HAL
open science

NL-SAR: a unified Non-Local framework for resolution-preserving (Pol)(In)SAR denoising

Charles-Alban Deledalle, Loïc Denis, Florence Tupin, Andreas Reigber, Marc Jäger

To cite this version:

Charles-Alban Deledalle, Loïc Denis, Florence Tupin, Andreas Reigber, Marc Jäger. NL-SAR: a unified Non-Local framework for resolution-preserving (Pol)(In)SAR denoising. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53 (4), pp.2021 - 2038. 10.1109/TGRS.2014.2352555 . hal-00844118v3

HAL Id: hal-00844118
https://hal.science/hal-00844118v3

Submitted on 17 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NL-SAR: a unified Non-Local framework for resolution-preserving (Pol)(In)SAR denoising

Charles-Alban Deledalle, Loïc Denis, Florence Tupin, Andreas Reigber, and Marc Jäger

***Abstract*—Speckle noise is an inherent problem in coherent imaging systems like synthetic aperture radar. It creates strong intensity fluctuations and hampers the analysis of images and the estimation of local radiometric, polarimetric or interferometric properties. SAR processing chains thus often include a multi-looking (i.e., averaging) filter for speckle reduction, at the expense of a strong resolution loss. Preservation of point-like and fine structures and textures requires to adapt locally the estimation. Non-local means successfully adapt smoothing by deriving data-driven weights from the similarity between small image patches. The generalization of non-local approaches offers a flexible framework for resolution-preserving speckle reduction.**

We describe a general method, NL-SAR, that builds extended non-local neighborhoods for denoising amplitude, polarimetric and/or interferometric SAR images. These neighborhoods are defined on the basis of pixel similarity as evaluated by multi-channel comparison of patches. Several non-local estimations are performed and the best one is locally selected to form a single restored image with good preservation of radar structures and discontinuities.

The proposed method is fully automatic and handles single and multi-look images, with or without interferometric or polarimetric channels. Efficient speckle reduction with very good resolution preservation is demonstrated both on numerical experiments using simulated data, airborne and spaceborne radar images. The source code of a parallel implementation of NL-SAR is released with the paper.

***Index Terms*—Estimation, Non-local means, Synthetic Aperture Radar (SAR), Interferometry, Polarimetry**

I. Introduction

POLARIMETRIC and interferometric SAR images are increasingly used in remote sensing, for a broad variety of applications ranging from crisis management to biomass study. Several new high-resolution airborne and satellite sensors with full polarimetric and/or interferometric capabilities are now operating (e.g., F-SAR, TerraSAR-X and TanDEM-X, COSMO-SkyMed. . .).

Like other coherent imaging techniques, radar images are affected by speckle noise. Speckle in images results in strong signal-dependent variance. Local smoothing is thus often performed to mitigate these fluctuations in homogeneous regions. Furthermore, the computation of the interferometric and polarimetric signatures of a radar scene requires estimating local covariance matrices from several pixels. Prior to their analysis, SAR images then often undergo processing steps that degrade their resolution. Though a speckle reduction step and covariance estimation are unavoidable in many applications, special care must be taken to limit blurring of significant structures in SAR images.

The simplest approach to speckle reduction and covariance estimation, *spatial multi-looking*, computes a simple moving average with a (typically rectangular) window. Sufficient smoothing of homogeneous regions comes at the cost of a strong resolution loss.

Several improvements to multi-looking have been proposed in the literature. The common underlying idea is to adapt the selection of samples used for covariance estimation in order to avoid mixing pixels belonging to different structures (e.g., blurring edges and strong scatterers by averaging them with their surrounding background). We suggest a classification of these methods into 3 main categories.

A first category of approaches attempts to limit the loss of structural information due to multi-looking, with a post-processing that adaptively reintroduces part of the input image based on the validity of the local stationarity assumption [1]–[3]. Non-stationarity models have been introduced to take into account some prior knowledge concerning the distribution of pixel intensity values, see for instance [4], [5]. Touzi [6] proposed a structural-multiresolution framework to handle both stationary and non-stationary signals and improved the previous approaches by context adaptation.

A second category of approaches selects the image that achieves the best trade-off between data fidelity and regularity, as defined in terms of wavelets coefficients distribution [7]–[12] or total variation of the image [13]–[16]. Such methods are referred to as variational approaches, or, in a Bayesian perspective, as maximum *a posteriori* estimation. The smoothness of the solution is locally adapted depending on its fit to the prior. Noise variance is reduced at the cost of introducing a bias towards the prior. In the context of image denoising, this bias may result in undesirable artifacts such as spurious structures (with wavelets) or loss of contrast (with total variation minimization) [17]. While these methods are quite expressive, their adaptation to multi-dimensional SAR data is non trivial and usually leads to complex optimization problems [18].

The last category of approaches, relying on adaptive selection of pixels, has seen growing attention over the past years. The generic denoising method proposed in this paper builds on some of the most recent ideas introduced by these approaches.

C. Deledalle is with IMB, CNRS-Université Bordeaux 1, Talence, France, e-mail: charles-alban.deledalle@math.u-bordeaux1.fr

L. Denis is with Université de Lyon, F-42023, Saint-Etienne, France; CNRS, UMR5516, Laboratoire Hubert Curien, F-42000, Saint-Etienne, France; Université de Saint-Etienne, Jean Monnet, F-42000, Saint-Etienne, France. e-mail: loic.denis@univ-st-etienne.fr.

F. Tupin is with Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, Paris, France, e-mail: florence.tupin@telecom-paristech.fr.

A. Reigber and M. Jäger are with the Microwaves and Radar Institute, German Aerospace Center (DLR), Wessling D-82234, Germany, e-mail: {andreas.reigber,marc.jaeger}@dlr.de.

We give a short review of the various methods that have been proposed to perform adaptive selection:

Oriented windows: Lee *et al.* [19]–[21] suggested to locally select the best window among a few pre-defined windows (a rectangular window and 8 edge-aligned oriented windows). Window selection is based on the gradient of the amplitude image. This leads to good preservation of straight edges. However, abrupt change in the decision (from one window to another) at neighboring pixels creates artifacts. The limited number of considered pre-defined windows limits the adaptation to complex structures or textures, and thus the ability to correctly restore them.

Region growing: Vasile *et al.* [22] proposed to use region-growing to build an adapted neighborhood restricted to similar pixels only. Adjacent pixels are aggregated incrementally based on their intensity (hence the name "IDAN": intensity-driven adaptive-neighborhood). This approach is therefore more flexible than the use of pre-defined windows, and leads to better resolution preservation than the previous methods. By construction, adaptive neighborhoods are necessarily connected, and all pixels are given the same weight in the estimation. The method is known to suffer from a selection bias [22]: due to speckle noise, intensities follow a heavy-tailed distribution and selecting pixels with similar intensities discards large values which biases the subsequent maximum likelihood estimation.

Point-wise: The approach for pixel selection can be further generalized by considering extended non-connected neighborhoods by selecting potentially far apart pixels based on their intensities. This idea can be traced-back to the early 80s with the introduction of Lee's sigma filter [23] and latter popularized under the name "bilateral filter" [24], [25]. Similarly to IDAN, the extension of this approach to SAR data suffers from a selection bias that can be corrected using the so-called improved sigma filter [26].

Patch-wise: Rather than selecting pixels with similar intensities, the relative importance of pixels can be weighted by comparing their surrounding neighborhoods (a.k.a., patches). This idea has been popularized in the image processing community by the works of Buades, Coll and Morel [27] under the name NL-means (i.e., non-local means). Most state-of-the-art denoising techniques in image processing now derive from this idea [28], including BM3D considered as one of the most powerful approaches [29]. The method described in this paper is a descendant of NL-means, extended to perform resolution-preserving SAR image restoration. Three different paths have been followed in the literature to adapt non-local methods to SAR data [30]:

- The homomorphic approach, used to extend the NL-means [31] and BM3D [32], first applies a logarithmic transform to the data so that noise becomes additive, then performs a standard non-local filter, and finally applies an exponential transform to map the filtered data back to their original dynamic range. A bias-correction step is necessary to correct for non-Gaussianity in log-space [7].
- The Bayesian approach introduced in [33] interprets the non-local means as posterior means where the posterior densities are measured by comparing patches. This model assumes that a speckle-free image is available and therefore generally requires a pre-filtering step. In [34], the authors highlight that this strong assumption is prone to selection bias which can be corrected with a *sigma-range pre-selection* following the idea of the improved sigma filter [26].
- The statistical approach introduced with the PPB filter [35], and then extended to interferometric SAR [36] and polarimetric SAR [37], considers the pixel selection as a detection problem and builds a statistical test to perform selection. Once similar pixels are detected, the denoising is performed by (weighted) maximum likelihood estimation. This idea has independently been described in [38] for polarimetric SAR images. It has recently been extended following the principles of the successful BM3D approach with SAR-BM3D [39]. Different statistical tests have been proposed for the pixel selection, including: joint-likelihood criteria [35], [36], [40], generalized likelihood ratio tests [38], [41], stochastic and geodesic distances [25], [42]. Some of them are free of selection bias (see Section 3.8 in [43] for more details). Thus, unlike the Bayesian approach, neither pre-estimation of a speckle-free image nor a *sigma-range pre-selection* is mandatory to drive the denoising procedure. Under strong speckle noise, prefiltering can still help to discriminate low-contrast features, as shown with iterative methods [35], [36], [39].

In this paper, we describe a generic framework, called NL-SAR, for non-local denoising of radar images. The method handles amplitude (SAR), interferometric (InSAR), polarimetric (PolSAR) or polarimetric and interferometric (PolInSAR) images in a unified way. The proposed resolution-preserving denoising method brings several novel contributions:

1) *adaptivity to local structures*: our method automatically selects the best local estimate among several computed with different parameters, thus adapting to the scale and the contrast of local structures.
2) *unsupervised method*: by careful weighting of covariance matrices, parameters of the model do not require any tuning related to the noise statistic. Moreover, by considering a wide variety of parameters and automatically selecting locally the best ones, the method is fully automatic.
3) *genericity*: in contrast to approaches requiring either single-look images [36] or multi-looking [38], our method can process single-look and multi-look images without degrading the resolution prior to performing denoising. The identification of similar pixels is performed using the full interferometric and/or polarimetric information, introducing less blur than intensity-only criteria [22], [37].
4) *robustness to noise correlation*: side-lobes of strong echos are often reduced using spectral apodization in radar imagery. This operation correlates noise as a side-effect. Current non-local approaches cannot be applied on correlated noise and require sub-sampling to decorrelate noise. This paper introduces a new way to weight similarities using kernels learned from a homogeneous area selected in the

data. Thanks to these adapted kernels, the proposed method is shown to be robust to speckle correlation.

