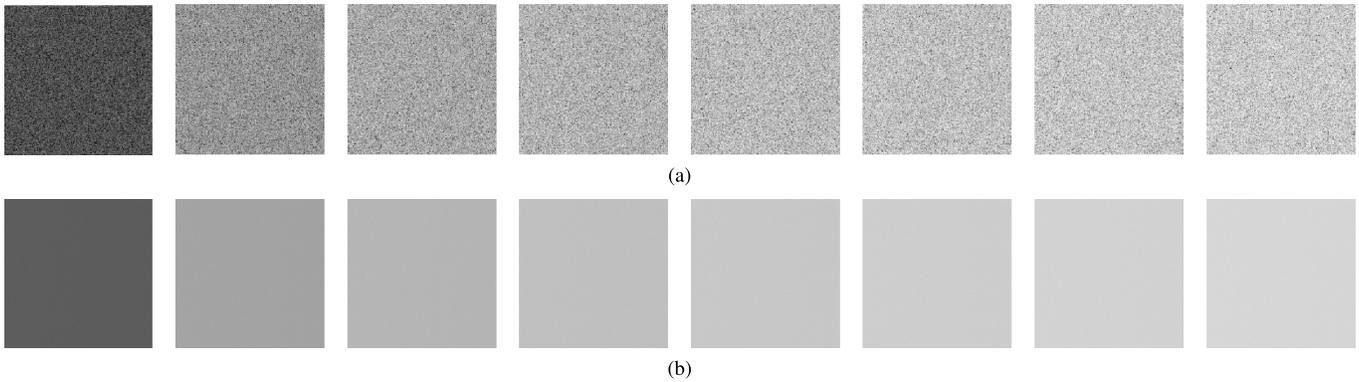
Fig. 2. Homogeneous varying data set. Band index increases progressively from left to right. (a) Single-look multitemporal image. (b) Reference multitemporal image. Images are displayed in logarithmic scale and share the same grayscale for a better visualization of the scene reflectivity variation over time.

images $z_{i,0}$ with $i = 1, \ldots, M$ are first generated according to the Homogeneous case described in [34]. Therefore, the illuminated scene for such images is a flat surface with homogeneous (i.e., constant in space) microscopic roughness, which is described via a fractional Brownian motion (fBm) stochastic process with the Hurst coefficient $H = 0.75$ and topothesy $T = 0.0625$ m, and homogeneous relative electrical permittivity $\varepsilon = 4$ and conductivity $\sigma = 0.001$ S/m [34]. Accordingly, the time series $\mathbf{z}_0$ consists of M i.i.d. homogeneous SAR images $z_{i,0}$. Hereafter we refer to $\mathbf{z}_0$ as the *original* time series.

Then, the ith component of the noisy multitemporal image z_i for the Homogeneous varying test case is obtained by modifying the intensity of $z_{i,0}$ according to the following expression:

$$z_i = \left[\frac{87.5}{M-1}(i-1) + 1 \right] z_{i,0} \quad i = 1, \ldots, M. \quad (13)$$

In the remainder of this section, we refer to the multitemporal image $\mathbf{z}$ defined in (13) as the *perturbed* time series.

The ith speckle-free image x_i is obtained through an analogous approach as follows:

$$x_i = \left[\frac{87.5}{M-1}(i-1) + 1 \right] x_1, \quad i = 1, \ldots, M \quad (14)$$

where x_1 is the reference image for each $z_{i,0}$ image, i.e., the reference image for the Homogeneous case in [34]. The perturbed time series and the corresponding reference multitemporal image are shown in Fig. 2(a) and (b), respectively.

Conditions (13) and (14) make average intensity of z_i and x_i increasing linearly through the time series with a factor ranging from 1 to 88.5 times the average intensity of the first band. Such an intensity dynamic is in agreement with data measured over agricultural areas [48]. Additionally, the choice of a linear variation of the average intensity can be justified by the following considerations. First, in most practical situations, seasonal variations of backscattering from agricultural fields can be linearized in a sufficiently limited observation time period, which, in our case, is the time period spanning from the acquisition of the first image to the acquisition of the last image of the time series. The interested reader is referred to, e.g., [48] for measured temporal profiles of backscattering

from rice canopy at various frequencies and incidence angles. Just as an example, in [48, Fig. 1], the seasonal variation of backscattering coefficient of rice plants in X-band and viewing angle of 35 degrees is reported. Even though the whole profile is far from being linear, it could be linearized in a narrower time period (up to several tens of days). Assuming an acquisition cycle of six days (as that offered by a two-sat constellation, e.g., Sentinel-1), it turns out that a time series of eight images are acquired in 43 days, which is compatible with the linear model.

Filter performance is measured through the indicators proposed in [34] for the Homogeneous case with some differences. In particular, for each band in intensity format, we first evaluate the MoI (corrected), MoI*, which is defined as the average value of the ratio image between the filtered band $\widehat{x}_i$ and the corresponding reference band x_i. This parameter is a modified version of the MoI introduced in [34] and takes into account for the average intensity variation among the bands due to the scaling factor in (13) and (14). The M MoI* measures are averaged in order to provide a synthetic indicator of potential overall (average) bias on the mean introduced by the filter. We call this performance parameter MoI*$_\mu$.

However, when dealing with multitemporal data, apart from an overall evaluation of the mean conservation, which is well captured by MoI*$_\mu$, it is also of key relevance to get an idea of the preservation of the average intensity temporal trend. Indeed, a filter might provide a good overall mean preservation, but introduce a significant distortion of the mean in each band. This behavior would not be completely captured by MoI*$_\mu$ and a second-order statistic is required. For this reason, we propose to measure also the standard deviation of the M MoI* measures, MoI*$_\sigma$. The MoI*$_\mu$ and MoI*$_\sigma$ parameters provide together a more comprehensive view of the mean preservation capability of the multitemporal filter. Speckle rejection capabilities and the accuracy of the filtered image are additionally measured via the DG and ENL parameters.

To further investigate the impact of the temporal variations of the scene reflectivity on the despeckling capabilities of the filter, it is useful to somehow relate the despeckling performance achieved with the perturbed time series $\mathbf{z}$ with that one achieved with the original time series $\mathbf{z}_0$. Indeed, the temporal variations in the perturbed time series might affect the averag-

ing along the time coordinate and then negatively impact the speckle reduction capability of the filter. This would result in an ENL lower than that achieved on the original multitemporal image, whose bands are i.i.d. Accordingly, a measure of such effects can be obtained by evaluating the ENL ratio (ENL_R), which is defined as

$$\text{ENL}_\text{R} = \frac{\text{ENL}(\widehat{x}_i)}{\text{ENL}(\widehat{x}_{i,0})} \tag{15}$$

where $\widehat{x}_{i,0}$ is the ith filtered band of the original multitemporal image.

2) Homogeneous With Corner: The perturbed noisy data set for this test case is created as follows: the first $M - 1$ images are simulated according to the Homogeneous case. Hereafter, they are referred to as the *unperturbed* bands. A point target is then introduced in the last image which is simulated according to the Corner test case, see Fig. 1(d). This is hereafter referred to as the *perturbed* band, while the whole multitemporal image is referred to as the *perturbed* time series. The corresponding reference multitemporal image is built by replicating the Homogeneous reference image [see Fig. 1(f)] over the first $M - 1$ bands, whereas the latter band is the Corner reference image, see Fig. 1(i). As will be better detailed in Section V-C, the temporal location of the perturbation does not appreciably affect the despeckling capabilities of the filters, and, therefore, it can be put on the last image without loss of generality.

Since the Homogeneous and the Corner images as originally presented in [34] exhibit different background intensity, a normalization of intensity data is performed so that the whole noisy data set has unitary background intensity average. The same normalization factor is then applied to the reference time series.

The presence of a bright feature in the perturbed band may affect despeckling capabilities in the unperturbed bands as multitemporal filters average along the time axis as well. Therefore, in such a scenario, it is reasonable to require that the multitemporal despeckling filter exhibits a sensitivity to the perturbation as low as possible in the unperturbed bands, while it should retain as accurate as possible the reflectivity variation features in the perturbed band.

Accordingly, proper metrics have to be defined for both the perturbed and the unperturbed bands. The capability of the filter to keep the corner features in the perturbed band is here evaluated with the corner measures C_NN and C_BG defined according to (5) and (6), respectively.

In order to assess despeckling capabilities in the unperturbed bands, it is useful to compare the performance achieved on the perturbed time series with that obtained with the original time series $\mathbf{z}_0$ defined in Section IV-B1. Here, we here propose a perturbation sensitivity (PS) measure on the ith unperturbed band ($i = 1, \ldots, M - 1$), which is defined as

$$\text{PS} = 10\log_{10}\frac{\text{MSE}(\widehat{x}_i, x_i)}{\text{MSE}(\widehat{x}_{i,0}, x_i)}, \quad i = 1, \ldots, M - 1 \tag{16}$$

where x_i is the Homogeneous reference image, $\widehat{x}_i$ and $\widehat{x}_{i,0}$ are the ith filtered band of the perturbed and original multitemporal images, respectively. To reduce measure variance, the PS

for the overall despeckled image stack is averaged among the $M - 1$ unperturbed bands.

The PS parameter defined in (16) is an indicator of potential artifacts introduced by the filter in the unperturbed bands due to the temporal change. The larger the PS, the higher the sensitivity of the filter to the corner perturbation. To better catch and measure the reflectivity distortions in the unperturbed bands, the PS in (16) is evaluated in a 5×5 region centered on the corner site.

Potential distortions arising far away from the corner reflector are measured with the ENL_R, see (15), averaged through the unperturbed bands. The ENL ratio will provide quantitative information on how the temporal change influences the speckle reduction capabilities of the filter in a homogeneous region. To avoid the influence of the artifacts potentially introduced in correspondence of the corner site, the ENL values in (15) are measured on an upper-left region not affected by appreciable temporal changes of the reflectivity.

3) Homogeneous With Building: This test case is conceived in a way similar to the Homogeneous with Corner test case. Accordingly, the first $M - 1$ images are Homogeneous images (unperturbed bands), whereas the last band is generated according to the Building test case [perturbed band, see Fig. 1(e)] as defined in [34]. Again, the noisy data set is normalized so that it exhibits unitary background intensity average. The same normalization factor is applied to the reference multitemporal image, which includes the Homogeneous reference image in the first $M - 1$ images and the Building reference image in the last band, see Fig. 1(j).

This test case is aimed at evaluating the capability of the filter to retain temporal changes due to multiple scattering phenomena and their impact on the filtered unperturbed bands. The first objective is fulfilled by evaluating the building measures C_DR and BS on the filtered perturbed band $\widehat{x}_M$. As for the effects of the temporal change on the unperturbed bands, we proceed as in the Homogeneous with Corner test case and evaluate the PS and the ENL_R.

4) Squares Perturbed: The perturbed multitemporal image for this test case is built as follows: the first $M - 1$ images of the single-look data set are generated according to the Squares case as illustrated in [34]. In the last band, the reflectivity of the lower-right square is changed by modifying its dielectric constant, which is set equal to that of the upper-left square. This change brings to a significant attenuation of the backscattered energy which makes the square appearing much darker than in the previous bands. Accordingly, such a scenario may be representative of a flooding event. Example images are shown in Fig. 3. The first (unperturbed) single-look band is shown in Fig. 3(a), whereas its corresponding reference image is shown in Fig. 3(b). The perturbed single-look and reference bands are depicted in Fig. 3(c) and (d), respectively.

As for the unperturbed bands, at first we measure the normalized edge smearing ES* averaged through the $M - 1$ unperturbed bands. Then, we evaluate the impact of the temporal change on the capability of the filter to preserve the sharp edges. To this end, we compare the edge preservation achieved on the perturbed time series with that obtained with the original time series, which consists of M i.i.d. Squares

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

DI MARTINO *et al.*: BENCHMARKING FRAMEWORK FOR MULTITEMPORAL SAR DESPECKLING 11
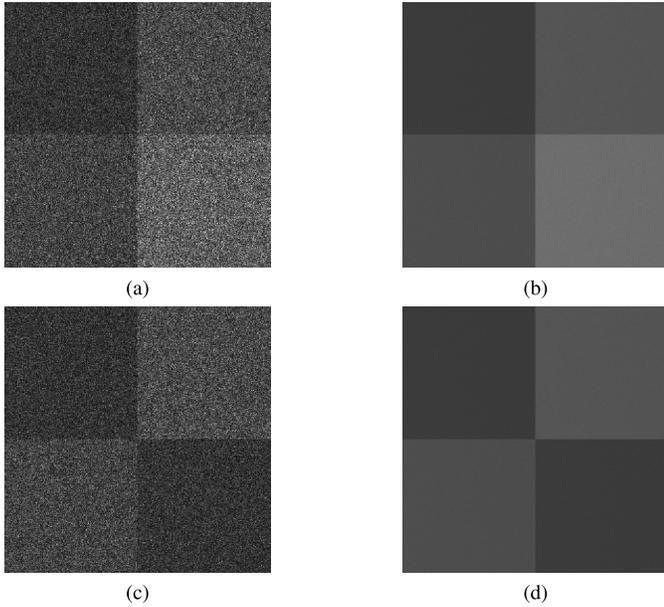

Fig. 3. Squares perturbed data set. (a) Single-look unperturbed. (b) Reference unperturbed. (c) Single-look perturbed. (d) Reference perturbed.

images simulated as in [34]. Accordingly, we define the edge smearing ratio ES_R for the ith band as

$$ES_R = \frac{ES^*(\widehat{x}_i)}{ES^*(\widehat{x}_{i,0})}, \quad i = 1, \ldots, M - 1 \qquad (17)$$

i.e., as the ratio between the ES* of the ith unperturbed filtered band $\widehat{x}_i$ and the ith original filtered band $\widehat{x}_{i,0}$. Among the different edges present in any band, we select the lower vertical edge as it is affected by the reflectivity temporal variation. By following the approach proposed in [34], multiple edge range profiles are averaged altogether to reduce measurement variations. Finally, ES_R is averaged among the $M - 1$ unperturbed bands to reduce measurement noise.

The normalized ES indicator is also adopted on the perturbed band to measure the filter capability to preserve the perturbation features. To this end, we measure the ES* on the same region used for the evaluation of ES_R.

V. EXPERIMENTAL RESULTS

In this section, we apply the benchmarking framework described in Section IV to a number of state-of-the-art multitemporal SAR despeckling techniques. This is basically done to provide, also by comparison with a visual inspection, the ability of the proposed quality indicators to catch satisfactorily the main limits and advantages of the denoising algorithms in presence of the defined canonical scenes.

The choice of the set of despeckling filters has been primarily driven by free availability of their software code. Accordingly, for each selected filter, we report the website containing the algorithm source codes used in this work. Another important criterion was also the reputation of the filter within the scientific community which led us to prefer most highly-cited works. Additionally, we selected filters operating either in the time domain only or also including a spatial

filtering step. This allowed us to shed light on the potentials and limits of both approaches. In the end, we selected the following algorithms:

1) Temporal multilook. Its implementation is straightforward and has been accomplished by the authors.
2) UTA [32] (http://www.grip.unina.it/).
3) NLTF [31] (http://www.grip.unina.it/).
4) MSAR-BM3D [31] (http://www.grip.unina.it/).
5) RABASAR [33] (https://www.charles-deledalle.fr/pages/).
6) BM4D adapted to speckle noise through the homomorphic approach (log-BM4D) [45] (http://www.cs.tut.fi/ foi/).

The set of filters comprises three time-domain (multilook, UTA, and NLTF) and three spatiotemporal (MSAR-BM3D, RABASAR, and BM4D) algorithms. The original BM4D filter presented in [45] is designed to face additive white Gaussian noise, and it is therefore adapted to multiplicative speckle noise by providing in input the logarithm of the noisy data set in intensity format. Such an operation causes a systematic bias on the mean which is equal to the opposite of the Euler–Mascheroni constant. It is compensated in the output image before taking its exponential.

All selected filters have been already described briefly in Section III-B. For each filter, we chose the default parameter settings proposed in the corresponding original work or in the software code.

Unless otherwise stated, all intensity test images are normalized to the average value of the corresponding intensity reference image, which, therefore, has unitary mean by definition. Additionally, by following the approach in [34], the clean image is created by temporal multilook of 512 single-look images. Additionally, for the multilook filter, the performance indicators are evaluated by replicating the multilook image M times. This is especially useful for measuring the parameters relevant to the time-varying test cases, where the reference image is not the same for all the bands. Finally, unless otherwise stated, we consider time series of $M = 8$ images. Recall that open-source codes and data are available at http://wpage.unina.it/alessio.disimone/download/download.htm.

A. Stationary Reflectivity

Here, we discuss the results obtained in presence of multitemporal SAR images of scenes with stationary reflectivity.

1) Single-Image Benchmarking Framework: This test case is of key relevance as it allows for quantitatively evaluating the despeckling capabilities of the filters under several respects, including the suppression of speckle noise in homogeneous areas, and the preservation of fine details, e.g., texture, edge, and strong scatterers. Synthetic quality indicators for the five canonical scenes are reported in Table I, while related images and graphs are shown in Figs. 4–11.

The Homogeneous test case (see Fig. 4) provides useful information on the capability of the filter to reduce speckle variance. The bias indicators MoI, MoR, and VoR reveal that a significant distortion of the average intensity is introduced

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12

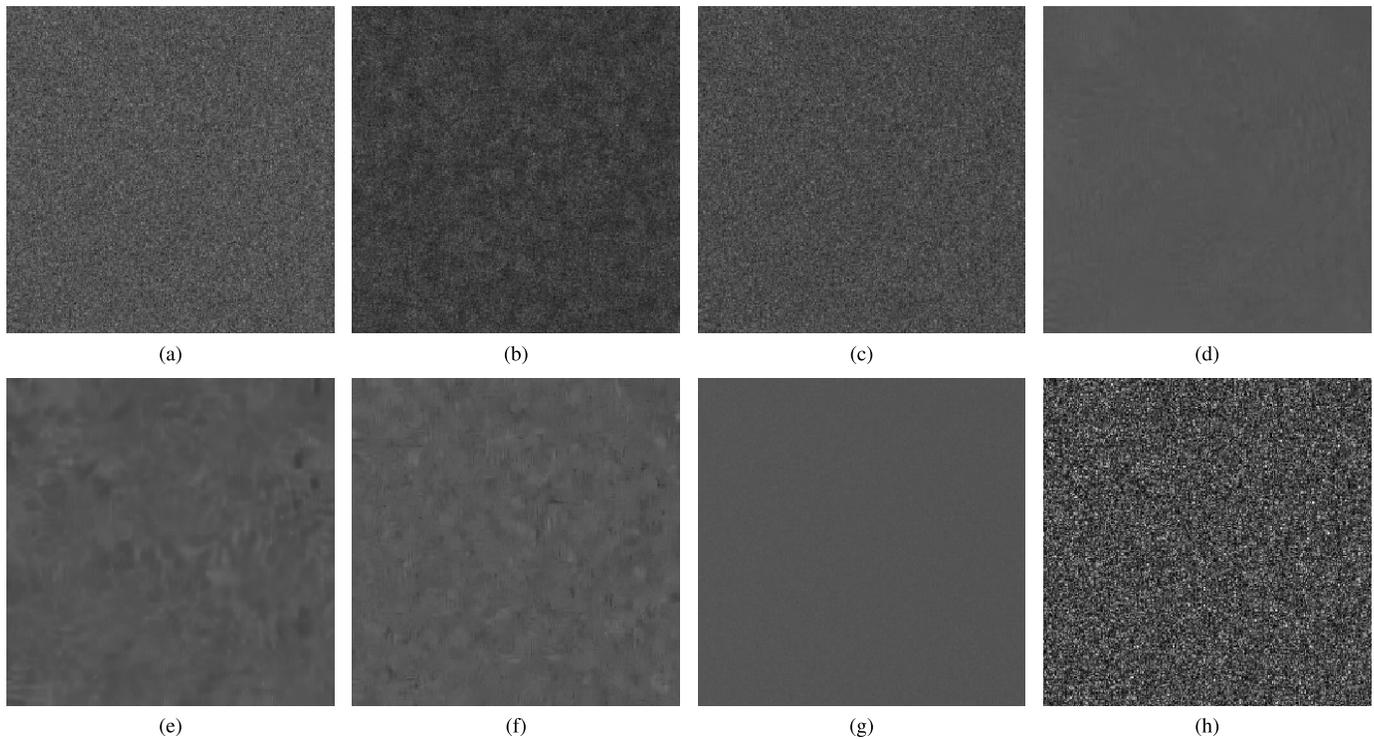IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Fig. 4. Results for the single-image benchmark Homogeneous test case (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.

TABLE I

MEASURES FOR THE STATIONARY TEST CASES

		Noisy	Reference	Multilook	UTA	NLTF	MSAR-BM3D	RABASAR	log-BM4D
HOMOGENEOUS	MoI	0.998	1.000	0.998	0.997	0.812	1.000	0.968	1.052
	MoR	-	0.998	1.000	1.000	1.239	0.990	1.009	0.917
	VoR	-	0.987	0.775	0.756	1.340	0.948	0.906	0.772
	DG	0	-	9.13	8.59	9.05	24.26	18.70	18.13
	ENL	1.01	436.97	8.09	7.15	7.39	428.24	86.18	99.86
DEM	MoI	1.003	1.000	1.003	1.001	0.882	1.014	0.632	0.959
	MoR	-	1.001	1.000	1.000	1.162	0.894	1.402	0.941
	VoR	-	0.999	0.779	0.770	0.918	0.572	2.633	0.876
	C_x	3.54	2.40	2.57	2.59	3.07	2.55	2.19	2.46
	DG	0	-	9.15	7.72	3.60	7.70	5.44	8.44
SQUARES	ES* (up)	0.0154	-	0.0007	0.0029	0.0026	0.0022	0.0018	0.0122
	ES* (down)	0.0173	-	0.0017	0.0033	0.0021	0.0048	0.0024	0.0236
	FOM	0.763	0.926	0.921	0.797	0.900	0.914	0.783	0.861
CORNER	C_{NN}	7.77	7.75	7.77	7.77	7.77	7.76	7.60	7.80
	C_{BG}	36.51	36.56	36.51	36.51	37.41	37.52	35.95	38.78
BUILDING	C_{DR}	65.90	65.90	65.90	65.91	66.80	66.94	66.24	68.18
	BS	0.092	-	0.048	0.052	0.066	3.411	0.614	8.099

by only NLTF. The measured bias is in accordance with that reported in the original work [31]. The temporal multilook filter exhibits an excessive speckle residual due to the relatively low number of bands averaged as it is witnessed by the VoR lower than one and by the low ENL and DG values achieved. More, in general, filters operating in the time domain only, namely multilook, UTA, and NLTF, provide poorer speckle suppression compared to spatiotemporal filters, MSAR-BM3D, RABASAR, and log-BM4D, which can exploit spatial content redundancy to boost denoising accuracy. This is

well described by speckle suppression indicators, DG and ENL. Best speckle reduction is provided by MSAR-BM3D, followed by log-BM4D and RABASAR which offer similar results. In this scenario, a simple visual inspection of the output images will likely lead to a similar overall filter rank.

The DEM test case (see Fig. 5) is of key relevance due to the rich spectrum of the surface reflectivity. As a matter of fact, separation of the high-frequency spectrum from speckle noise might be a challenging task. In this case, a visual inspection might be much less informative as most filters seem to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