5) *efficient implementation*: the re-use of some computations to derive estimates with different parameters and our parallel implementation lead to an efficient algorithm that can be applied to large images.
6) *reproducible research*: to facilitate further comparisons and a broad usage of our denoising method, we release under public license the source code of NL-SAR.

The paper is organized as follows: Section II recalls the statistical properties of speckle noise in various SAR image modalities (SAR, InSAR, PolSAR and PolInSAR). Section III then presents the proposed denoising framework. Local adaptivity to the local structures of SAR images is discussed in Section IV. The method is then validated in Section V on airborne and spaceborne images provided by ONERA, Airbus Defence and Space, DLR and ESA.

II. SAR IMAGE STATISTICS

Due to interference among elementary scatterers inside each resolution cell, single-look complex (SLC) SAR images suffer from undesirable fluctuations called speckle.

Single-look SAR images: From D co-registered SLC SAR images, a D-dimensional scattering vector $\mathbf{k}$ is formed at each pixel x, with entries corresponding to the complex amplitudes of the different acquisitions at location x. Under the assumption of fully developed speckle and homogeneity of elementary scatterers inside a resolution cell, the so-formed scattering vector $\mathbf{k}$ follows a D-dimensional circular complex Gaussian distribution [44]:

$$p(\boldsymbol{k}|\boldsymbol{\Sigma}) = \frac{1}{\pi^D|\boldsymbol{\Sigma}|}\exp\left(-\boldsymbol{k}^\dagger\boldsymbol{\Sigma}^{-1}\boldsymbol{k}\right) \tag{1}$$

where $\boldsymbol{\Sigma} = \mathbb{E}\{\boldsymbol{k}\boldsymbol{k}^\dagger\}$ is a $D \times D$ complex covariance matrix with $\mathbb{E}$ the mathematical expectation, † indicates the Hermitian transpose and $|\boldsymbol{\Sigma}|$ stands for the determinant of matrix $\boldsymbol{\Sigma}$. Diagonal elements of $\boldsymbol{\Sigma}$ relate to the radar cross-section of scatterers in each channel. Off-diagonal elements define the complex correlation between channels, and depending on the SAR configuration (InSAR, PolSAR, PolInSAR), correspond to interferometric and/or polarimetric information.

Single-channel images are defined by complex-valued scalars $\mathbf{k} \equiv z$ with z following a zero-mean complex circular Gaussian distribution. The phase of z is thus uniformly distributed, and only its modulus $|z|$, the amplitude, is informative. From eq. (1), it follows that the intensity $|z|^2$ is distributed according to an exponential law.

Hereafter, we will consider that multi-channel images are pre-processed (flat earth fringes removal and orbital inaccuracies correction) so that a flat and homogeneous region is modeled by a constant covariance matrix $\boldsymbol{\Sigma}$.

Multi-look SAR images: Due to the high variability caused by speckle, SLC images have long been spatially averaged at the price of a loss of resolution. Multi-look complex (MLC) images result from the computation of the sample covariance matrix of L scattering vectors $\boldsymbol{k}^{(1)}, \ldots, \boldsymbol{k}^{(L)}$ from a spatial neighborhood centered at pixel x:

$$\boldsymbol{C}(x) = \frac{1}{L}\sum_{t=1}^{L}\boldsymbol{k}^{(t)}\boldsymbol{k}^{(t)\dagger}, \tag{2}$$

where L is referred to as the equivalent number of looks. Note that when $L = 1$, eq. (2) provides a (rank-deficient) covariance matrix representation of SLC data without resolution loss. When $L \geq D$, the distribution of MLC data is described by a complex Wishart distribution given by:

$$p(\boldsymbol{C}|\boldsymbol{\Sigma}) = \frac{L^{LD}|\boldsymbol{C}|^{L-D}}{\Gamma_D(L)|\boldsymbol{\Sigma}|^L}\exp\left(-L\ \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{C})\right) \tag{3}$$

where $\mathrm{tr}(\cdot)$ is the matrix trace. With $D = 1$, eq. (3) simplifies to gamma distribution. When $L < D$, the complex covariance matrix $\boldsymbol{C}$ is singular ($|\boldsymbol{C}| = 0$), and therefore cannot be characterized by a density defined on the open cone of positive definite hermitian matrices. In this case, $\boldsymbol{C}$ is said to have a degenerate distribution. Note that the elements of $\boldsymbol{C}$ are however described term by term by a pdf.

III. GENERIC NON-LOCAL DENOISING OF SAR IMAGERY

Analysis and high-level processing of SAR images require the estimation of the covariance matrix at each pixel of the image. As underlined in section II, covariance matrices carry all the information on the local radiometric, polarimetric and interferometric properties. Due to the dimensionality of the covariance matrices (up to 6×6 matrices in PolInSAR) and the high level of speckle noise in the observed scattering vectors, estimation of covariance matrices requires many samples. Special care must be taken during the selection of those samples, since mixing samples from pixels with distinct radiometric, polarimetric or interferometric properties leads to biased estimations.

We describe in this section the general scheme of our method to perform non-local estimation of covariance matrices. Fig. 1 summarizes the main steps of the method. Starting from a single-look complex SAR image, or a multi-look complex image, we begin by computing a pre-estimation of empirical covariances (section III-A). This pre-estimation is then used to identify similar pixels in a search window. Similarity is defined based on a detection test of identical covariance matrices. Weights are derived from the similarities and used to balance samples in a weighted maximum likelihood estimation procedure (section III-B). In order to achieve a good bias/variance trade-off, non-local estimation is followed by a bias reduction step similar to to the local linear minimum mean square estimation (LLMSE, section III-C). A key ingredient to the performance and robustness of the method comes from the unsupervised selection at each pixel of the best estimate among several estimates obtained by varying parameters of the pre-estimation and non-local estimation. We postpone the presentation and analysis of this last step until section IV.

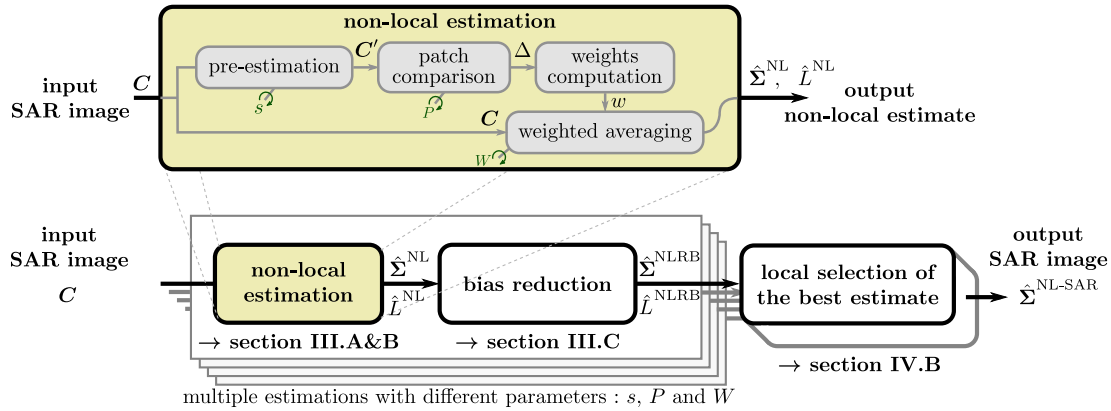


Fig. 1. General scheme of the proposed denoising method

A. Pre-estimation of empirical covariance

Our non-local estimation method performs a weighted estimation where weights are derived from similarity between covariance matrices. It is thus necessary in a first step to compute a pre-estimation of covariance matrices. This pre-estimation will be used only for weights computation. At the non-local estimation step described in section III-B, original (full resolution) input data will be processed. Two issues must be addressed when defining a pre-estimation of empirical covariance: estimation with few samples, and the trade-off between the discrimination power of covariance matrices and the accuracy of their localization.

Estimation of covariance matrices from few samples: When the number of looks L of the original image is smaller than the dimensionality D of scattering vectors, the empirical covariance matrix $\boldsymbol{C}$ is singular and its probability distribution is no longer given by the Wishart distribution (see Section II). In order to derive in Section III-B the similarity between two covariance matrices starting from Wishart distribution, we discuss methods to enforce full-rank to empirical covariance matrices.

Our attempts to regularize $\boldsymbol{C}$ using diagonal loading methods did not provide satisfying results. In [37], we had suggested canceling off-diagonal elements to ensure $\boldsymbol{C}$ to be diagonal. Good performance was already obtained, even though most interferometric and polarimetric information was lost after this processing. This restricted the discrimination capability of covariance similarity criteria based on those matrices.

Instead of canceling off-diagonal elements, the solution proposed here consists in rescaling the off-diagonal elements to ensure that $\boldsymbol{C}$ has full rank. The resulting matrix $\boldsymbol{C}'$ is expressed as follows

$$\forall i, \quad \boldsymbol{C}'_{i,i} = \boldsymbol{C}_{i,i} \quad \text{and} \quad \forall i \neq j, \quad \boldsymbol{C}'_{i,j} = \gamma \boldsymbol{C}_{i,j} \qquad (4)$$

for $\gamma \in [0, 1[$. When $\boldsymbol{C}$ has almost full-rank, i.e. $L \approx D$, $\boldsymbol{C}'$ can be chosen with γ close to 1. When $\boldsymbol{C}$ is strongly rank deficient, i.e. $L \ll D$, γ should be close to 0 to ensure the well-conditioning of $\boldsymbol{C}'$. To achieve this behavior, we have chosen to use the setting $\gamma^3 = \min(L/D, 1)$.

Discrimination vs localization trade-off: Due to strong speckle noise, weakly contrasted structures are difficult to discriminate from surrounding regions. Differences in covariance matrices resulting from such geometrical structures are masked by large fluctuations due to the variance of estimation. Effective restoration however depends on the capability to perform such discrimination. It is then necessary to introduce some pre-filtering to enhance covariance estimation prior to similarity evaluation. One possibility is to re-iterate the non-local estimation procedure, i.e., to use the output of the denoising method as a pre-estimation of covariances, and apply again the non-local procedure using these refined covariances, as done in [35], [36]. A limit of such approach is that the first pre-estimation must be good enough to make it possible to discriminate low-contrast features, otherwise they will be smoothed out after the first non-local step, a "chicken and egg" dilemma. It is faster and more robust to perform such pre-filtering with simple spatial averaging. This averaging improves the discriminative power of covariance matrices by reducing the estimation variance, at the cost of a loss of spatial resolution (i.e., increased bias). Rather than setting a single constant value for the averaging operation (hereafter denoted 'scale' s, corresponding to the radius of the low-pass filter, with $s = 0$ to indicate the absence of pre-filtering), we discuss in section IV how to select locally the best pre-processing among several scales of averaging, thereby achieving an optimal trade-off between discrimination power and localization accuracy.