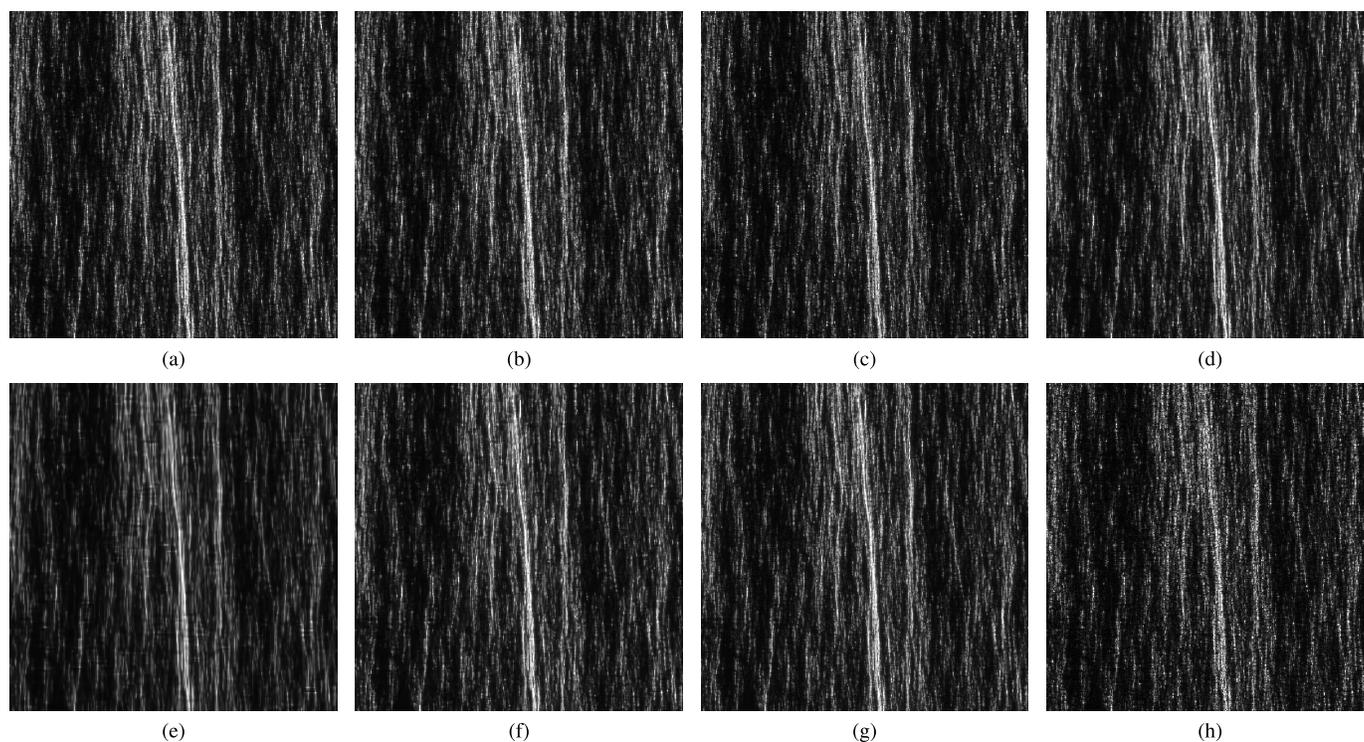DI MARTINO *et al.*: BENCHMARKING FRAMEWORK FOR MULTITEMPORAL SAR DESPECKLING 13

Fig. 5. Results for the single-image benchmark DEM test case (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.
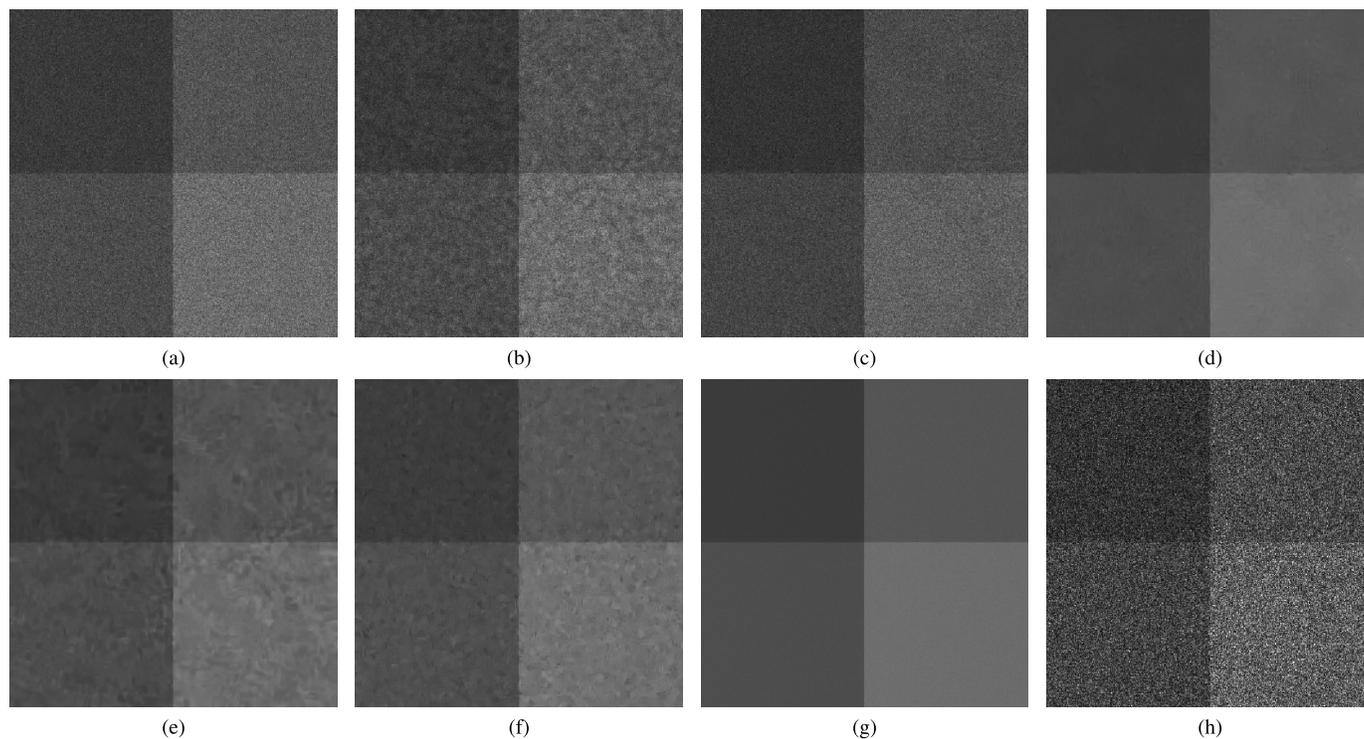

Fig. 6. Results for the single-image benchmark Squares test case (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.

exhibit similar despeckled image quality. The only exception is RABASAR, which leads to a visible oversmoothing of the image. This behavior is well captured by the VoR, which

is much larger than one, and by DG. Apart from NLTF, in this DEM test case, a severe mean bias is introduced by RABASAR, whereas the best and worst texture preservation,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

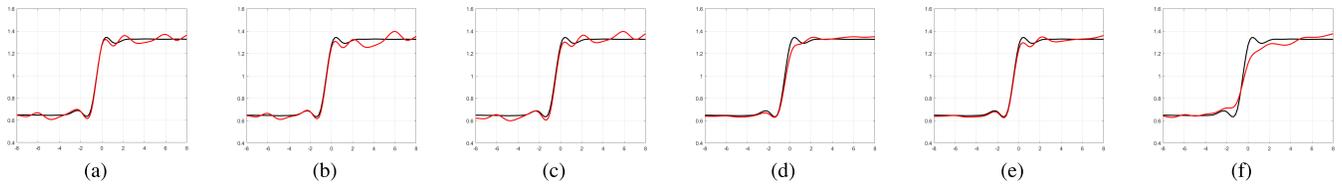14   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Fig. 7. Normalized (i.e., mean-bias corrected) edge range profile (lower edge) for the Squares test case. Despeckling filter (red line) and reference (black line). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D.
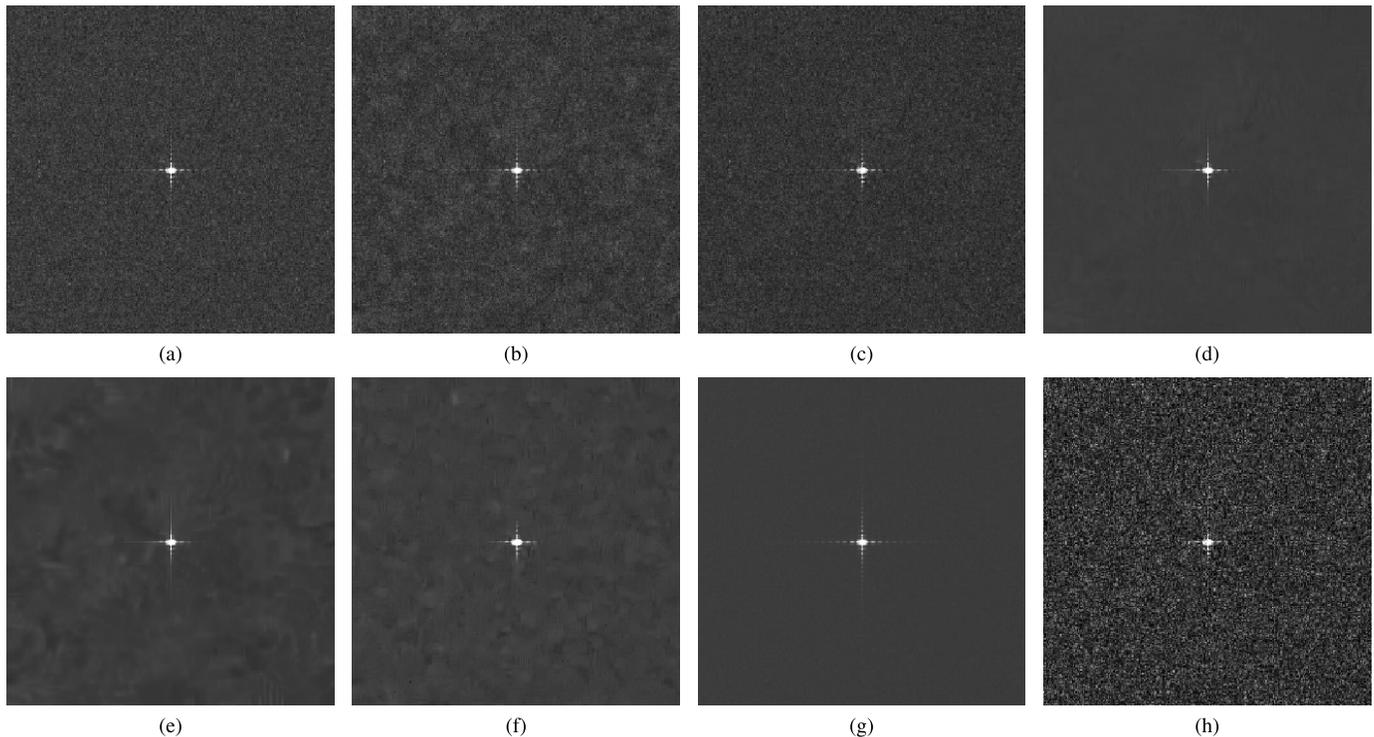

Fig. 8. Results for the single-image benchmark Corner test case (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.
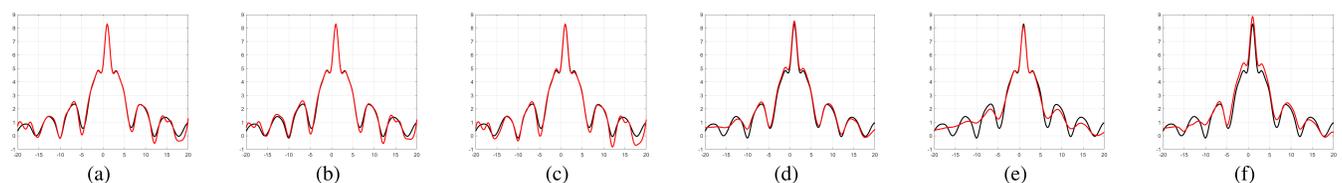

Fig. 9. Corner range profile in logarithmic scale for the Corner test case. Despeckling filter (red line) and reference (black line). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D.

captured by C_x, is offered by log-BM4D and NLTF, respectively. The spatial variability of the scene reflectivity on small scales reduces the image redundancy and, in turn, the effectiveness of nonlocal filters in suppressing speckle noise. As a result, for the DEM case, the multilook filter represents the best solution for speckle reduction, as witnessed by the DG indicator.

In the Squares test case (see Figs. 6 and 7), the benefits of the nonlocal approach on edge preservation are evident in the spatiotemporal filters. As can be seen in Fig. 6, all the tested nonlocal filters offer satisfactory preservation of all the

edges in the image. This is also confirmed by the ES* values reported in Table I and by the normalized edge profiles shown in Fig. 7. Among spatiotemporal filters, RABASAR offers the best edge preservation and provides ES* values comparable to UTA and NLTF, which, conversely, operate in the time domain only. However, the filtering in the spatial domain makes the spatiotemporal filters less effective in retaining the sharp edge with respect to the simple multilook filter which provides the best edge preservation capability in this stationary case.

In the presence of corner reflectors (see Figs. 8 and 9), such as trihedral structures, it is important to keep as unal-

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

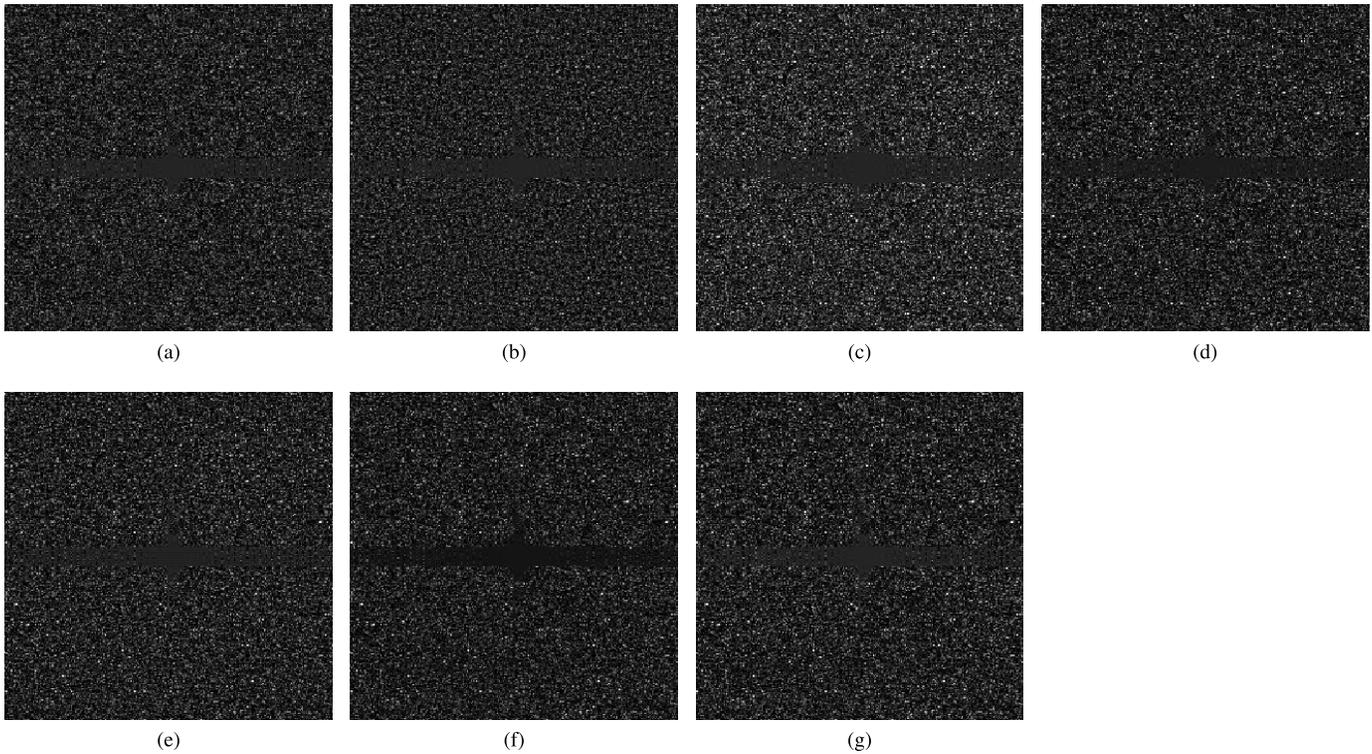DI MARTINO *et al.*: BENCHMARKING FRAMEWORK FOR MULTITEMPORAL SAR DESPECKLING
15

Fig. 10. Results for the single-image benchmark Building test case. Ratio image (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference.
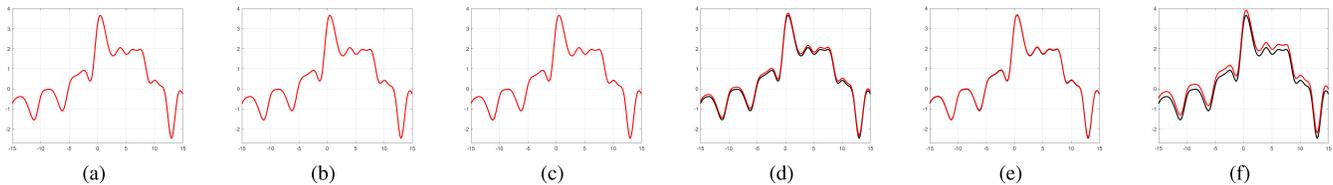

Fig. 11. Normalized (i.e., mean-bias corrected) building range profile in logarithmic scale for the Building test case. Despeckling filter (red line) and reference (black line). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D.

tered as possible, the strong scatterer response. Accordingly, an effective despeckling procedure will leave unfiltered the corner reflector image feature, which comprises the corner peak and the surrounding sidelobes. As is evident from Figs. 8 and 9, all the analyzed filters offer a reliable preservation of the sinc-squared shape of the corner reflector. This is well captured by the corner contrast measures C_{NN} and C_{BG} reported in Table I. Also in this case, best quality measures are provided by the multilook filter, whereas the log-BM4D offers worst performance due to a slight overestimation of the corner reflector response. Major differences among the filters are in the background region, where the same comments made for the Homogeneous test case apply, and, therefore, are not of interest in this test case.