B. Non-local estimation of covariance

The main mechanism underlying our denoising method is the non-local processing described in this section. In contrast to local estimation techniques, non-local approaches do not select samples that are spatially the closest, but rather samples that most likely follow the same distribution within an extended neighborhood. It is assumed in non-local methods that samples following the same distribution may be identified by finding a collection of patches in an extended search area that are 'similar' to the reference patch centered on the pixel to denoise.

We derive in the following a criterion that measures the closeness of two empirical covariance matrices. Values of this criterion are then transformed into normalized weights thanks to a specially designed kernel. Based on those weights, non-

local estimation is performed with the weighted maximum likelihood estimator.

Dissimilarity criterion: Following a previous study of several methods for deriving similarity criteria adapted to a specified noise distribution [41], we rephrase the problem of evaluating the (dis)similarity of two empirical covariance matrices $\boldsymbol{C}'_1$ and $\boldsymbol{C}'_2$ as a hypothesis test (i.e., a parameter test):

$$\mathscr{H}_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \equiv \boldsymbol{\Sigma}_{12} \qquad \text{(null hypothesis)}, \tag{5}$$

$$\mathscr{H}_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2 \qquad \text{(alternative hypothesis)}. \tag{6}$$

A criterion answering the above hypothesis test with independent Wishart distributed covariance matrices $\boldsymbol{C}'_1$ and $\boldsymbol{C}'_2$ is given by the generalized likelihood ratio (GLR) [45], [46]:

$$\mathcal{L}_G(\boldsymbol{C}'_1, \boldsymbol{C}'_1) = \frac{|\boldsymbol{C}'_1|^{L'} \cdot |\boldsymbol{C}'_2|^{L'}}{|\frac{1}{2}(\boldsymbol{C}'_1 + \boldsymbol{C}'_2)|^{2L'}} \ . \tag{7}$$

Equation (7) provides a measure of the similarity between two empirical covariance matrices $\boldsymbol{C}'_1$ and $\boldsymbol{C}'_2$: large values of the GLR test $\mathcal{L}_G$ indicate that the covariance matrices likely arise from a common distribution. Since the pre-estimation step described in section III-A ensures that matrices $\boldsymbol{C}'_1$ and $\boldsymbol{C}'_2$ be full rank, their distribution can be modeled by the Wishart distribution defined in equation (3), with a number of looks L' higher than the number of looks L of the original data, hence eq. (7) applies.

As mentioned in [41], this criterion boils down to the same expression as the criterion used in the PPB filter for amplitude SAR images [35]. Furthermore, compared to the criterion used in the NL-InSAR filter [36] that is restricted to pairs of complex images, (7) enjoys the same desirable properties [43] and can also be used on multi-looked data. Note that unlike IDAN [22] or the sigma filter [23], this similarity criterion does not suffer from selection bias: it leads to a selection rule of pixel intensities such that the subsequent maximum likelihood estimation is unbiased (see Section 3.8 in [43]). In line with other non-local approaches, similarities are computed over whole patches instead of single pixels in order to decrease the variance of the test. We then define the dissimilarity between two patches centered at the two pixels x and x' as:

$$\Delta(x, x') = \sum_{\tau} -\log \mathcal{L}_G\big[\boldsymbol{C}'(x+\tau), \boldsymbol{C}'(x'+\tau)\big] \ , \tag{8}$$

where $\tau \in [-p, p]^2$ is a 2-D shift indicating the location within each patch of size $P = (2p+1)\times(2p+1)$. The dissimilarity Δ corresponds to the negative logarithm of the GLR expressed on patches, under a pixel-wise independence assumption. Although this independence assumption is not strictly fulfilled in practice due to correlations introduced by the pre-estimation step, dissimilarities can be relatively compared. The translation of dissimilarities into weights through an adapted kernel alleviates the effect of intra-patch correlations, as discussed next.

From dissimilarities to weights: The relative importance of each sample used in the non-local estimation is balanced by weights w derived from the dissimilarities. The mapping of dissimilarities Δ into weights w is done with kernel $\varphi : \mathbb{R}^+ \to [0, 1]$:

$$w(x, x') = \varphi[\Delta(x, x')] \ . \tag{9}$$

The usual way to define the weights from the dissimilarity measure Δ is to use an exponential kernel $w(x, x') = \exp[-\Delta(x, x')/h]$, where $h > 0$ is a filtering parameter. This definition however does not guarantee that a given value of the parameter h produces the same amount of smoothing in homogeneous areas when the pre-estimation is changed (e.g., the scale s of the spatial averaging), or after modification of either the size of the patches P, the number of looks L, or the dimensionality D of scattering vectors. Moreover, the kernel function must account for correlations between empirical covariance matrices to reduce the weight of dissimilar samples.

In [47], the authors have suggested adapting the kernel by setting the parameter h as $h = \mathbb{F}^{-1}(\zeta)$, where $\mathbb{F}^{-1}$ is the inverse cumulative distribution function of Δ under $\mathscr{H}_0$, and ζ is typically set to 0.99. This method ensures that $\mathbb{P}[w(x, x') < e^{-1} \,|\, \mathscr{H}_0] = 1 - \zeta$ whatever the values of P, s, L and D. In this paper, we suggest going one step further by controlling probabilities $\mathbb{P}[w(x, x') < \kappa \,|\, \mathscr{H}_0]$ for all $\kappa > 0$, i.e., we define the kernel φ so that the distribution of weights $w(x, x')$ under $\mathscr{H}_0$ be independent of the patch size P, the reference scale s, the number of looks L and the number of channels D. Our motivation is that in a homogeneous area, the distribution of the weights should depend neither on the noise properties (as L and D) nor on parameters used to retrieve structural features (as P and s). To this end, we define $\varphi = \psi \circ \mathbb{F}$, i.e.,

$$w(x, x') = \psi\left\{\mathbb{F}[\Delta(x, x')]\right\} \tag{10}$$

where $\mathbb{F}$ is the cumulative distribution function of Δ under $\mathscr{H}_0$ and $\psi : [0,\, 1] \to [0,\, 1]$ is another mapping. The values of $\mathbb{F}[\Delta(x, x')]$ within a homogeneous image region follow a uniform distribution on $[0, 1]$, independently of parameters P, s, L and D. Due to the established performance of the exponential kernel in the Gaussian noise context [48], the mapping ψ is chosen so that $\psi \circ \mathbb{F}$ corresponds to the exponential kernel in the case of Gaussian noise. Under Gaussian noise, the dissimilarity is naturally defined as the sum of the square differences and dissimilarities are distributed according to a χ^2 random variable with P degrees of freedom. Let $\mathbb{G}$ denote the cumulative distribution function of the χ^2 law with P degrees of freedom. Dissimilarities Δ are mapped into weights with the same distribution as in the classical Gaussian case with:

$$\psi : \ \mathbb{F}[\Delta(x, x')] \ \mapsto \ w(x, x') = \exp\left[-\frac{\mathbb{G}^{-1}\{\mathbb{F}[\Delta(x, x')]\}}{h}\right], \tag{11}$$

where $\mathbb{G}^{-1}$ is the reciprocal of the χ^2 distribution $\mathbb{G}$. Once $\mathbb{G}$ and h are set, weights computed on images with different number of looks L or dimensionality D can be compared. A homogeneous area is then smoothed similarly whatever the patch size, number of looks, dimensionality or pre-estimation.

As noted by several authors, the weight for the central pixel (i.e., self-similarity $w(x, x)$) should be considered separately [49]. Motivated by the studies in [33], [49]–[51], we use the

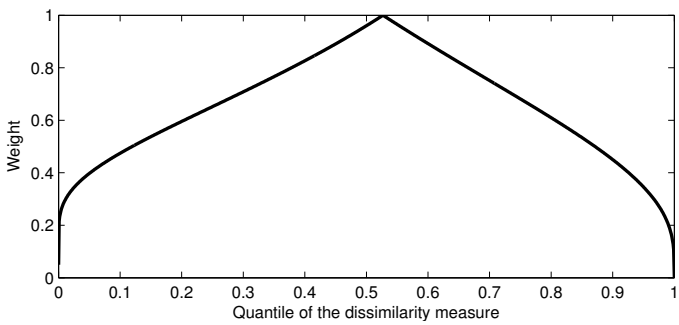


Fig. 2. Mapping $\psi : \mathbb{F}[\Delta(x,x')] \mapsto w(x,x')$ used to impose a target distribution to the weights.

following modified kernel:

$$w(x,x') = \begin{cases} \exp\left(-\frac{\left|\mathbb{G}^{-1}\{\mathbb{F}[\Delta(x,x')]\}-c\right|}{h}\right) & \text{if } x \neq x' \\ 1 & \text{if } x = x', \end{cases} \quad (12)$$

with $c = \mathbb{E}\left[\mathbb{G}^{-1}\{\mathbb{F}[\Delta(x,x')]\}\middle|\mathscr{H}_0\right]$, i.e., the expected transformed dissimilarity of two patches following the same distribution. Since $\mathbb{F}[\Delta(x,x')]$ has a uniform distribution, the expression of c boils down to the degrees of freedom of $\mathbb{G}$. Unlike eq. (11), this kernel ensures that, when h tends to 0, there is almost surely only one non-zero weight: the central one $w(x,x)$, and the output tends to the noisy input image (no denoising). This definition also prevents any pixel from having a larger weight than the central pixel.

The mapping ψ defined by eq. (12) has been set in NL-SAR with $h=1/3$ and $\mathbb{G}$ the χ^2 distribution with 49 degrees of freedom (up to a change of variables, this corresponds to the set of parameters for Gaussian NL-means used in [52] with 7×7 patches) resulting in the curve given in figure 2. Unlike h in the traditional NL-means, the choice of ψ can be kept fixed even if P, s, L and D vary. The adaptation to these latter parameters is implicitly done through $\mathbb{F}[\Delta(x,x')]$. The distribution of dissimilarities Δ is modified when noise is correlated. By sampling the probability distribution of Δ directly on the data (on a homogeneous area selected by the user), a kernel adapted to noise distribution and robust to noise correlation can be learned (see sections IV-C and V-C). Unlike the traditional exponential kernel, the proposed kernel $\varphi = \psi \circ \mathbb{F}$ can thus be considered as noise-aware. Note that the kernel φ can be learned once and re-used on several images coming from the same sensor with similar observation conditions.

Non-local estimation with weighted maximum likelihood: The original non-local means method [27] has been introduced to denoise images corrupted by additive white Gaussian noise. After computation of weights $w(x,x')$ based on squared differences between patches and an exponential kernel, the non-local means algorithm performs a weighted averaging. This method has been extended to more general estimation problems in [35], [53] by introducing a weighted maximum likelihood:

$$\hat{\boldsymbol{\Sigma}}^{\mathrm{NL}}(x) = \arg\max_{\boldsymbol{\Sigma}} \sum_{x'} w(x,x')\, p\big(\boldsymbol{C}(x') \mid \boldsymbol{\Sigma}\big), \quad (13)$$

where weights are defined based on similarities as described in previous paragraph, and the sum is carried out over all pixels x' in an extended neighborhood of size $W = (2w+1)\times(2w+1)$ centered at pixel x and called the search window. Note that equation (13) involves only covariance matrices $\boldsymbol{C}$ from the full resolution input image, not the pre-estimated covariances $\boldsymbol{C}'$ defined in section III-A. The resolution of the original image is thus preserved.

Since empirical covariance matrices $\boldsymbol{C}(x')$ all have the same number of looks L, the weighted maximum likelihood estimator is given by the weighted means:

$$\hat{\boldsymbol{\Sigma}}^{\mathrm{NL}}(x) = \frac{\sum_{x'} w(x,x')\boldsymbol{C}(x')}{\sum_{x'} w(x,x')}. \quad (14)$$

C. Bias reduction step

A peculiarity of SAR images is their very high dynamic range. Bright targets have intensities several orders of magnitude larger than their surrounding background. Even though a patch containing such a bright target at pixel x' and a patch with background-only are very dissimilar and the corresponding weights $w(x,x')$ are very low, the weighted mean given in equation (14) creates some blurring of the bright target. When estimating the covariance at a pixel x in the background, the contribution $w(x,x')\boldsymbol{C}(x')$ is indeed not negligible due to the large magnitude of the bright target at pixel x'.

In order to reduce the spreading of bright structures, we add a bias-reduction step after the non-local estimation. Following [19], bias is reduced by performing a convex combination between the (possibly over-smoothed) non-local estimation and the noisy empirical covariance:

$$\hat{\boldsymbol{\Sigma}}^{\mathrm{NLRB}}(x) = \hat{\boldsymbol{\Sigma}}^{\mathrm{NL}}(x) + \alpha\left[\boldsymbol{C}(x) - \hat{\boldsymbol{\Sigma}}^{\mathrm{NL}}(x)\right] \quad (15)$$

where $\hat{\boldsymbol{\Sigma}}^{\mathrm{NLRB}}(x)$ is referred to as the Non-Local Reduced Bias (NLRB) estimate. Values of the weight α close to zero keep the non-local estimate while values close to one replace the non-local estimate with the original (noisy) empirical covariances. The value of α must then be set according to the confidence in the non-local estimate.

The quality of non-local estimates depends on whether candidates $\boldsymbol{C}(x')$ have been selected correctly, i.e., if the candidates $\boldsymbol{C}(x')$ and the empirical matrix $\boldsymbol{C}(x)$ are samples arising from the same distribution. Over-smoothed images and spread bright targets appear when the non-local averaging has mixed heterogeneous samples. According to speckle statistics, the variance of the intensity in a homogeneous region is equal to $\bar{I}^2/L$, with $\bar{I}$ the average intensity and L the number of looks in the original data. If the variance of the intensity in a collection of samples is much larger than $\bar{I}^2/L$, then this is an indication that all the variance cannot be attributed to speckle and that part of the variance comes from heterogeneity of the samples. A natural test is then to compare the (weighted) variance with the squared (non-local) intensity. Let $j \in \{1,\ldots,D\}$ be a given channel. We define the weighted variance at pixel x for channel j by:

$$\widehat{\mathrm{Var}\,[I_j]}^{\mathrm{NL}}(x) = \frac{\sum_{x'} w(x,x') I_j(x')^2}{\sum_{x'} w(x,x')} - \hat{I}_j^{\mathrm{NL}}(x)^2, \quad (16)$$

where $I_j(x) = \boldsymbol{C}_{j,j}(x)$ denotes the j-th diagonal element of $\boldsymbol{C}(x)$, and correspondingly $\hat{I}_j^{\mathrm{NL}}(x) = \hat{\boldsymbol{\Sigma}}_{j,j}^{\mathrm{NL}}(x)$. The non-local variance should be close to $\hat{I}_j^{\mathrm{NL}}(x)^2/L$ if most significant weights $w(x,x')$ correspond to pixels x' following the same law as pixel x (i.e., if $\boldsymbol{\Sigma}(x') \approx \boldsymbol{\Sigma}(x)$).

The value α in (15) is then chosen according to this homogeneity test. This is indeed the strategy of the Local Linear Minimum Mean Square Estimator (LLMMSE) introduced in [19] that defines a value α^{LLMMSE} following this principle and that achieves an optimal bias-variance trade-off (i.e., minimum mean square error). Here, we suggest using a slightly different definition of α given by

$$\alpha^{\mathrm{NLRB}} = \max_j \left[\max \left(0, \frac{\widehat{\mathrm{Var}\,[I_j]}^{\mathrm{NL}}(x) - \hat{I}_j^{\mathrm{NL}}(x)^2/L}{\widehat{\mathrm{Var}\,[I_j]}^{\mathrm{NL}}(x)} \right) \right] \quad (17)$$

related to the LLMMSE parameter through $\alpha^{\mathrm{NLRB}} = (1 + 1/L)\alpha^{\mathrm{LLMMSE}}$. Our definition of α works then at a different bias-variance trade-off than the LLMMSE procedure. It provides a lower bias and a higher variance. In the mono-dimensional case ($D = 1$), while α^{LLMMSE} lies in the range $[0, 1/(1 + 1/L)[$, α^{NLRB} lies in $[0, 1[$. Hence, unlike the definition of LLMMSE, ours ensures that the non-local estimate is preserved ($\alpha \approx 0$) when candidates $I(x')$ have been selected correctly, while it is fully rejected ($\alpha \approx 1$) when $\widehat{\mathrm{Var}\,[I_j]}^{\mathrm{NL}}(x) \gg \hat{I}_j^{\mathrm{NL}}(x)^2/L$. In the multi-dimensional case ($D > 1$), our strategy differs from the LLMMSE procedure described in [20] which is performed on the span of the covariance matrices. A drawback of the span image is that its expected variance is unknown since it depends on the inter-channel correlations. Our approach expects a variance of $\hat{I}_k^{\mathrm{NL}}(x)^2/L$ independently on each channel. It has proved more effective in our experiments as it involves no prior knowledge of inter-channel correlations.

Fig. 3 illustrates some results of the non-local estimation and bias-reduction steps on a synthetic image. This test image represents 2 different geometrical structures: two bright points of different sizes with a large magnitude of 30dB (the actual contrast is not seen on the images due to the display range); and a periodic curved linear structure with a contrast of 5dB. The first 4 images in the first row of figure 3 show the result of non-local estimations with different settings of the patch size P, and the scale s used for the averaging operation of the pre-estimation step. It is observed that if the scale s is too small, the non-local estimation leads to a blurry image. The reason is that dissimilarities have poor discriminative power when pre-estimations are performed with too few samples, as discussed in section III-A. The non-local estimation step then mixes pixels from different distributions (i.e., from dark and light areas), resulting in either a blurry estimation (e.g., the curve in the image on the first column), or aberrant values (e.g., around the left bright target in the image on the second column). The 4 images in the third row of figure 3 show the result of the bias-reduction procedure just described. It is observed in the cases of a too small scale s that much of the original resolution is restored by the bias-reduction step. Non-local estimations for other settings are also improved, especially near the edges of structures that were slightly blurred. Next section further analyzes the effect of each parameter and describes a method for unsupervised local selection of the best parameters, providing the result shown on the last column of the figure.

IV. Unsupervised adaptation to local structure

The non-local denoising procedure described in section III requires to set several parameters: the scale s of the pre-estimation, the size of patches P and the size of the search window W. We show in section IV-A that, depending on the considered geometrical structure, different parameters should be preferred. Section IV-B describes how to automatically select locally the best parameters.

A. No global tuning can preserve all structures

A major limitation of evolved denoising techniques such as non-local techniques is the number of parameters that must be jointly tuned, and even more critical, the need to adapt those parameters to the nature of the image (modality, scale) and its content (smooth natural regions or man-made areas). We illustrate in this section that, even if the user is prepared to tune all parameters by hand, a unique set of parameters cannot preserve correctly the diversity of structures met in an entire radar image.

Fig. 3, first described in section III-C to illustrate the gain of the bias-reduction step, gives the result of non-local estimation for different tunings of parameters and various geometrical structures. Among others, the figure illustrates the case of two strong scatterers (30dB). Very bright point-like structures are numerous in SAR images. They are problematic for patch-based non-local methods since no patch similar to a patch containing an isolated bright point is found in its neighborhood, a problem referred to as the *rare patch effect* in the literature [54]–[56]. As a consequence, not only the covariance at the point itself but all surrounding covariances in an area of size $(2p+1)\times(2p+1)$ are left unchanged (since no similar covariance matrices have been found to perform the average). It results in very noisy regions surrounding the two bright dots in the non-local estimations, especially for larger values of the scale s (see, the column 2, 3, 4 figure 3). Since in those areas, the variance is too large, the bias-reduction step is of no help to reduce this phenomenon (still visible in the third row of figure 3). Using a small patch size significantly reduces the size of the affected region. When curved linear structures are considered, the use of the same small patch size leads to artifacts because the patches are then too small to capture the local geometry (curvature). The size of the patches thus cannot be kept constant on images containing different kinds of structures, as already shown in [55].

Thanks to the procedure described in section III-B to enforce an invariant distribution of weights, no smoothing parameter needs to be adapted each time a parameter is changed and the same amount of smoothing is kept in homogeneous areas. This is observed in the background areas that show similar variances with very different settings.