For the Building test case (see Figs. 10 and 11), comments similar to the Corner test case do apply. Indeed, the building profile is preserved very well by all filters, see Fig. 11 and the building measures in Table I. Again, the log-BM4D slightly overestimates the building response as it is captured by the C_{DR} parameter and, even more, by BS. The negligible spatial redundancy of the building image feature makes the pure temporal filters, multilooking, UTA, and NLTF, perform better
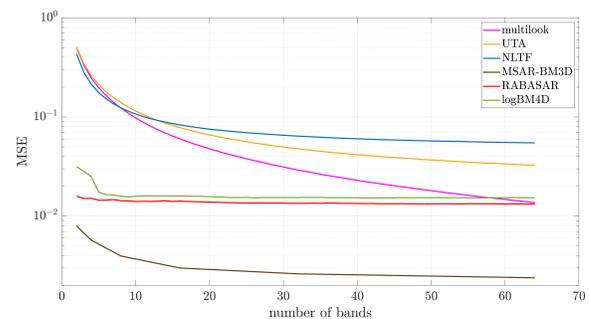

Fig. 12. MSE for convergence properties evaluation.

than spatiotemporal nonlocal filters. Again, the multilook filter ensures the best performance in terms of both building contrast measure and profile preservation.

2) Convergence: Synthetic measures relevant to the convergence properties analysis are reported in Table II, whereas Fig. 12 shows the average MSE defined in (12) as a function of the number of bands for the different filters analyzed here. The MSE_{64} parameter is evaluated according to (12) by applying the filter on a 64-band Homogeneous image.

TABLE II

MEASURES FOR CONVERGENCE ($\alpha_{TH} = 0.05$)

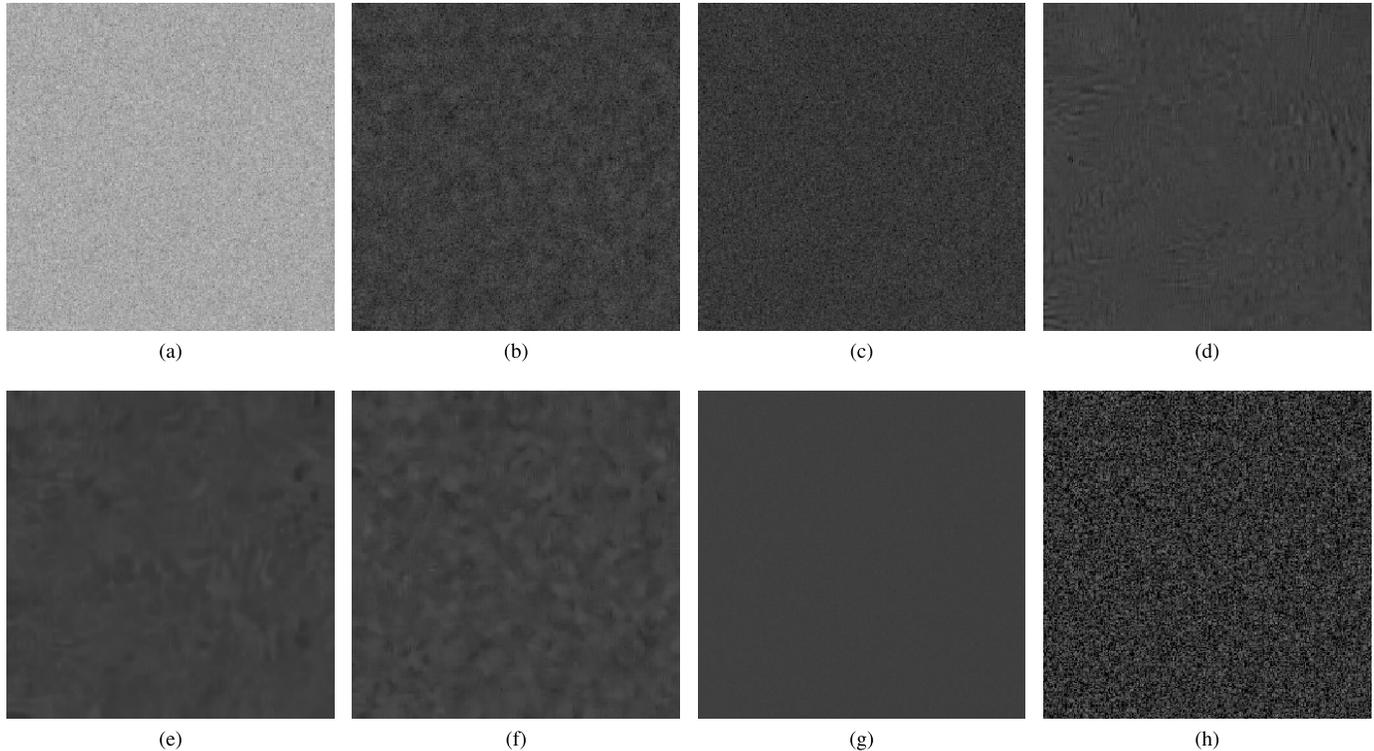|	Noisy	Reference	Multilook	UTA	NLTF	MSAR-BM3D	RABASAR	log-BM4D
MSE_{64}	0.999	-	0.0137	0.0325	0.0549	0.0024	0.0133	0.0153
$M_{\alpha_{TH}}$	-	-	21	16	12	N.A.	4	7

Fig. 13. Results for the Homogeneous varying test case (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.

Among the selected filters, the only exception is MSAR-BM3D that is currently implemented to work with a number of bands which is a power of two. Accordingly, for this filter, only the MSE_{64} parameter can be evaluated. Numerical results in Table II are obtained by setting $\alpha_{TH} = 0.05$. The highest convergence rates are measured by RABASAR and log-BM4D which converge with three and six bands, respectively. Lower convergence rates are achieved by filters averaging along the time coordinate only, due to the less information exploited by such filters. It is evident that MSAR-BM3D ensures the best performance in terms of MSE for any number of bands, see the MSE graph in Fig. 12. This is partially captured by the MSE_{64} performance measure which confirms the larger effectiveness of MSAR-BM3D with respect to the competitors analyzed. Worst MSE are provided by UTA and NLTF, due to their lighter smoothing.

Finally, it is worth mentioning that the results presented here are consistent with those presented in [49], where the convergence properties are measured on the Squares test case. The slightly higher convergence rate achieved by RABASAR is related to the larger threshold ($\alpha_{TH} = 0.1$) adopted there.

B. Time-Varying Reflectivity

In this section, we discuss the test cases relevant to scenes with temporal changes in reflectivity.

1) Homogeneous Varying: This test case is relevant in applications related to the monitoring of cultivated fields. The EM energy backscattered by such areas varies over time due to the growth of vegetation, whose EM response depends on the plant growth stage. In each stage, the reflectivity can be reasonably assumed homogeneous, i.e., constant in space. Accordingly, in this scenario, it is important that the filter reduces speckle noise without altering the reflectivity temporal trend of the illuminated surface. It is also interesting comparing the despeckling performance with those achieved on a time series acquired over a surface with homogeneous and stationary reflectivity.

Quantitative performance measures are listed in Table III, whereas Fig. 13 shows the first (darkest) band of the single-look, reference, and filtered images.

A pure temporal multilooking approach is obviously badly suited to scenarios with significant temporal changes in reflectivity. In this case, a severe distortion of the radiometric profile

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

DI MARTINO *et al.*: BENCHMARKING FRAMEWORK FOR MULTITEMPORAL SAR DESPECKLING 17

TABLE III

MEASURES FOR HOMOGENEOUS VARYING

	Noisy	Reference	Multilook	UTA	NLTF	MSAR-BM3D	RABASAR	log-BM4D
MoI*$_\mu$	1.001	1.000	6.712	1.000	0.814	1.007	0.974	1.040
MoI*$_\sigma$	0.0089	0	15.42	0.0089	0.0079	0.0178	0.0081	0.0130
DG	0	-	-1.12	8.54	9.07	22.83	18.25	15.19
ENL	1.001	454.1	5.68	7.05	7.32	379.31	75.12	40.00
ENL$_R$	0.996	1.04	0.702	0.986	0.990	0.891	0.876	0.405

TABLE IV

MEASURES FOR HOMOGENEOUS WITH CORNER

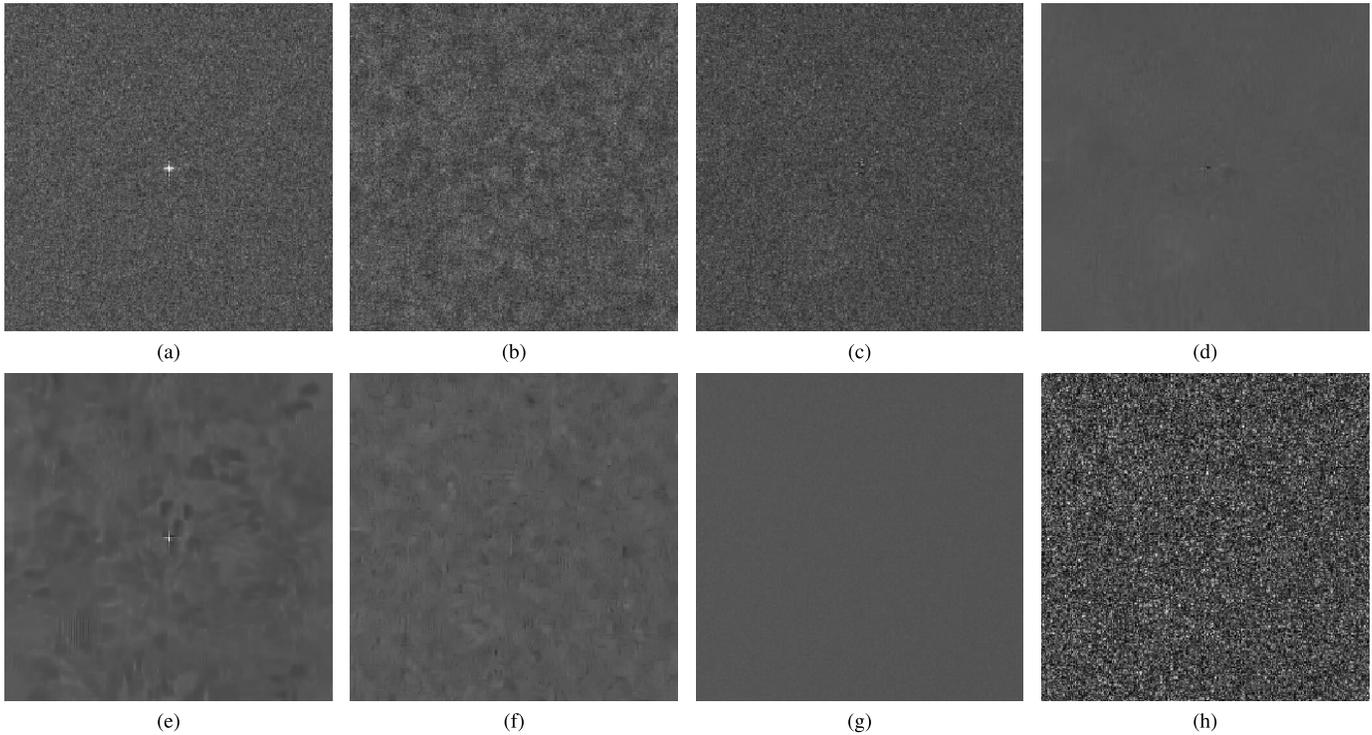|	Noisy	Reference	Multilook	UTA	NLTF	MSAR-BM3D	RABASAR	log-BM4D
C_{NN}	7.76	7.75	7.72	5.60	7.76	7.77	7.91	5.63
C_{BG}	36.49	36.56	27.51	27.99	37.36	37.08	37.16	27.84
PS	0	-	51.28	11.15	5.45	16.01	42.62	2.13
ENL$_R$	1.000	1.000	0.995	0.993	0.990	0.985	0.984	0.990

Fig. 14. Results for the Homogeneous with Corner test case (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.

is obtained with the multilook filter, see MoI*$_\mu$ and MoI*$_\sigma$. Conversely, more advanced approaches, including UTA and spatiotemporal filters, offer an accurate preservation of the reflectivity temporal profile, see MoI*$_\mu$ and MoI*$_\sigma$. The MoI*$_\mu$ indicator reveals that NLTF introduces a nonnegligible bias, which is similar to that introduced in the Homogeneous test case. However, such a bias can be related to a multiplicative factor as it can be inferred from the very low MoI*$_\sigma$ and could be easily compensated.

Finally, time-domain filters exhibit a negligible degradation of performance in terms of ENL with respect to the Homogeneous test case discussed in Section V-A1, see ENL$_R$. A significant ENL reduction is obtained with the multilook filter, due to the reflectivity variation which increases the variance of the averaged samples. A nonnegligible ENL reduction

is also experienced by spatiotemporal filters where largest ENL degradation is obtained by log-BM4D which here reaches ENL equal to 40.00 significantly smaller than 99.86 achieved in the homogeneous test case.

Similar performance degradation is measured on DG for all filters but RABASAR and log-BM4D which offer a minor sensitivity to temporal changes in terms of DG. This is much more evident for log-BM4D, whose degradation in DG is about 16%, much less of the reduction in ENL. It is also worth noting that the multilook filter leads to a negative DG, meaning that the average MSE is increased after filtering.