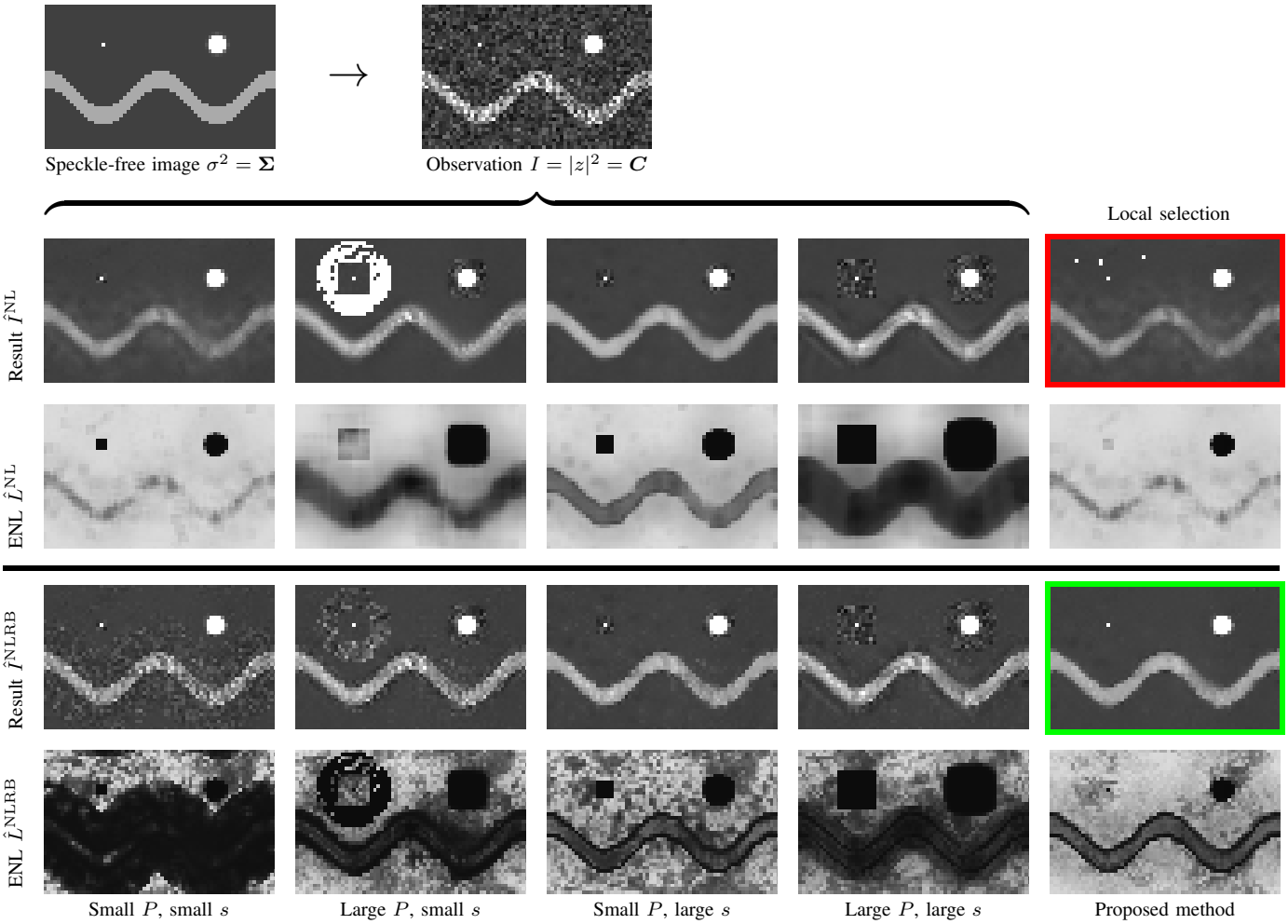


Fig. 3. Local section of the best estimate among several non-local estimations obtained without (upper part of the figure) or with a bias reduction step (lower part). The first 4 columns show the denoising results and estimated equivalent number of looks (ENL) for 4 different parameter settings. The last column gives the final result obtained by selecting at each pixel the estimate with the largest ENL (in red: from denoising results computed without bias reduction, in green: proposed approach with bias reduction).

Here, the same large search area has been kept to produce the first four columns in figure 3. Larger search windows seem always preferable to smaller ones in our simple test cases. In more complex cases, it is beneficial to adapt locally the search window size to the local content since larger search windows would necessarily bring in dissimilar patches that would degrade the estimation [47].

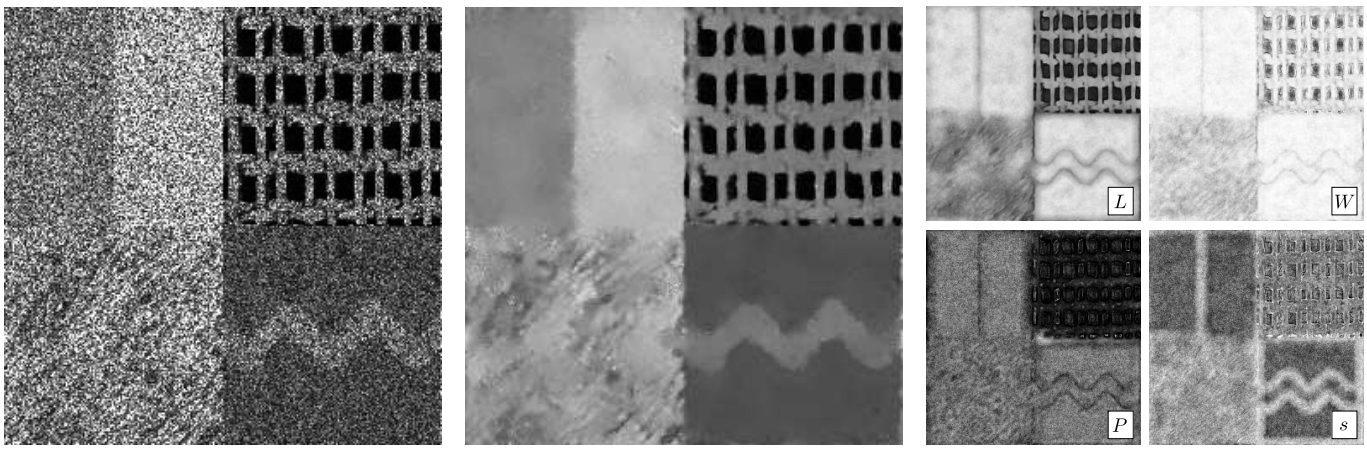
B. Local selection of the best estimate

Building a new estimate out of several is a well studied problem in statistics referred to as "aggregation" [57]. The new estimator may be formed as a convex or linear combination of existing estimators, or by selecting the best one according to a given criterion (an approach called "model selection"). In non-local filtering, several selection/aggregation procedures have been used. For selecting locally the best search window size, the authors of [47] employ the so-called ICI rule selection (a.k.a., Lepski method). This method requires that the estimators can be sorted in terms of bias-variance trade-off. This order is easily induced when only the search window size varies, but does not exist in our context. Other approaches have used statistical estimators of the reconstruction error (the so-called SURE for Stein unbiased risk estimator). The authors of [58], [59] suggest selecting the best global parameters w.r.t. SURE while in [55], [60] the selection is locally adaptive. Such estimators are difficult to use in our context since, first, local estimations of risk generally suffer from a large variance, and, next, they are not trivial to extend to non-scalar data and non-Gaussian noise distributions. Here, we follow an approach based on variance reduction similar to [61] in order to select locally the estimate considered as the most reliable. We first derive our criterion for selecting the best estimate, then illustrate its performance on a synthetic image.

Performing the average of L independent and identically distributed (i.i.d.) samples reduces the variance by a factor L:

$$\mathrm{Var}\left[\frac{1}{L}\sum_{t=1}^{L} I^{(t)}\right] = \frac{\mathrm{Var}\left[I\right]}{L}\,. \tag{18}$$

If a weighted average is done instead, the variance is reduced



(a) (b) (c)

Fig. 4. Local adaptivity of the proposed approach: (a) Noisy image ($L = 1$); (b) Result of the proposed approach; (c) Averages on 20 noise realizations of the resulting equivalent number of looks $\hat{L}^{\mathrm{NLRB}}$ (range: $[0, \pi(25/2)^2]$), the search window size W (range: $[0, \pi(25/2)^2]$), the patch size P (range: $[3^2, 11^2]$), and the reference scale s (range: $[0, 2]$).

by a factor $\hat{L}^{\mathrm{NL}}$ depending on the weights (see Appendix A):

$$\hat{L}^{\mathrm{NL}}(x) = \frac{\left(\sum_{x'} w(x,x')\right)^2}{\sum_{x'} w(x,x')^2} \approx \frac{\mathrm{Var}\left[I(x)\right]}{\mathrm{Var}\left[\hat{I}^{\mathrm{NL}}(x)\right]}\,. \tag{19}$$

$\hat{L}^{\mathrm{NL}}(x)$ represents an equivalent number of looks at pixel x after the non-local estimation step but before the bias-reduction procedure. Equation (19) corresponds to the variance reduction of the intensity $I(x)$ taken on the diagonal of the covariance matrix $\boldsymbol{C}$ under the assumption that all samples have the same variance and by considering weights as deterministic.

As discussed in section III-C, the non-local estimation step tends to over-smooth in some regions due to the lack of discriminative power of the weights. Samples are then no longer identically distributed, but distributed according to distributions with different parameter values, and the resulting estimation is biased. The confidence on a given estimate cannot rely on $\hat{L}^{\mathrm{NL}}(x)$ only but must account for the bias. After the bias-reduction step, bias is strongly reduced and a bias-variance trade-off is achieved. The equivalent number of looks after bias-reduction depends on the weight α of the noisy covariance in equation (17) (see Appendix A):

$$\hat{L}^{\mathrm{NLRB}}(x) = \frac{\hat{L}^{\mathrm{NL}}(x)}{(1-\alpha)^2 + \left(\alpha^2 + \frac{2\alpha(1-\alpha)}{\sum_{x'} w(x,x')}\right)\hat{L}^{\mathrm{NL}}(x)} \tag{20}$$

where we consider $w(x,x) = 1$. This equivalent number of looks gives an indication of the quality of an estimate. Values of $\hat{L}^{\mathrm{NLRB}}(x)$ are computed at each pixel x for all sets of parameters and the estimate leading to the largest value is selected. This defines a local selection method of the best parameters. Our selection follows the spirit of [61], the main difference being that we prevent from large bias (i.e., over-smoothing) by applying a bias-reduction step before selection.

The map of equivalent numbers of looks for each non-local estimate is displayed in figure 3. After bias reduction, the equivalent number of looks is reduced in all areas that were over-smoothed. Locally selecting the estimation with the largest equivalent number of looks gives the result displayed in the last column "Local selection" on the right of the figure. Applying this selection rule directly to non-local estimates (i.e., before bias reduction) favors the strongest smoothing and then gives a blurry result with artifacts around the isolated point source (image framed in red in figure 3). By contrast, the bias-reduction step produces a constant bias-variance trade-off and selection based on the equivalent number of looks gives an obviously preferable result (image framed in green in figure 3). This result obtained with local selection is of better quality than any result using a fixed set of parameters.

To further analyze the local adaptivity of the parameters, we consider another synthetic example shown in figure 4. The original image is composed of 4 different regions: (i) two homogeneous regions separated by a straight boundary; (ii) a texture with quasi-periodic structure; (iii) a textured area; (iv) a thick curved line. The noisy image is shown in sub-figure (a) and the restored one in sub-figure (b). Sub-figure (c) gives maps of the equivalent number of looks, search window size W, patch size P, and smoothing scale s, averaged over 20 noise realizations. It is noticeable that none of the selected parameters are constant over the whole image, which confirms *a posteriori* the necessity of local adaptivity. Unsurprisingly, noise is most efficiently reduced in homogeneous areas where large search windows are preferred. At a discontinuity, such as the vertical line between the two homogeneous regions in the first quadrant or the edges of the curved line, noise is less reduced because fewer similar patches are found. When increasing the size of the search window, only two more similar patches are found while many dissimilar patches are added to the search area (a number proportional to the size of the search window). Intermediate search window sizes are thus preferred. To decrease the variance of the estimation close to discontinuities it would be necessary to consider anisotropic search windows (straightforward using our selection method but computationally costly). To reduce as much as possible the influence of the many dissimilar patches of the search window, it is necessary to strongly smooth in the pre-estimation step

TABLE I
PROCESSING TIME ON A CLUSTER OF INTEL® XEON® X5550 2,66 GHZ

$M \times N$	D	1 core	8 cores	8×8 cores
256^2	1	55s	14s	7s
	2	1min 6s	17s	9s
	3	1min 29s	24s	13s
1024^2	1	15min	2min 27s	38s
	2	18min	3min 16s	48s
	3	26min	4min 51s	1min 13s
4096^2	1	4h 21min	40min	6min 4s
	2	5h 18min	54min	9min
	3	7h 24min	1h 20min	14min

in order to improve the discriminative power of the similarity measure (large s are selected). Finally, in textured areas, the patch size must be set so that enough similar patches are found. Smaller patches are thus preferred in the second quadrant so as to capture the edges and prevent their blurring.

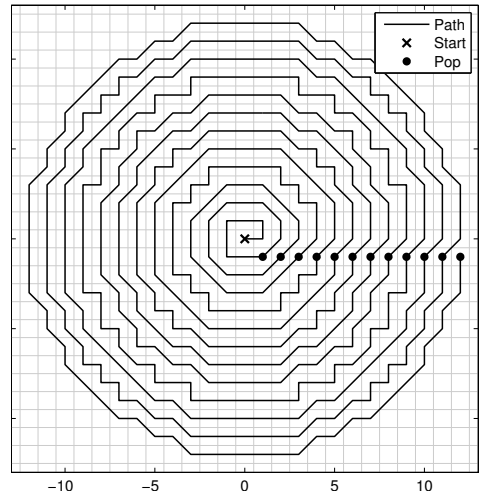


Fig. 5. Spiral path inside the circular search window. The search starts at the center and the results are "popped" after each rotation.

C. Implementation details

Thanks to the use of a carefully designed mapping of dissimilarities into weights and our unsupervised procedure to locally select the best set of parameters, our denoising method can be considered as fully automatic. No adaptation of parameters is required when going from SAR intensity denoising to PolSAR or InSAR images. When changing the scene and image type (from vegetated areas to man-made structures, or with different sensor resolutions), the method selects the optimal parameters provided that the sets $\mathbb{W}$ of search window sizes, $\mathbb{P}$ of patch sizes, and $\mathbb{S}$ of scales are large enough. A key condition for the practical usability of our denoising framework is the possibility to perform all the estimations involved within a reasonable time-frame. We describe in this section an implementation that can process a 2048×2048 pixels polarimetric image (requiring the estimation of over 150 million values) in less than a quarter of an hour on a cluster of 64 cores. Compared to methods that require human interaction and several runs of the denoising algorithm to tune the parameters, this computation time seems reasonable. The source code of our parallel implementation is released under public license together with the paper[1]. The code can be used in command line, called from `IDL`, from `Matlab`, from `PolSARpro`, from `Python` or used as a `C` library.

In our denoising framework, many non-local estimations must be performed. Our efficient implementation is based on 3 ingredients:

- computational savings when evaluating the dissimilarities thanks to integral tables [62];
- re-use of estimates computed with smaller search windows;
- natural parallelization of non-local methods, in contrast to regularization methods that require joint estimation of all pixels.

The proposed algorithm works as follows: non-local estimations are computed for increasing sizes of the search window. In order to have isotropic search windows, the central reference patch is compared with patches extracted following a spiral path. For each position in the spiral, non-local estimates for all patch sizes in $\mathbb{P}$ and all pre-processing scales in $\mathbb{S}$ are updated using integral tables [62]. After each rotation of the spiral, a new set of non-local estimates for the corresponding search window size is obtained (see figure 5). The equivalent number of looks is computed and if it improves on the best-so-far number of looks, the corresponding best non-local estimate is updated.

The computation of the weights is performed as follows. A (noisy) homogeneous area is either provided by the user or simulated based on speckle statistics. Patches are extracted from that homogeneous area in order to sample the probability distribution of Δ. Dissimilarity values $\Delta(x, x')$ are estimated off-line for all pairs of patches (x, x'). These dissimilarities are stored in a separate table sorted in increasing order for each possible choice of the patch size P and pre-estimation scale s. On-line, when the weight $w(x, x')$ must be evaluated from the dissimilarity $\Delta(x, x')$ evaluated with given parameters P and s, the closest value to $\Delta(x, x')$ is found in the corresponding sorted table by binary search. The rank of that value in the table (i.e., its index) gives the corresponding quantile $\mathbb{F}[\Delta(x, x')]$. The final value of the weight $w(x, x') = \psi\{\mathbb{F}[\Delta(x, x')]\}$ is then obtained from a look-up table built off-line by uniform sampling of ψ function on the range $[0, 1]$. Overall, the evaluation of the weight $w(x, x')$ from $\Delta(x, x')$ requires $\log_2(K)$ iterations where $K = 2^{10}$ is the size of the sorted table encoding $\mathbb{F}$. Our experiments have shown that the computation time to evaluate our kernel function is of the same order as the computation time required to evaluate the exponential kernel based on the floating exponential function `expf` of the `C` library.

For an image of size $M \times N$, the original NL-means algorithm has an algorithmic complexity of order $\mathcal{O}(MNWP)$ and a memory complexity of order $\mathcal{O}(MN)$. Our non-local method is based on the local selection of the best estimate among $|\mathbb{W}| \cdot |\mathbb{S}| \cdot |\mathbb{P}|$ non-local estimates. Thanks to the re-use of previous computations, the algorithmic complexity of our algorithm is $\mathcal{O}(MNW_{\max} \cdot |\mathbb{P}| \cdot |\mathbb{S}|)$, where $W_{\max}$ is the maximum search window size. Since the number of pairs of

[1] http://www.math.u-bordeaux1.fr/~cdeledal/nlsar.php

parameters (s, P) that are considered is of the same order of magnitude as the number of pixels in a patch (from a few tens to a few hundreds), the complexity of our algorithm is on a par with the complexity of the original NL-means algorithm. In terms of memory usage, our algorithm requires the storage of $\mathcal{O}(MN|\mathbb{P}||\mathbb{S}|)$ values. In the case of scattering vectors of dimension D, covariance matrices of size $D \times D$ must be estimated and all complexities are multiplied by a factor D^2.

Our `C` implementation has been parallelized with `OpenMP` and distributed with `Portable Batch System` leading to computation times reported in Table I.

V. Experiments and results

We illustrate in this section the capability of the proposed non-local denoising framework to handle different kinds of images (intensity only, InSAR, PolSAR or PolInSAR). We first evaluate the quantitative performance of the proposed estimator on numerical simulations and compare it to other state-of-the-art approaches. We next give qualitative results on airborne InSAR and PolSAR data. For a more accurate assessment of the qualitative performance of NL-SAR, additional results can be found in [63]. We also invite the reader to visit the dedicated web page http://www.math.u-bordeaux1.fr/~cdeledal/nlsar.php.

All results are obtained with the unsupervised local selection method of the best non-local estimate. Except in the specific case of correlated noise discussed in section V-C, we used the same set of parameters to get all the results:

- $\mathbb{W} = \{3^2, 5^2, 7^2, \ldots, 25^2\}$,
- $\mathbb{P} = \{3^2, 5^2, 7^2, 9^2, 11^2\}$,
- $\mathbb{S} = \{0, 1, 2\}$.

The pre-estimation step at scale s is performed by convolution with a truncated Gaussian defined by $K \exp\left(-\frac{\pi(x^2+y^2)}{(s+0.5)^2}\right)$ if $-s \leq x, y \leq s$ and 0 otherwise, where K is a normalization constant. Recall that $s = 0$ indicates the pre-filtering is off.

A. Bias-variance characterization of the estimator

Restoration methods in SAR imaging can be compared based on their ability to reduce noise variance while preserving the resolution. In order to evaluate this bias-variance trade-off, we performed a Monte Carlo study of several speckle reduction methods. Fig. 6 illustrates the capability of 4 different methods to restore bright sources or repeating squares. The input intensity-images provided to each speckle reduction method were generated using Goodman's speckle model (i.e., intensities follow an exponential law, see section II). Fig. 6 illustrates both the typical output of each method, as grayscale images, and the bias-variance characteristics on a line profile. The 4 denoising methods are: (b) the pretest non-local method described in [38], (c) the iterative version of probabilistic patch-based denoising (PPB-it) [35], (d) the SAR-BM3D method proposed in [39] as an extension to radar imaging of the original BM3D Gaussian denoising method [29], and (e) the NL-SAR method that we propose in this paper. Line profiles show the expectation of each estimator as well as 0.98% confidence intervals. The line profiles drawn in the top row correspond to two bright targets. Given the very high dynamic range, a logarithmic scale is used for the intensity axis of the line profiles and the corresponding intensity images are saturated. The images with repeating squares have a much weaker contrast and a linear scale is used.

It is observed from the expectation of the estimators in the case of the two bright targets (first row of figure 6) that pretest method (b) and BM3D-based method (d) introduce some bias around the point-like bright source. PPB-it and the proposed method introduce almost no bias. For all patches containing the point-like target, no other similar patch is found ("rare patch" phenomenon). This results in large residual variances around the bright targets for pretest and PPB-it methods. Thanks to the adaptive selection of the best parameters, our method effectively reduces speckle noise in the background without spreading bright targets. The Monte Carlo simulations performed on images of repeating squares with weak contrast illustrate also that the proposed method leads to the strongest reduction of noise variance with introduction of very limited bias. Compared to PPB-it, our method slightly blurs the edges but offers a more efficient reduction of noise.

B. Denoising performance on numerical simulations

We further compare our method to state-of-the-art speckle reduction techniques using images of a resolution target corrupted by speckle. We considered different types of data: intensity images, interferometric images and polarimetric images.