2) Homogeneous With Corner: In Figs. 14 and 15, we show band 1 (unperturbed) and band 8 (perturbed), respectively, for the noisy, reference, and the selected filters. Performance indicators for the perturbed band, C_{NN} and C_{BG}, and for the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

18   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
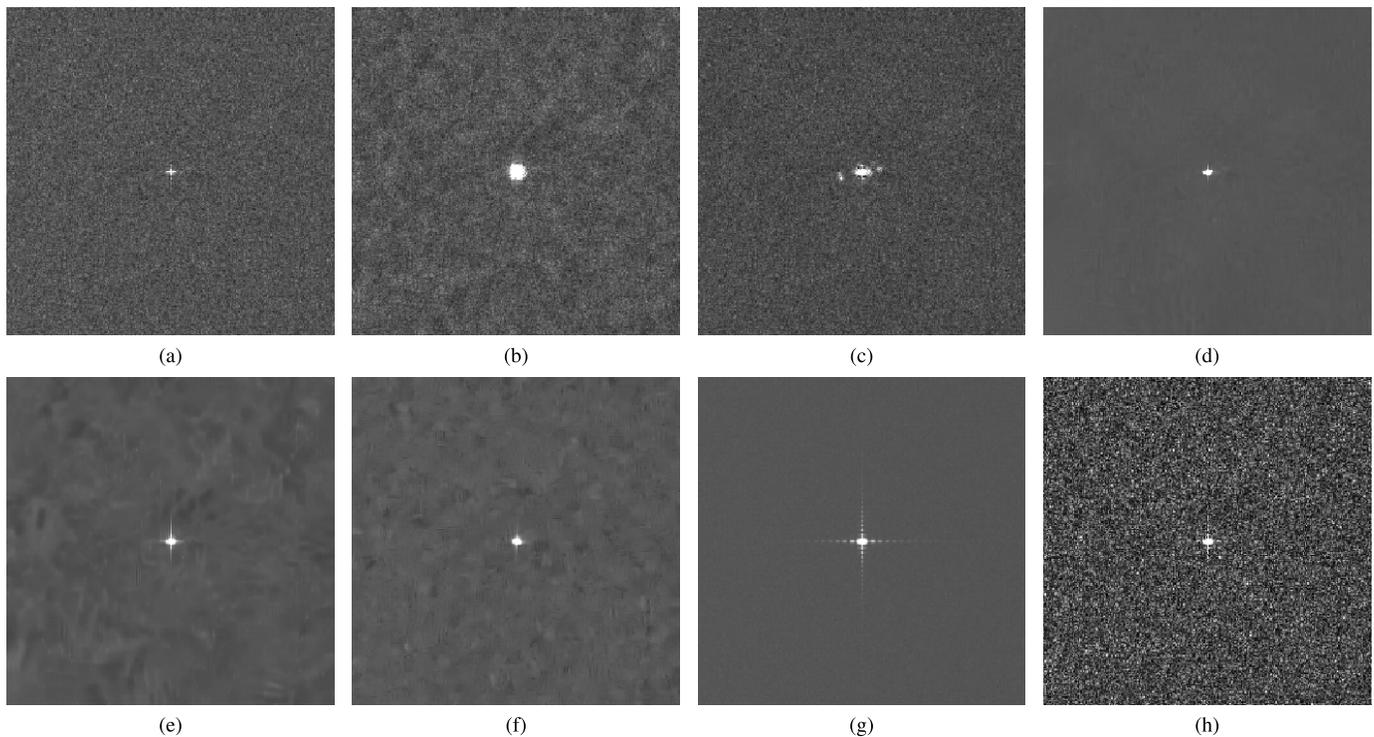

Fig. 15.   Results for the Homogeneous with Corner test case (band 8). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.
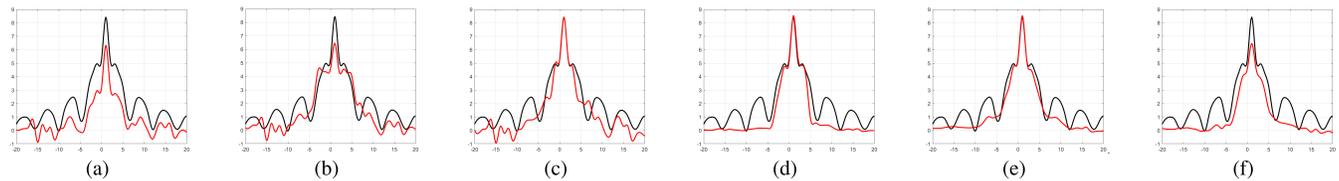

Fig. 16.   Corner range profile in logarithmic scale for the Homogeneous with Corner test case (band 8). Despeckling filter (red line) and reference (black line). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D.

unperturbed bands, PS and ENL_R, are listed in Table IV. It is worth noting that the corner quality indicators C_{NN} and C_{BG} are here measured on the single perturbed band without any averaging through multiple bands as done in the single-image benchmarking test case. This causes slight differences of such parameters for the noisy data set with respect to the values obtained in the stationary Corner test case, see Table I. Conversely, the PS and ENL_R parameters are averaged through the seven unperturbed bands. Additionally, ENL_R is measured over an upper-left 100×100 region. By visual inspection of the unperturbed band in Fig. 14, it is evident that the majority of filters introduce visible artifacts that, in multilook and RABASAR, closely recall the corner reflector feature. The latter case might have detrimental effects in some applications, e.g., change detection algorithms, where it would be highly desirable to keep as unaltered as possible the temporal local-ization of the reflectivity perturbation. Notwithstanding, any significant distortion of the reflectivity map in the unperturbed bands might impair a correct interpretation and processing of the SAR image. Conversely, log-BM4D and UTA do not exhibit visible artifacts in band 1. Quantitative values of PS are consistent with a visual inspection: the largest values are measured with multilook and RABASAR, whereas log-

BM4D and NLTF offer the lowest sensitivity to the corner perturbation, with UTA and MSAR-BM3D in the middle. However, PS is based on MSE and is therefore not well suited to identify corner, such as artifacts.

In such scenario, it is also interesting to evaluate the impact of the temporal change on the filtering capabilities in regions surrounding the disturbance. This is addressed by ENL_R, which measures the ENL degradation with respect to the Homogeneous test case. It is evident that for filters operating only on the temporal coordinate, i.e., multilook, UTA, and NLTF, the influence of the perturbation is limited to the image portion where the perturbation response is non-negligible. Therefore, for these filters, the ENL_R is practically one. Conversely, spatiotemporal filters are influenced also in regions where the perturbation response is negligible, the size of the region depending on the search window size. This brings to ENL_R values slightly inferior to one.

Moving to the perturbed band, it is interesting to note that UTA and log-BM4D, which exhibits low PS values, introduce a severe distortion of the corner reflector response and perform worst in terms of C_{NN} and C_{BG}, see also the corner profiles in Fig. 16. Similarly, the RABASAR filter, which has one of the highest PS, is able to satisfactorily preserve the corner

TABLE V
MEASURES FOR HOMOGENEOUS WITH BUILDING

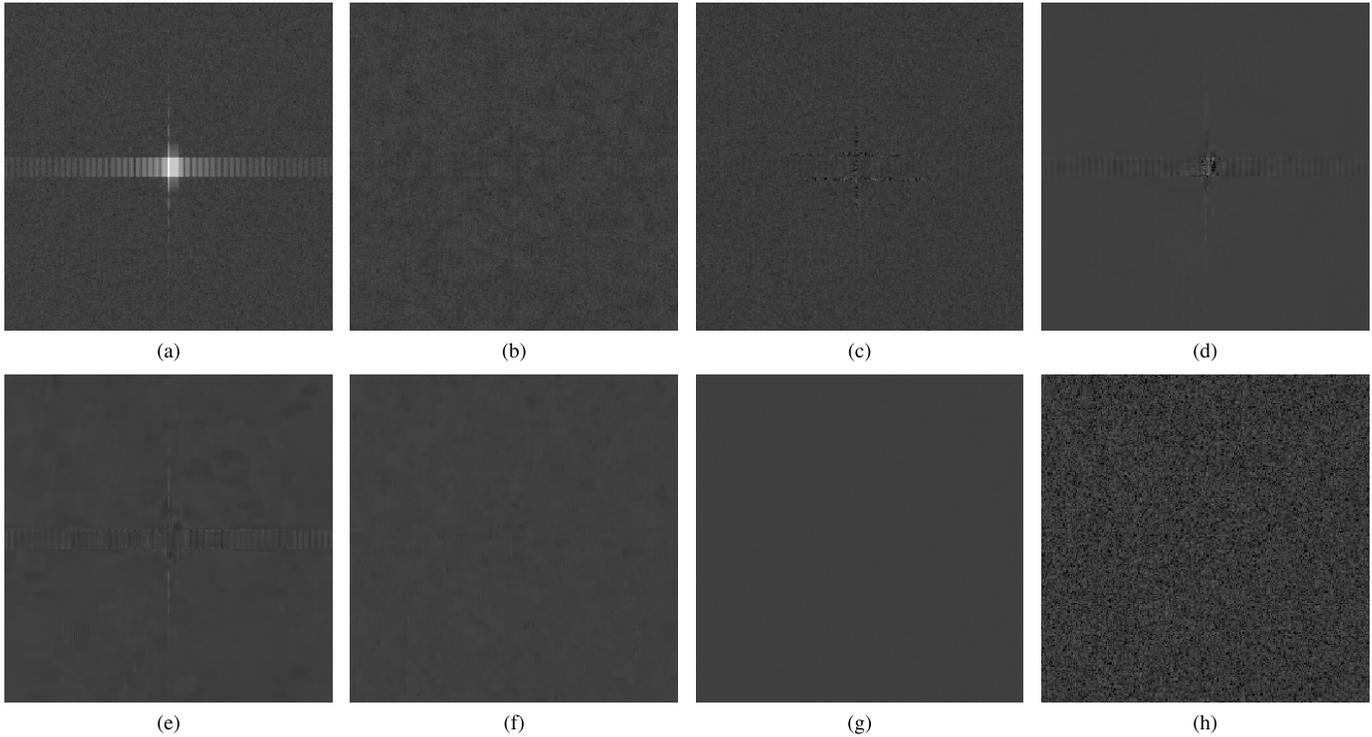|	Noisy	Reference	Multilook	UTA	NLTF	MSAR-BM3D	RABASAR	log-BM4D
C_{DR}	65.96	65.90	56.88	60.47	61.47	66.88	66.19	66.65
BS	0.095	-	28.84	22.17	11.84	4.80	0.75	7.46
PS	0	-	109.40	0.16	5.36	39.71	10.96	2.61
ENL_R	1.00	1.00	0.992	1.000	0.992	1.069	1.006	0.988

Fig. 17. Results for the Homogeneous with Building test case (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.

profile. It is also noteworthy that the multilook filter reduces the corner response by a factor that, in the regions where it is much larger than the average background, can be approximated with the number of bands (8 in our case). This rescaling factor is the same for the corner peak and the surrounding sidelobes. Accordingly, C_{NN} is close to the reference value, whereas C_{BG} is reduced by about 8, which corresponds to approximately 9 dB. MSAR-BM3D and NLTF offer very good preservation of the corner response, despite a visible attenuation of far sidelobes not captured by C_{NN}.

3) Homogeneous With Building: The results relevant to this test case are shown in Fig. 17 (band 1, unperturbed), Fig. 18 (band 8, perturbed), and Fig. 19 (normalized building profiles). Synthetic quality measures are listed in Table V. Similar to the Homogeneous with Corner test case, the building parameters C_{DR} and BS are measured on the perturbed band only, leading to slight differences in the noisy image with respect to the Building test case. The PS is measured on a 21×36 region centered on the building, whereas the ENL_R is measured on the same area used in the Homogeneous with Corner test case. Both parameters are then averaged through the seven unperturbed bands.

Also, in this case, the reflectivity perturbation causes the presence of visible artifacts in the unperturbed bands, see Fig. 17, especially in the output image of the multilook filter. MSAR-BM3D inherits the time averaging strategy adopted in NLTF, which exhibits a nonnegligible PS. The subsequent nonlocal spatial averaging strengthens the artifacts introduced in the time averaging process and leads to a large PS. RABASAR achieves intermediate performance in terms of PS, whereas best results are offered by log-BM4D and UTA. All filters are negligibly affected far from the perturbation as witnessed by ENL_R which is practically unitary. Remarkably, MSAR-BM3D and RABASAR exhibit an ENL_R slightly larger than one, likely due to residual measurement noise.