The first part of Fig. 9 illustrates the denoising performance in the case of intensity-only SAR images. The original resolution target is shown in figure 9(a) and a single-look noisy realization in figure 9(b). The output of 4 iterations of the probabilistic patch-based PPB-it non-local method [35] is shown in figure 9(c). This method strongly reduces noise variance in homogeneous areas but blurs the smallest details and tends to introduce some artifacts along discontinuities. The second considered state-of-the-art method is the recent SAR-BM3D technique [39]. This method produces a strong variance reduction with excellent preservation of smallest details, see figure 9(d). It can be considered as the best speckle reduction method for intensity images to date [64]. Some artifacts can be noticed on some discontinuities with artificial edges created parallel to the actual edges. The extension of this method to multi-channel images (InSAR, PolSAR or PolInSAR) is not straightforward and has not been proposed yet. Finally, the proposed method produces a result with quality close to that of SAR-BM3D, see figure 9(e). Tiniest details are lost and edges tend to be slightly jagged compared to the output of SAR-BM3D. No other systematic artifact are noted.

The performance in an interferometric configuration is illustrated on the second part of figure 9 where the same underlying grayscale image is used to generate a two-channel dataset with two-level coherence and phase. Pointwise estimation of the radiometry, interferometric phase and coherence is very noisy, as shown in figure 9(g). We applied Lee's refined filter [21], figure 9(h), and the intensity-driven region growing method IDAN [22], figure 9(i). The output of the proposed method is shown on the last column (j). Both Lee's refined filter and

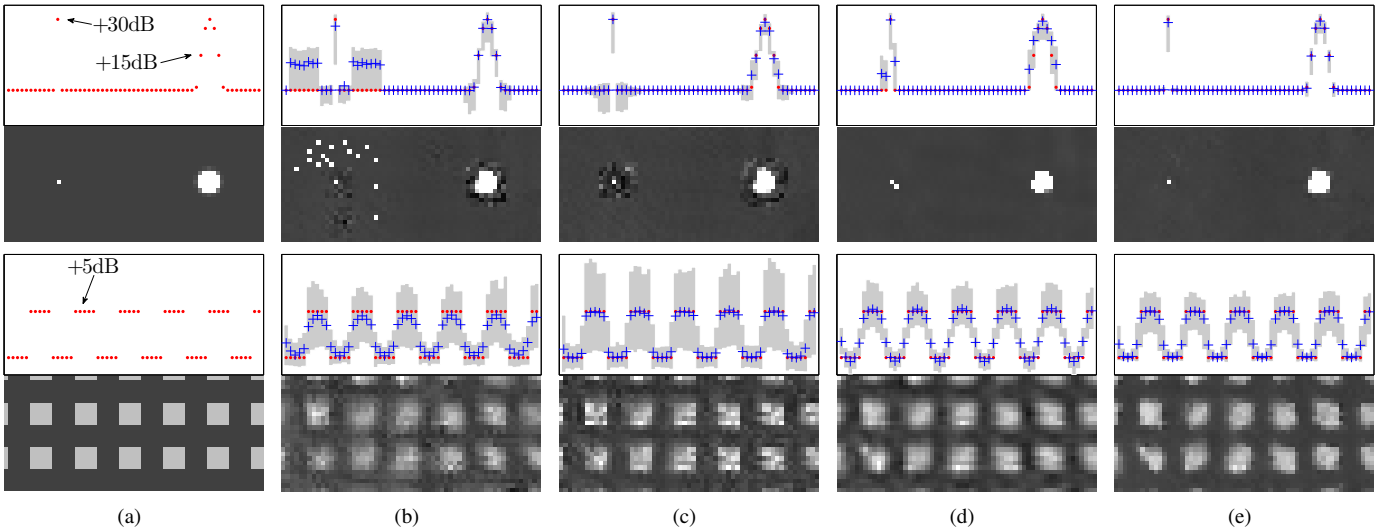


Fig. 6. Bias-variance characterization of the several speckle reduction methods for single look SAR intensity denoising: (a) underlying signal, (b) Pretest non-local filter [38], (c) PPB-it [35], (d) SAR-BM3D [39], (e) NL-SAR proposed in this paper. Two types of structures are analyzed: bright targets (upper half) and repeating squares (lower half). Grayscale images show the output of each denoising method for a given noisy realization. Above each grayscale image, line profiles corresponding to the expectation (blue crosses) and 0.98% confidence intervals (gray area) of each estimator are drawn. Line profile intensities on the top row are drawn in log-scale to adapt to the high dynamic range.

IDAN leave noticeable fluctuations in homogeneous regions. Lee's filter does not restore small details. IDAN better restores small details and leaves less residual variance. Tiniest details however disappear due to selection bias. The proposed method offers very strong noise reduction with good preservation of even the tiniest details. In this numerical experiment, the co-location of edges in radiometry, interferometric phase and coherence favors our method since the definition of the similarity criterion jointly on the interferometric covariance matrix leads to stronger discrimination power than a criterion based solely on intensities.

The performance of the method on polarimetric images is illustrated on the third part of figure 9 where polarimetric covariance matrices have been set to represent 3 different configurations found in PolSAR images: vegetation-type areas that fully de-polarize the incident wave (appear in green when using Pauli basis coloring and have a high polarimetric entropy as defined in [65]), surfaces or corner reflectors that produce single or triple bounces (appear in red in Pauli basis coloring), and ground-wall creating double-bounces (appear in blue). Pointwise estimation of polarimetric properties shown in figure 9(l) are too noisy to be used directly. The output of IDAN filter is shown in figure 9(m). Noise variance is reduced at the cost of a loss of small details. The pretest non-local method introduced in [38] offers both a stronger variance reduction and a better preservation of details. The proposed method reaches the strongest variance reduction with notably improved estimation of entropy in the area with the smallest details.

C. Robustness to noise correlation

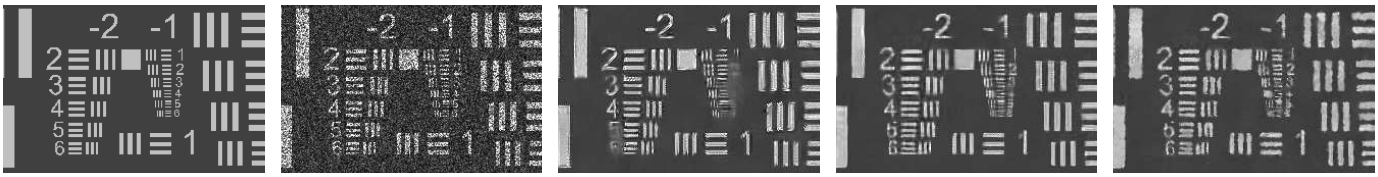
The applicability of a SAR restoration method depends on its robustness to changes in the image source (sensor and processing chain used for SAR synthesis). One major issue for existing restoration methods is speckle correlation. Noise correlation has several impacts on non-local methods: first, similarities are over-estimated and less robust for a given patch-size (the effective number of independent measurements being smaller than the number of pixels in each patch), second averaging N neighboring pixels decrease the variance by a factor less than N so that more samples are needed to provide the same amount of smoothing.

Over-estimation of similarities is addressed in our method by learning the kernels used to map similarities into weights on a homogeneous area provided by the user. When noise is correlated, the probability distribution of dissimilarities Δ is modified. This is accounted for by using the cumulative distribution function $\mathbb{F}$ in the definition of the kernels. Noise correlation is easily detected by analyzing a homogeneous area. If correlation is detected, our algorithm adapts the range of parameters:

- $\mathbb{W} = \{3^2, 7^2, 11^2, \ldots, 49^2\} \div 4$
- $\mathbb{P} = \{3^2, 7^2, 11^2, 15^2, 19^2\}$

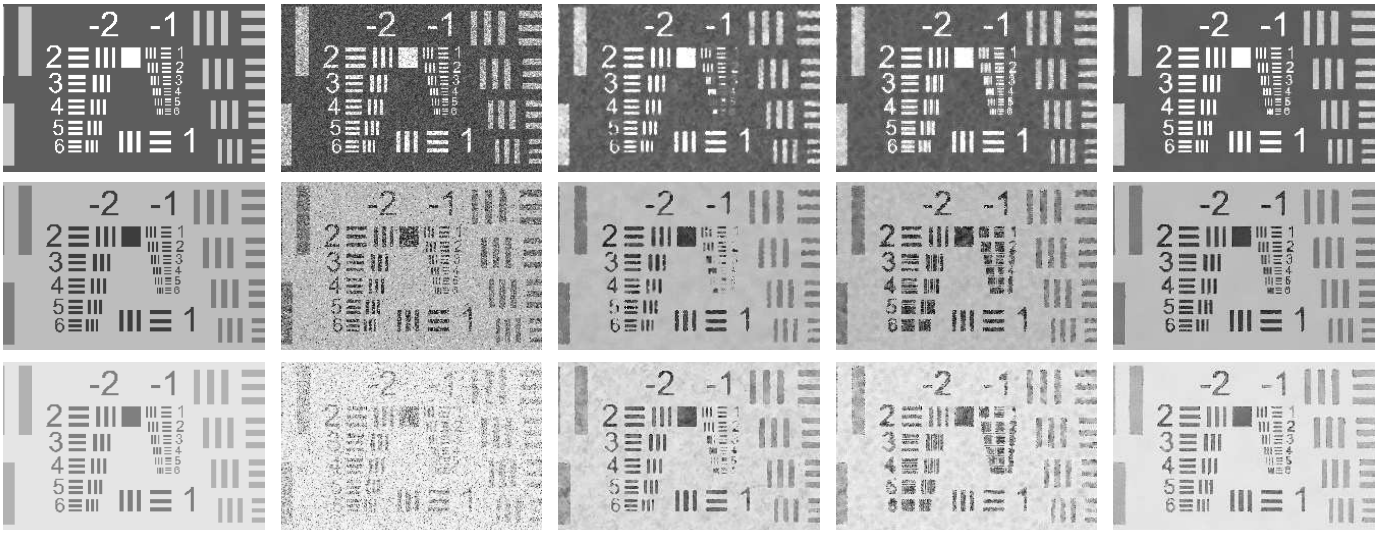
where $\div 4$ means that we keep one pixel over 4 in the search window during the spiral path in order to save time. Larger patches provide more robust patch comparisons while extended search areas are used to obtain satisfying noise reduction. Note that thanks to sub-sampling along the spiral path of the search window, the computation time is unchanged.

Fig. 10 illustrates the impact of noise correlation present on airborne radar images obtained by RAMSES sensor from ONERA. Raw images are shown on column (a). Decorrelated images were computed by decimation by a factor 2 in each direction. Results obtained by PPB-it and the proposed method on these decimated images are shown on columns (b) and (c). Both results are satisfying, although PPB-it has a tendency to over-smooth the image. When the algorithms are applied to the original correlated data, PPB-it produces an output with strong artifacts (column (d)), while the proposed method gives a satisfying image. Compared to the decimated image, homogeneous regions are equally smoothed.