Nonlocal spatiotemporal filters ensure reliable preservation of the building SAR response, see Fig. 18 and the building measures in Table V. Hence, all such filters preserve the double reflection line quite accurately, with a slight overestimation of C_{DR} within 1 dB. Best performance is here provided by RABASAR which offers the C_{DR} value closest to the reference one and also the lowest BS value. The high accuracy of RABASAR with respect to competitors is also confirmed by
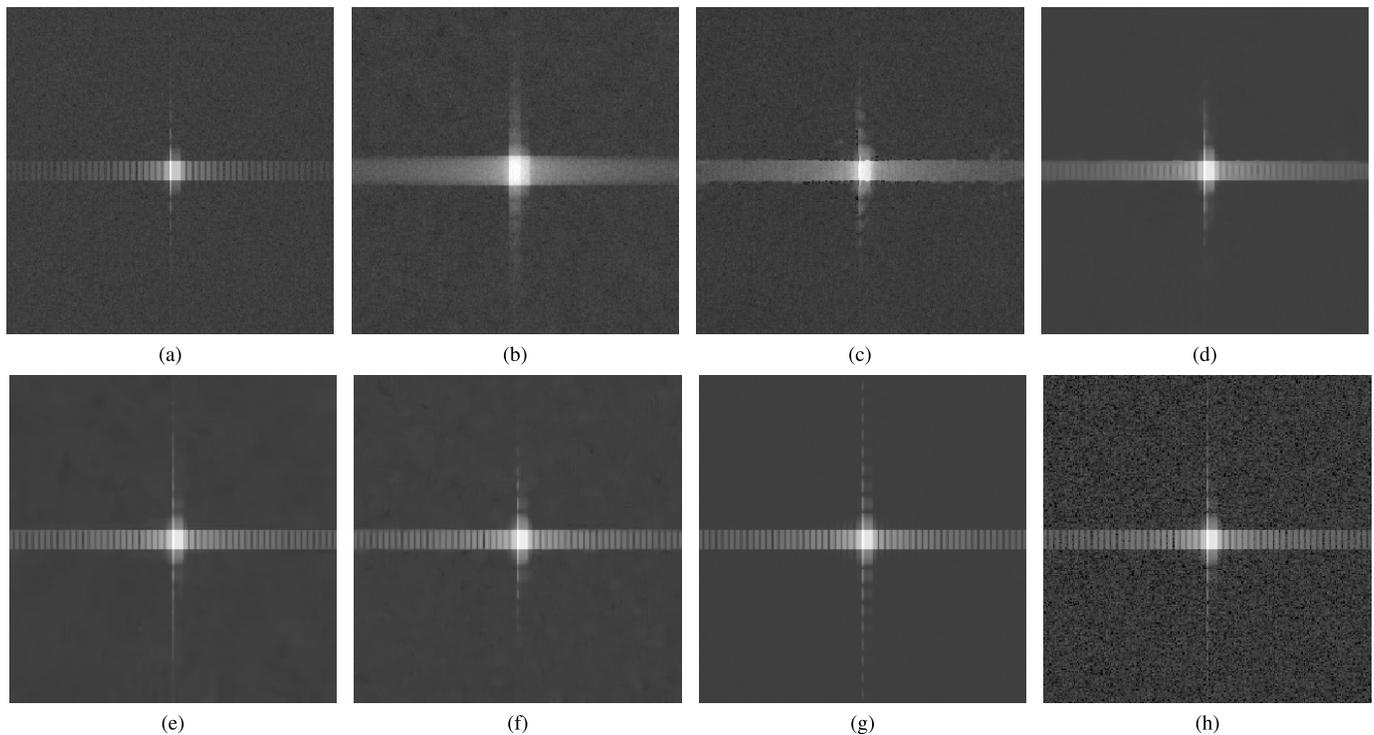
Fig. 18. Results for the Homogeneous with Building test case (band 8). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.
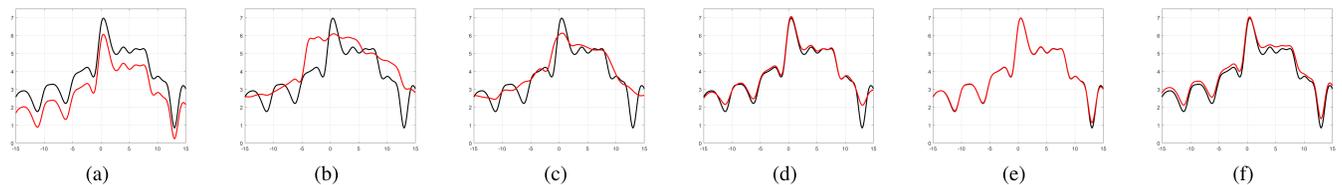

Fig. 19. Normalized (i.e., mean-bias corrected) building range profile in logarithmic scale for the Homogeneous with Building test case (band 8). Despeckling filter (red line) and reference (black line). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D.

TABLE VI
MEASURES FOR SQUARES PERTURBED (BOTTOM EDGE)

	Noisy	Reference	Multilook	UTA	NLTF	MSAR-BM3D	RABASAR	log-BM4D
ES* (perturbed band)	0.013	-	1.384	0.610	0.519	0.404	0.012	0.083
ES* (unperturbed bands)	0.018	-	0.009	0.014	0.012	0.029	0.012	0.035
ES_R	1.000	-	5.460	4.332	5.976	6.356	5.510	1.559

a visual inspection of Figs. 18 and 19, where the capability of the filter in retaining the fine spatial details of the building response can be appreciated. MSAR-BM3D and log-BM4D apply a stronger smoothing, which leads to larger BS values, the former inheriting the poor preservation of fine spatial details provided by NLTF. Worst performance is achieved by UTA and multilook, the former applying an excessive smoothing, the latter losing about 9 dB in C_{DR} as in the Homogeneous with Corner test case.

4) Squares Perturbed: Figs. 20 and 21 show an unperturbed and the perturbed bands, for the noisy, clean and the selected filters, respectively. For the sake of clarity, we also show the normalized range profiles of the bottom vertical edge

in Figs. 22 and 23 in the unperturbed and perturbed bands, respectively. Performance metrics are listed in Table VI.

While being an effective and efficient solution for despeckling of stationary time series, in the presence of significant temporal changes, the multilook filter reveals its limits and justifies the adoption of more sophisticated temporal filtering approaches, such as the nonlocal paradigm. Indeed, a pure temporal averaging of all the band filters out the perturbed square as it is clearly visible in Figs. 20–23. This behavior is well captured by ES* whose largest value is obtained by the multilook filter.

But for the multilook filter, a visual inspection of Fig. 20 does not reveal appreciable differences with respect to the
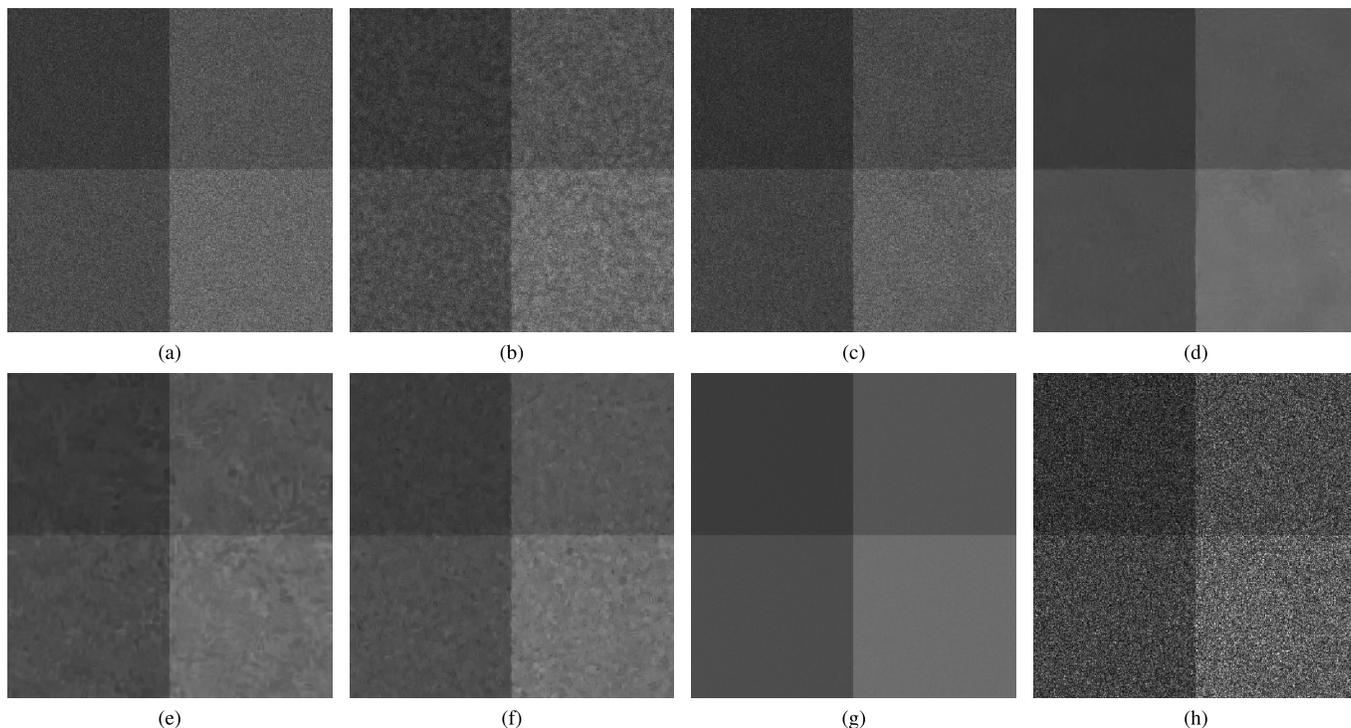
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

DI MARTINO *et al.*: BENCHMARKING FRAMEWORK FOR MULTITEMPORAL SAR DESPECKLING 21

Fig. 20. Results for the Squares perturbed test case (band 1). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.
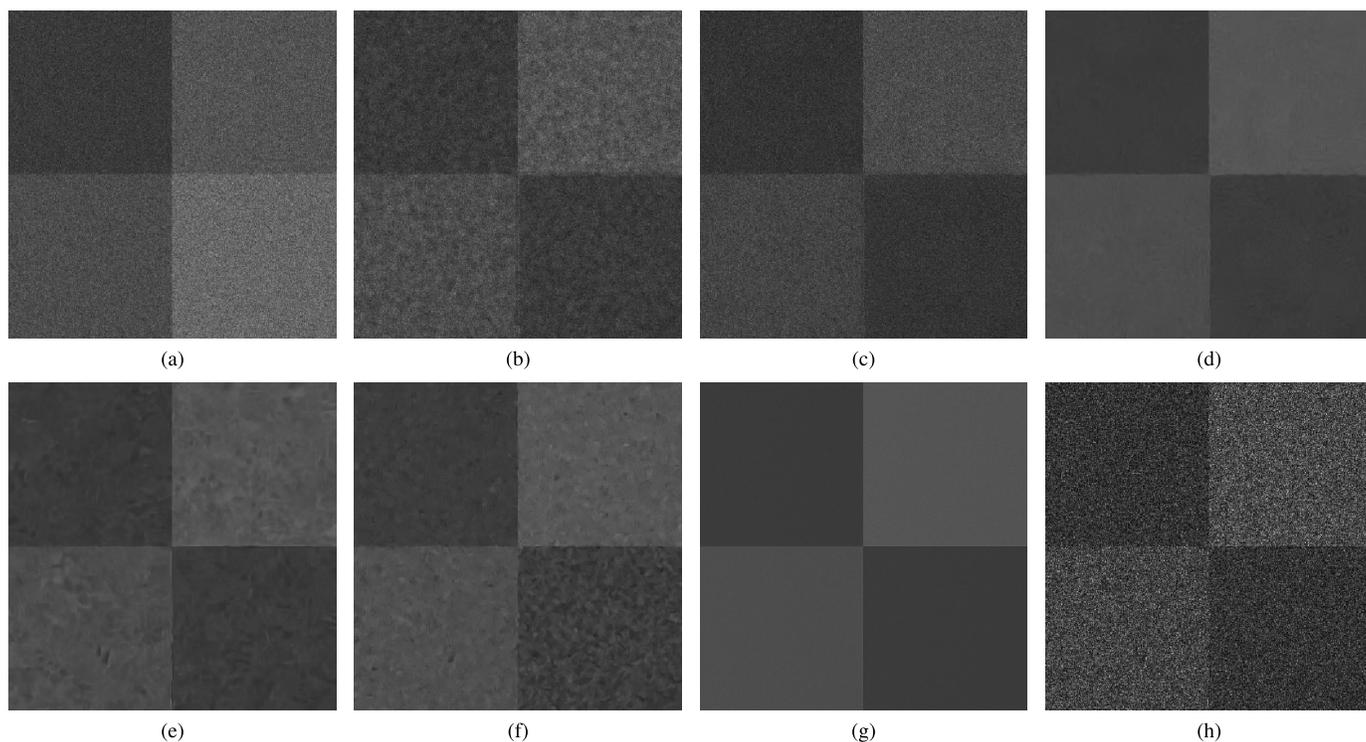

Fig. 21. Results for the Squares perturbed test case (band 8). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D. (g) Reference. (h) Noisy.

Squares test case shown in Fig. 6. However, deeper insights into the despeckling capabilities of the filters can be gathered by looking at the ES* and ES$_R$ parameters and by comparing the edge profiles in Fig. 22 with those in Fig. 7. The weighted temporal filtering makes UTA and NLTF more effective in retaining the perturbed feature compared to multilook. However, a significant distortion of the lower edge is still visible in the perturbed band (as is witnessed by ES* and visible in Fig. 23). An additional filtering in the space domain aids in the preservation of the perturbed edge thanks to the much
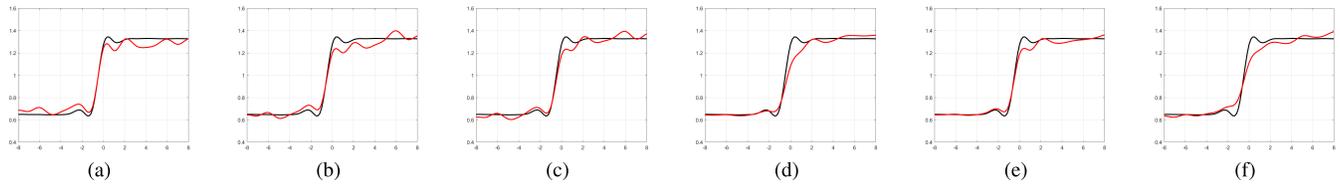
Fig. 22.   Normalized (i.e., mean-bias corrected) edge range profile for the Squares perturbed test case (band 1). Despeckling filter (red line) and reference (black line). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D.
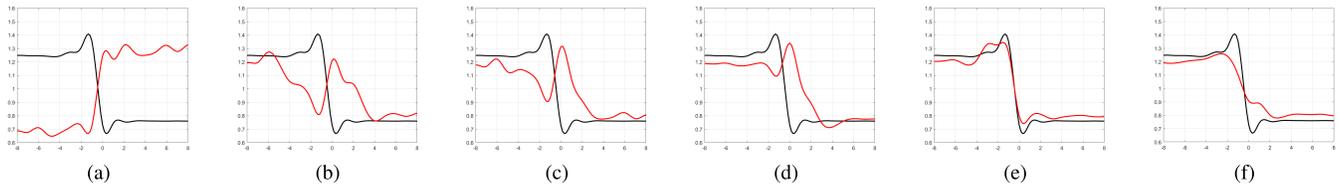

Fig. 23.   Normalized (i.e., mean-bias corrected) edge range profile for the Squares perturbed test case (band 8). Despeckling filter (red line) and reference (black line). (a) Multilook. (b) UTA. (c) NLTF. (d) MSAR-BM3D. (e) RABASAR. (f) Log-BM4D.

larger number of samples used in the despeckling process. This is clearly visible in comparing the perturbed profiles and ES* of NLTF and MSAR-BM3D, with the former being the temporal filtering step of the latter. Among spatiotemporal filters, RABASAR offers the best performance in terms of ES*, whereas log-BM4D and MSAR-BM3D suffer from an excessive smoothing and edge shift, respectively.

Despeckling quality in the unperturbed bands is assessed by evaluating ES* and ES_R on the lower edge. Remarkably, it emerges that, apart the log-BM4D filter, all the selected filters exhibit a large PS in terms of ES, i.e., a large ES_R value. The ES lose is also evident by comparing Figs. 22(d) and 7(d). However, it is worth noting that a low (high) ES_R does not imply a good (bad) preservation of the edge, i.e., a low (high) ES*, but only that the filter is scarcely (largely) influenced by temporal changes in reflectivity. Indeed, despite the low ES_R, log-BM4D suffers from an excessive smoothing of the edge, as revealed by ES*. On the other hand, despite a large ES_R, the multilook filter still offers the best contour preservation in the unperturbed bands.

C. Flexibility of the Benchmarking Tool

The results shown in Section V-A and V-B have been obtained assuming time series of eight bands (with the exception of the Convergence test case), where the perturbation, if present, is located in the last image. However, these assumptions are not mandatory for the correct usage of the proposed methodology. To better investigate this point, here we discuss how the proposed benchmarking framework can be adapted (as far as possible) to specific applications or SAR user needs. To this end, we considered three aspects potentially affecting the despeckling capabilities of a filter: temporal location of the perturbation within the time series; temporal duration of the perturbation; observation time period. For each of these points, we explain the flexibility offered by our tool and, additionally, we briefly discuss the results obtained by applying the selected despeckling filters to properly modified test cases.

1) Temporal Location of the Perturbation: Potential users of the proposed benchmarking tool might be interested in assessing filters' performance on time series affected by tem-

poral changes of the scene reflectivity in specific dates. If this is the case, the researcher will be free to modify the perturbed band index according to his/her needs by moving the perturbed band within the time series.

As an example, the impact of the temporal location of the perturbation on the performance of the selected filters has been analyzed on the Homogeneous with Corner test case. In particular, we run this test case multiple times by changing each time the position of the perturbed band within the time series. The result of such an analysis is that the considered filters are not appreciably influenced by the time instant where the perturbation takes place, in the sense that the quality of the despeckled multitemporal image is scarcely influenced by the temporal location of the perturbation. Such a behavior can be justified by the lack of time ordering in the despeckling processing chain of the selected filters. However, it cannot be ruled out that other filters could exhibit a larger sensitivity to the actual temporal location of the perturbation.

2) Temporal Duration of the Perturbation: Changes in the scene during the observation time period due for example to new built-up areas or flooding might persist over time leading to temporal changes in multiple subsequent bands, with potentially severe effects on the quality of the despeckled time series. The possibility of setting up new test cases by selecting a number of perturbed bands suited to the scenario of interest represents another strength of the proposed tool.

To better assess the effects of the temporal duration of the perturbation on the despeckling capabilities of the filters, we slightly modified the Homogeneous with Corner test case presented in Section V-B by considering a data set of eight images, the last four out of which were perturbed by the corner target. In presence of a lasting perturbation, most filters exhibit a larger PS due to the higher number of perturbed bands. Conversely, RABASAR experiences a significant reduction of PS, due to a more accurate estimation of the superimage for the unperturbed bands. Additionally, the presence of more perturbed patches results in a lowering of the ENL_R for MSAR-BM3D and log-BM4D. Finally, multilook, UTA, and log-BM4D offer greater accuracy in preserving the corner response with respect to the single perturbed band case.

3) Observation Time Period: In specific applications, e.g., change detection, it might be required the denoising of a data set as limited as two or three images. The methodology described in Section IV can be applied to an arbitrary number of bands $M \geq 2$, the only exception being the Convergence test case which requires a large number of bands $M = 64$. The user can then freely define the length of the time series according to the application of interest. Obviously, in the case of temporal changes of the reflectivity, the bands where the change takes place must be less than M.

The sensitivity of the selected filters to the number of bands has been assessed on the time-varying test cases for $M = 2$ and $M = 4$. In both cases, the Homogeneous varying test case has been built according to (13) and (14), while in the remaining time-varying test cases the perturbation has been put on the last band. As for the unperturbed bands, it emerges that time-domain filters exhibit negligible effects on ENL_R due to the filtering along the time coordinate only. Additionally, they offer a significant improvement of the perturbation measures which are closer to corresponding reference values meaning that a large number of unperturbed bands make the filter underestimate the perturbation response. Conversely, for spatiotemporal filters, the reduced number of bands leads to a severe reduction of the speckle rejection capabilities as it is measured by ENL_R which lowers below 0.9 for $M = 2$. This is due to the much lower amount of similar patches available for averaging. Finally, for all filters but MSAR-BM3D an increased sensitivity to the perturbation is obtained reducing the number of available bands.

D. Effects of Speckle Temporal Correlation

In this section, we analyze the impact of temporal correlation on the despeckling capabilities of multitemporal filters. Indeed, the assumption of uncorrelated bands made in the proposed benchmarking framework is rather extreme as the coherence time of both natural surfaces and urban areas is typically larger than the revisit time of most spaceborne SAR missions. As a matter of fact, some SAR applications, e.g., SAR interferometry (InSAR), differential InSAR, strongly rely on the temporal coherence of the illuminated scene and their performance degrades with decreasing correlation among the bands. Notwithstanding, for despeckling purposes, best performance is reasonably expected with uncorrelated time series, since the temporal filtering (however, it is intended) will be more effective with weaker correlations.

In order to investigate the effects of the temporal correlation on the performance of multitemporal filters, the adopted SARAS simulator should be properly modified in order to account for time-correlated speckle noise, which in our physical-based simulator undergoes the same processing chain as the clean signal does. However, the design of a SAR simulator of correlated time series goes outside the scope of this work and will be the subject of future works. Accordingly, we here rely on the Homogeneous case, as it allows for a much easier simulation of time-correlated bands, thanks to the homogeneous reflectivity. Nevertheless, as it will be better motivated in the following of this section, the main results of
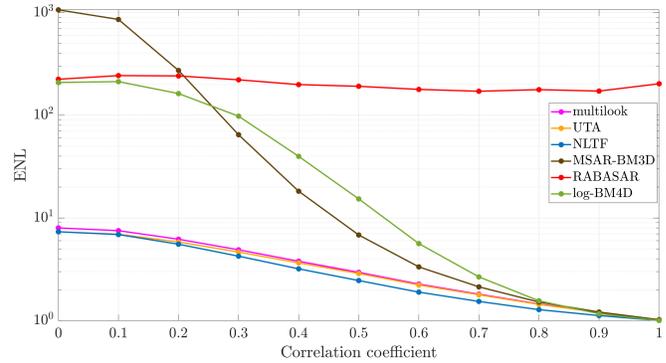

Fig. 24. ENL for $M = 8$ as a function of the temporal correlation coefficient for different multitemporal SAR despeckling algorithms.

this analysis are reasonably independent of the spatial features of the scene reflectivity and, therefore, similar results are expected in other test cases. We considered $M = 8$ bands with size 256×256 and simulated multiple time series, each one exhibiting a constant temporal correlation profile. An example of single-look and reference image is shown in Fig. 1(a) and (f), respectively. According to the homogeneous reflectivity, despeckling performance is evaluated through the ENL (averaged over all eight bands) and results are shown in Fig. 24 for the selected multitemporal filters.

It is noticeable that two different behaviors can be highlighted: most algorithms analyzed here exhibit significant sensitivity to the temporal correlation and offer poorer performance with increasing correlation. In the limiting case of perfectly correlated time series ($\rho = 1$), they do not provide any speckle reduction (ENL equals one). This is easily explained for time-domain filters: if $\rho = 1$, image sample values at any location are the same among all the bands and, therefore, no filtering takes place along the time coordinate. For spatiotemporal filters, results reveal a strong impact of the temporal characteristics of the scene on the spatial filtering step. However, spatiotemporal filters offer stronger speckle rejection compared to time-domain filters for $\rho < 1$. Such a negative impact of the temporal correlation is likely due to the less information which is available to the filters with respect to the case of uncorrelated bands. Accordingly, it is reasonable to expect such negative effects are present irrespective of the spatial features of the image. The only exception is the RABASAR filter which is negligibly affected by temporal correlation among bands, thus offering a nearly constant ENL regardless of the correlation coefficient. This peculiar behavior demonstrates the benefits of the spatial filtering step for the generation of the superimage. Similarly, some other filters reported in the literature are able to take advantage of temporal correlation among the bands, see e.g., [47].

VI. Conclusion

In this article, we presented a benchmarking framework for the evaluation of the performance of multitemporal SAR despeckling algorithms. The proposed approach is aimed at supporting the despeckling and image processing community by providing an objective and standard procedure for a

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

24   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

quantitative assessment of the despeckling capabilities of multitemporal SAR filters. To this end, we conceived a set of simulated test images relevant to scenes with both stationary and time-varying reflectivity. The aim of the stationary-reflectivity test cases is twofold: first, to evaluate basic features of multitemporal filters, such as speckle suppression in homogeneous areas and the preservation of fine spatial details (edges, textured areas, peculiar scattering mechanisms); second, to analyze the convergence properties (steady-state error, convergence rate) of the filters, i.e., its behavior for a number of bands going to infinity. The sensitivity of the filter to reflectivity temporal perturbations is assessed in the time-varying reflectivity test cases, where different temporal changes are simulated. Such temporal variations of the reflectivity have been inspired by physical considerations about the scene dynamics in time, such as agricultural activities, flooding, vegetation growth, and appearance of strong scatterers. For each scene, proper quality measures have been introduced in order to assess despeckling capabilities in several respects, namely, speckle reduction, texture, and fine details preservation (both in space and time), edge and scattering properties preservation. All such metrics along with their target value are listed in Table VII.

The proposed tool offers some degrees of flexibility as it can be tuned by the SAR user to specific applications or scenarios. For instance, despeckling filters could be tested on time series of few images for change detection applications or on long time series for long-term monitoring applications; analysis on persistent variations of reflectivity, e.g., new buildings and flooded areas, can be carried out by considering multiple perturbed bands.

An experimental analysis has been carried out in order to gain insight into the actual potential of this tool and test its efficacy to catch the main despeckling weaknesses and strengths of a filter. To this end, the test cases were run over a representative set of state-of-the-art multitemporal filters, including algorithms working on the time coordinate only or somehow supporting temporal averaging with spatial filtering. The numerical results demonstrate a good consistency between synthetic quality indicators and visual inspection of filtered images.

The applicability of the proposed methodology to despeckling filters based on the deep learning paradigm deserves some comments. Indeed, the performance of such filters is strictly related to the training data set and training method, which are rarely made available or even properly discussed in the related works, thus impairing reproducible research. Unfortunately, so far, there are no standardized procedures for the training of deep neural networks, which might be desirable for fair and meaningful comparisons, as suggested in [46]. However, this aspect, despite its interest for the despeckling community, goes outside the scope of this work, which is addressed to ready-to-use filters. Notwithstanding, pre-trained deep learning filters, whatever the training data set and method, can still take advantage of this benchmarking tool for the assessment of their performance.

Additionally, the proposed benchmarking framework has been built over the fundamental hypothesis of temporally

TABLE VII
PERFORMANCE METRICS FOR THE PROPOSED MULTITEMPORAL BENCHMARKING FRAMEWORK

Test case	Parameter	Target value
HOMOGENEOUS	MoI	1
	MoR	1
	VoR	1
	DG	∞
	ENL	∞
DEM	MoI	1
	MoR	1
	VoR	1
	C_x	2.40
	DG	∞
SQUARES	ES* (up)	0
	ES* (down)	0
	FOM	1
CORNER	C_{NN}	7.75
	C_{BG}	36.56
BUILDING	C_{DR}	65.90
	BS	0
CONVERGENCE	MSE_{64}	0
	$M_{\alpha_{TH}}$	-
HOMOGENEOUS VARYING	$\mathrm{MoI}*_\mu$	1
	$\mathrm{MoI}*_\sigma$	0
	DG	∞
	ENL	∞
	ENL_R	1
HOMOGENEOUS WITH CORNER	C_{NN}	7.75
	C_{BG}	36.56
	PS	0
	ENL_R	1
HOMOGENEOUS WITH BUILDING	C_{DR}	65.90
	BS	0
	PS	0
	ENL_R	1
SQUARES PERTURBED (lower edge)	ES* (perturbed band)	0
	ES* (unperturbed bands)	0
	ES_R	1

uncorrelated bands. This assumption allowed us to analyze filter performance under the most favorable despeckling conditions. However, a preliminary analysis of the impact of temporal correlation among bands on the despeckling capabilities of the selected filters has been carried out. It has pointed out that most filters are negatively affected by temporal correlation and that negligible speckle reduction is offered for highly correlated time series. The only exception among the selected filters is represented by RABASAR, whose despeckling capabilities seem to be practically independent of temporal correlation.

Notwithstanding, the development of a benchmarking framework explicitly taking into account realistic temporal correlation profiles in the definition of the canonical scenes represents an interesting research path to be explored in the near future.

In summary, the proposed benchmarking framework might be fruitfully exploited to:

1) Perform a fair and objective comparison of different multitemporal filters.
2) Get meaningful and wide insights into the despeckling capabilities of the filter.

3) Build a proper application-dependent benchmark tool by somehow combining the different measures introduced.
4) Pave the way for creating a benchmarking framework using real-world SAR imagery.

REFERENCES

[1] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*. Boca Raton, FL, USA: CRC Press, 2017.

[2] F. Xue, X. Lv, F. Dou, and Y. Yun, "A review of time-series interferometric SAR techniques: A tutorial for surface deformation analysis," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 1, pp. 22–42, Mar. 2020.

[3] N. Pierdicca, L. Pulvirenti, and G. Pace, "A prototype software package to retrieve soil moisture from Sentinel-1 data by using a Bayesian multitemporal algorithm," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 153–166, Jan. 2014.

[4] S. Paloscia, G. Macelloni, P. Pampaloni, and E. Santi, "The contribution of multitemporal SAR data in assessing hydrological parameters," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 3, pp. 201–205, Jul. 2004.

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

[6] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.

[7] C. Schuster, T. Schmidt, C. Conrad, B. Kleinschmit, and M. Förster, "Grassland habitat mapping by intra-annual time series analysis–Comparison of RapidEye and TerraSAR-X satellite data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 34, pp. 25–34, Feb. 2015.

[8] K. Zalite, O. Antropov, J. Praks, K. Voormansik, and M. Noorma, "Monitoring of agricultural grasslands with time series of X-band repeat-pass interferometric SAR," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 8, pp. 3687–3697, Aug. 2016.

[9] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Small reservoirs extraction in semiarid regions using multitemporal synthetic aperture radar images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3482–3492, Aug. 2017.

[10] L. Zeng, M. Schmitt, L. Li, and X. X. Zhu, "Analysing changes of the poyang lake water area using Sentinel-1 synthetic aperture radar imagery," *Int. J. Remote Sens.*, vol. 38, no. 23, pp. 7041–7069, Dec. 2017.

[11] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "A new framework for SAR multitemporal data RGB representation: Rationale and products," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 117–133, Jan. 2015.

[12] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "An end-user-oriented framework for the classification of multitemporal SAR images," *Int. J. Remote Sens.*, vol. 37, no. 1, pp. 248–261, Jan. 2016.

[13] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1321–1334, Jun. 2004.

[14] H. Skriver *et al.*, "Crop classification using short-revisit multitemporal SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 2, pp. 423–431, Jun. 2011.

[15] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Rijeka, Croatia: SciTech, 2004.

[16] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, "A tutorial on speckle reduction in synthetic aperture radar images," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 3, pp. 6–35, Sep. 2013.

[17] D. Kuan, A. Sawchuk, T. Strand, and P. Chavel, "Adaptive restoration of images with speckle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 373–383, Mar. 1987.

[18] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 157–166, Mar. 1982.

[19] M. I. H. Bhuiyan, M. O. Ahmad, and M. N. S. Swamy, "Spatially adaptive wavelet-based method using the cauchy prior for denoising the SAR images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 4, pp. 500–507, Apr. 2007.

[20] S. Solbo and T. Eltoft, "Homomorphic wavelet-based statistical despeckling of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 711–721, Apr. 2004.

[21] T. Bianchi, F. Argenti, and L. Alparone, "Segmentation-based MAP despeckling of SAR images in the undecimated wavelet domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 9, pp. 2728–2742, Sep. 2008.

[22] M. Dai, C. Peng, A. K. Chan, and D. Loguinov, "Bayesian wavelet shrinkage with edge detection for SAR image despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1642–1648, Aug. 2004.

[23] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 606–616, Feb. 2012.

[24] F. Argenti and L. Alparone, "Speckle removal from SAR images in the undecimated wavelet domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2363–2374, Nov. 2002.

[25] C.-A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2661–2672, Dec. 2009.

[26] G. Di Martino, A. Di Simone, A. Iodice, and D. Riccio, "Scattering-based nonlocal means SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3574–3588, Jun. 2016.

[27] G. Di Martino, A. Di Simone, A. Iodice, G. Poggi, D. Riccio, and L. Verdoliva, "Scattering-based SARBM3D," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2131–2144, Jun. 2016.

[28] G. F. De Grandi, M. Leysen, J. S. Lee, and D. Schuler, "Radar reflectivity estimation using multiple SAR scenes of the same target: Technique and applications," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. Remote Sens. Sci. Vis. Sustain. Develop. (IGARSS)*, vol. 2, Aug. 1997, pp. 1047–1050.

[29] S. Quegan and J. Jiong Yu, "Filtering of multichannel SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2373–2379, Nov. 2001.

[30] X. Su, C.-A. Deledalle, F. Tupin, and H. Sun, "Two-step multitemporal nonlocal means for synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6181–6196, Oct. 2014.

[31] G. Chierchia, M. El Gheche, G. Scarpa, and L. Verdoliva, "Multitemporal SAR image despeckling based on block-matching and collaborative filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5467–5480, Oct. 2017.

[32] S. Quegan, T. Le Toan, J. J. Yu, F. Ribbes, and N. Floury, "Multitemporal ERS SAR analysis applied to forest mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 741–753, Mar. 2000.

[33] W. Zhao, C.-A. Deledalle, L. Denis, H. Maitre, J.-M. Nicolas, and F. Tupin, "Ratio-based multitemporal SAR images denoising: RABASAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3552–3565, Jun. 2019.

[34] G. Di Martino, M. Poderico, G. Poggi, D. Riccio, and L. Verdoliva, "Benchmarking framework for SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1596–1615, Mar. 2014.

[35] G. Franceschetti, M. Migliaccio, D. Riccio, and G. Schirinzi, "SARAS: A synthetic aperture radar (SAR) raw signal simulator," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 1, pp. 110–123, Jan. 1992.

[36] G. Franceschetti, A. Iodice, D. Riccio, and G. Ruello, "SAR raw signal simulation for urban structures," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1986–1995, Sep. 2003.

[37] G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "A novel approach for disaster monitoring: Fractal models and tools," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1559–1570, Jun. 2007.

[38] G. Franceschetti, M. Migliaccio, and D. Riccio, "SAR raw signal simulation of actual ground sites described in terms of sparse input data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 6, pp. 1160–1169, Nov. 1994.

[39] G. Franceschetti, M. Migliaccio, and D. Riccio, "On ocean SAR raw signal simulation," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 1, pp. 84–100, Jan. 1998.

[40] S. G. Dellepiane and E. Angiati, "Quality assessment of despeckled SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 691–707, Feb. 2014.

[41] L. Gomez, R. Ospina, and A. Frery, "Unassisted quantitative evaluation of despeckling filters," *Remote Sens.*, vol. 9, no. 4, p. 389, Apr. 2017, doi: 10.3390/rs9040389.

[42] R. Touzi, "A review of speckle filtering in the context of estimation theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2392–2404, Nov. 2002.

[43] M. Walessa and M. Datcu, "Model-based despeckling and information extraction from SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2258–2269, Sep. 2000.

[44] F. Aspert, M. Bach-Cuadra, A. Cantone, F. Holecz, and J.-P. Thiran, "Time-varying segmentation for mapping of land cover changes," in *Proc. ENVISAT Symp.*, 2007.

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

[46] G. Fracastoro, E. Magli, G. Poggi, G. Scarpa, D. Valsesia, and L. Verdoliva, "Deep learning methods for SAR image despeckling: Trends and perspectives," 2020, *arXiv:2012.05508*. [Online]. Available: https://arxiv.org/abs/2012.05508

[47] X. Ma, C. Wang, Z. Yin, and P. Wu, "SAR image despeckling by noisy reference-based deep learning method," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8807–8818, Dec. 2020.

[48] Y. Inoue *et al.*, "Season-long daily measurements of multifrequency (Ka, Ku, X, C, and L) and full-polarization backscatter signatures over paddy rice field and their relationship with biological variables," *Remote Sens. Environ.*, vol. 81, nos. 2–3, pp. 194–204, Aug. 2002.

[49] G. Di Martino, A. Di Simone, A. Iodice, D. Riccio, and G. Ruello, "Assessing performance of multitemporal SAR image despeckling filters via a benchmarking tool," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 1536–1539.

Gerardo Di Martino (Senior Member, IEEE) was born in Naples, Italy, in 1979. He received the Laurea degree *(cum laude)* in telecommunication engineering and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples Federico II, Naples, in 2005 and 2009, respectively.

From 2009 to 2016, he was with the University of Naples Federico II, working on projects regarding applied electromagnetics and remote sensing topics. From 2014 to 2016, he was with the Italian National Consortium for Telecommunications (CNIT), Parma, Italy, and the Regional Center Information Communication Technology (CeRICT), Naples. He is a tenure-track Assistant Professor of Electromagnetics with the Department of Electrical Engineering and Information Technology, University of Naples Federico II. His research interests include microwave remote sensing and electromagnetics, with focus on electromagnetic scattering from natural surfaces and urban areas, SAR signal processing and simulation, information retrieval from SAR data, and electromagnetic propagation in urban areas.

Dr. Di Martino is an Associate Editor of IEEE JOURNAL OF SELECTED TOPICS ON APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and IEEE ACCESS, and a Section Editorial Board Member of *Remote Sensing* (MDPI).

Alessio Di Simone (Member, IEEE) was born in Torre del Greco, Italy, in 1989. He received the B.Sc. and M.Sc. (Laurea) degrees *(cum laude)* in telecommunication engineering and the Ph.D. degree in information technology and electrical engineering from the University of Naples Federico II, Naples, Italy, in 2011, 2013, and 2017, respectively.

In 2016, he joined the Universitat Politècnica de Catalunya, Barcelona, Spain, as a Visiting Researcher. From 2017 to 2018, he joined the NATO Science and Technology Organization Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy, as a Visiting Researcher. Since 2017, he has been a Research Fellow with the Department of Electrical Engineering and Information Technology, University of Naples Federico II. His research interests include microwave remote sensing and electromagnetics, including modeling of the electromagnetic scattering from natural surfaces, urban areas, and artificial targets, and simulation and processing of synthetic aperture radar (SAR) and GNSS-R data.

Dr. Di Simone was awarded a prize for the Best Master Thesis in Remote Sensing by the IEEE South Italy Geoscience and Remote Sensing Chapter in 2015.

Antonio Iodice (Senior Member, IEEE) was born in Naples, Italy, in 1968. He received the Laurea degree *(cum laude)* in electronic engineering and the Ph.D. degree in electronic engineering and computer science from the University of Naples Federico II, Naples, in 1993 and 1999, respectively.

In 1995, he was with the Research Institute for Electromagnetism and Electronic Components of the Italian National Council of Research (IRECE-CNR), Naples. From 1999 to 2000, he was with Telespazio S.p.A., Rome, Italy. He was with the University of Naples Federico II as a Research Scientist from 2000 to 2004 and then, from 2005 to 2018, as a Professor of Electromagnetics. He is a Full Professor of Electromagnetics with the Department of Electrical Engineering and Information Technology, University of Naples Federico II. He has been involved as a Principal Investigator or a Co-Investigator in several projects funded by the European Union (EU), the Italian Space Agency (ASI), the Italian Ministry of Education and Research (MIUR), the Campania Regional Government, and private companies. He is the author or coauthor of more than 300 articles, of which over 90 published on refereed journals. His research interests include microwave remote sensing and electromagnetics: modeling of electromagnetic scattering from natural surfaces and urban areas, simulation and processing of synthetic aperture radar (SAR) signals, and electromagnetic propagation in urban areas.

Prof. Iodice received the "2009 Sergei A. Schelkunoff Transactions Prize Paper Award" from the IEEE Antennas and Propagation Society, for the best paper published in 2008 on the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION. He was recognized by the IEEE Geoscience and Remote Sensing Society as a 2015 Best Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is the Chair of the IEEE South Italy Geoscience and Remote Sensing Chapter.

Daniele Riccio (Fellow, IEEE) was born in Naples, Italy. He received the Laurea degree *(cum laude)* in electronic engineering from the University of Naples Federico II, Naples, in 1989.

His career has been developing at the University of Naples Federico II, where he is a Full Professor of electromagnetic fields with the Department of Electrical Engineering and Information Technology. He was a Research Scientist with the Italian National Research Council, Institute for Research on Electromagnetics and Electronic Components, Naples, from 1989 to 1994, a Guest Scientist with the German Aerospace Centre (DLR), Munich, Germany, from 1994 to 1995, and a Research Affiliate to the National Aeronautics and Space Administration (NASA), Washington, DC, USA, from 2010 to 2018. He lectured abroad (Spain and Czech Republic) in Ph.D. Schools. At the University of Napoli Federico II, he is the Rector Delegate for Ph.D. Schools. He is the coordinator of the Ph.D. Schools in Information and Communication Technology for Health and in Information Technology and Electrical Engineering. He has authored three books, including *Scattering, Natural Surfaces, and Fractals* (2007), and over 450 scientific articles. His research interests include remote sensing, electromagnetic scattering from complex media and surfaces, synthetic aperture radar techniques, application of fractal geometry to remote sensing, and propagation of electromagnetic fields for wireless communication networks planning.

Prof. Riccio is a member of the Cassini Radar Science Team. He is a member of the Mission Advisory Group for the Platino-1 mission, Italian Space Agency. He is a Board Member for the Ph.D. programme on Space at the Scuola Superiore Meridionale, Naples. He is the Athenaeum Representative with the Assembly of the National Inter-University Consortium for Telecommunications (CNIT), Parma, Italy, where has been elected as a Board Member, and the Board of the Italian Society of Electromagnetism (SIEm), where he is appointed with the Secretary role. He was a recipient of the 2009 Sergei A. Schelkunoff Transactions Prize Paper Award for the best paper published in 2008 on the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION. He was the General Chair of the 5G International Ph.D. School from 2018 to 2020. He serves as an associate editor for some journals on remote sensing.