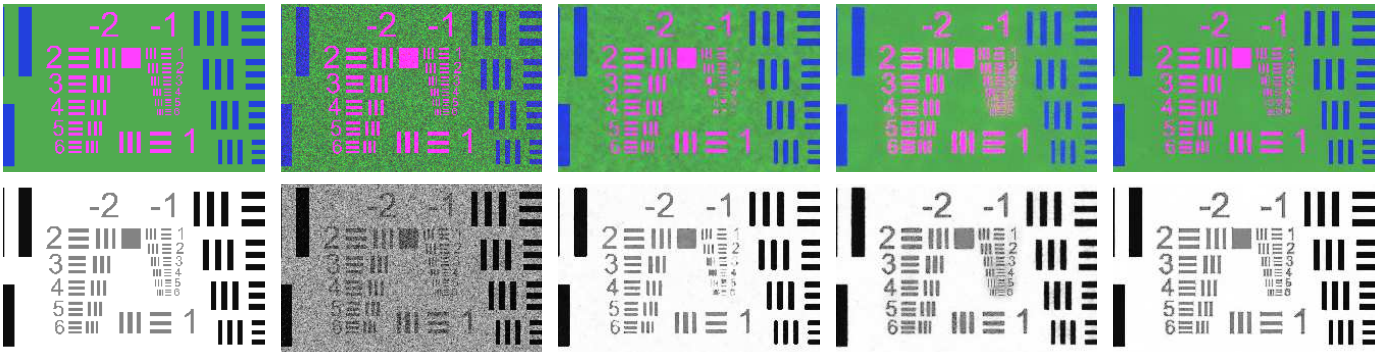
(a) (b) (c) (d) (e)

Fig. 7. Application to simulated amplitude images. (a) True image. (b) Noisy 1 look image. (c) 4 it. of PPB. (d) SAR-BM3D. (e) Our result.



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

Fig. 8. Application to simulated interferometric images. From top to bottom: amplitude, interferometric phase (range $[0, 2\pi]$) and coherence (range $[0, 1]$). (a) True image. (b) Noisy 2 looks image. (c) IDAN (d) Refined lee. (e) Our result.



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

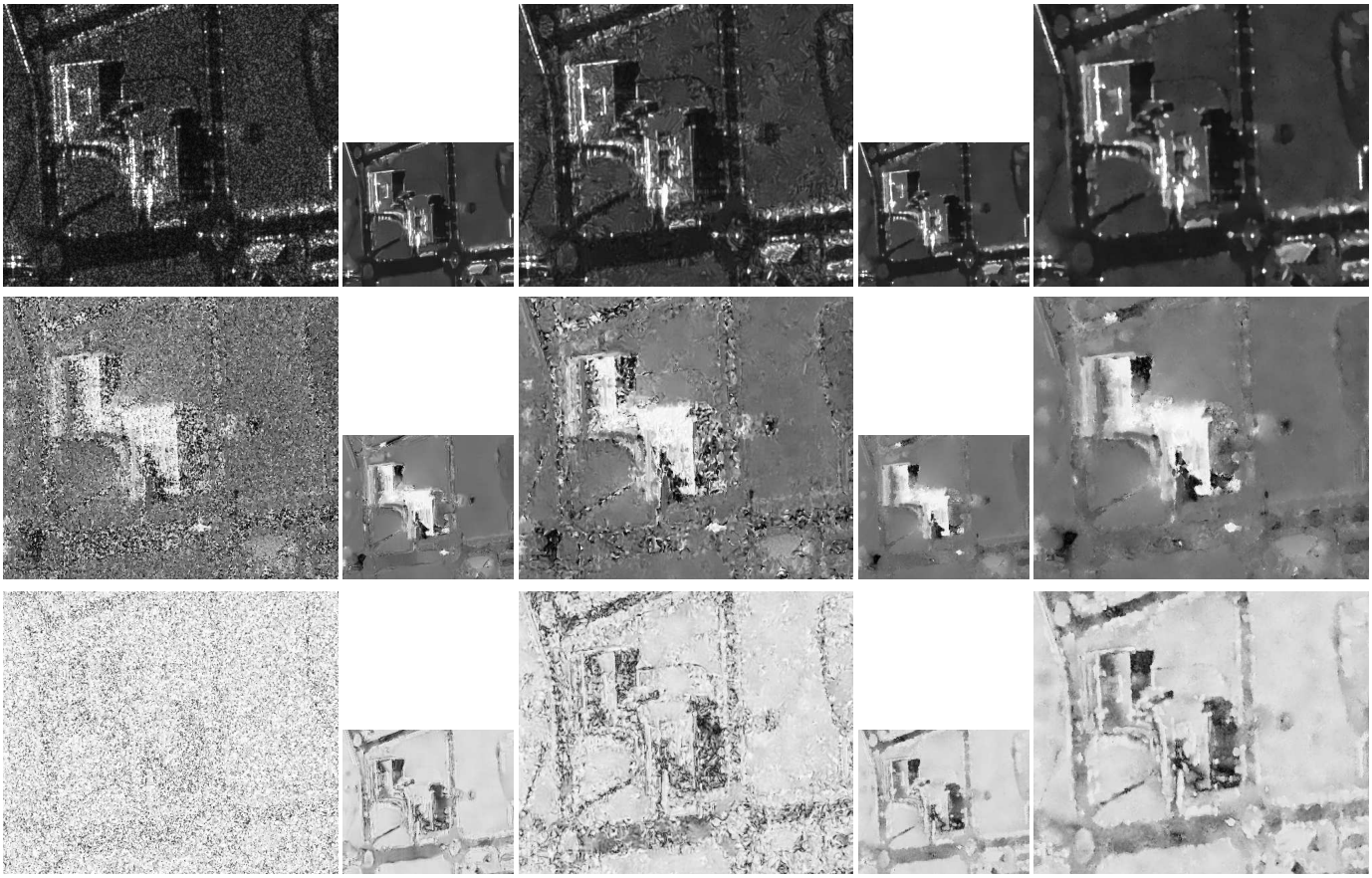
Fig. 9. Application to simulated polarimetric images. From top to bottom: RGB representation based on the Pauli basis (HH-VV,2HV,HH+VV) and Entropy (range $[0, 1]$). from top to bottom: amplitude, interferometric phase (range $[0, 2\pi]$) and coherence (range $[0, 1]$). (a) True image. (b) Noisy 3 looks image. (c) IDAN (d) Pretest. (e) Our result.

D. Restoration of interferometric and/or polarimetric images

We illustrate the performance of the proposed method on different SAR images with interferometric and/or polarimetric components. Many more examples are available in [63] and on the dedicated web page http://www.math.u-bordeaux1.fr/~cdeledal/nlsar.php where comparisons with other methods are also performed.

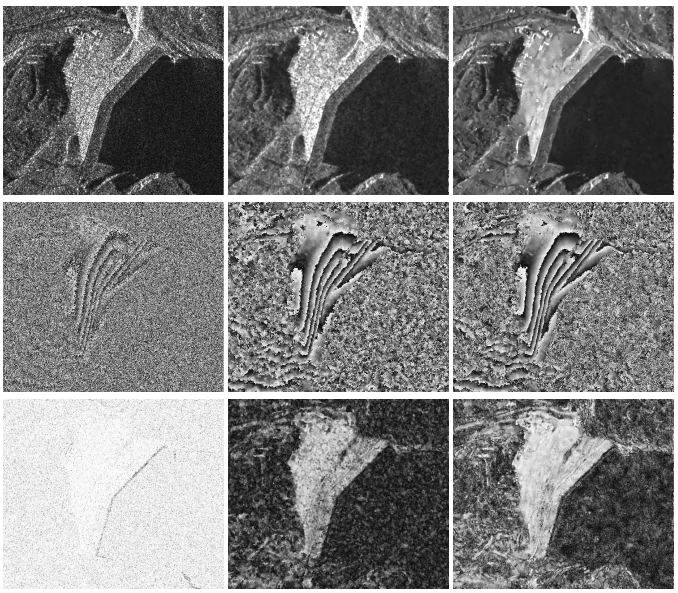
We first compare qualitatively NL-SAR with IDAN filter [22] for the estimation of the radiometry, interferometric phase and coherence from a pair of spaceborne images of a French Alpine dam sensed by TerraSAR-X. It is worth mentioning that the two images have been acquired with an interval of a few years. Hence, compared to Fig. 10, this image has a very poor coherence except on the dam which is extremely stable, and for which the phase is wrapped several time due to its high elevation. Fig. 11 depicts the restorations obtained with each method. By visual inspection, the restored radiometry, interferometric phase and coherence show smaller fluctuations and more details in NL-SAR output images.

We next consider a very high resolution airborne image captured in S-band by the German F-SAR system designed by the DLR. Fig. 12(a) shows an excerpt of a $0.5\text{m} \times 0.64\text{m}$ resolution polarimetric image. The image is color-coded using the Pauli basis. Strong fluctuations are noticed in all homo-



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

Fig. 10. (a) An interferometric image obtained from a pair of SAR images of an urban area in Toulouse (France) sensed by RAMSES ©ONERA with a mono-pass. From top to bottom: amplitude, interferometric phase (range $[0, 2\pi]$) and coherence (range $[0, 1]$). Speckle in this image is spatially correlated. Results obtained by NL-InSAR on decimated (b) and undecimated data (c). Results obtained with our approach on decimated (d) and undecimated data (e).



(a) (b) (c)

Fig. 11. (a) An interferometric image obtained from a pair of SAR images of the Serre-Ponçon dam (Alpes, France) (image courtesy of Airbus Defence and Space). From top to bottom: amplitude, interferometric phase (range $[0, 2\pi]$) and coherence (range $[0, 1]$). Results obtained by IDAN (b) and our approach (c).

geneous areas. The obtained restoration with the proposed method is shown in figure 12(b). Point-like features, linear structures and edges are well preserved without significant spreading while homogeneous areas are strongly smoothed. In contrast to denoising methods based on maximum a posteriori estimation with Markovian priors or collaborative filtering of blocks (BM3D), it is straightforward to produce a map of the equivalent number of looks of the restored image, see figure 12(c). As expected, while multi-looking averages samples in a non-adaptive way, the proposed non-local restoration is signal-adaptive and the equivalent number of looks is larger in homogeneous areas than close to geometrical structures. Low equivalent number of looks are observed as dark graylevels on figure 12(c). Point-like structures have the smallest equivalent number of looks, which is desirable since spatial averaging would spread those structures. Fig. 12(d,e,f) display the model based decomposition defined in [65], [66] where the polarimetric entropy H, the α parameter and the anisotropy A are computed from the polarimetric covariances estimated with our non-local restoration method. Areas with tall vegetation typically have the largest entropy with isotropic behavior. This is expected due to the high randomness of the polarimetric backscattering mechanisms in the forest canopy at S-band. Echoes created by point or ground-wall interaction appear mostly with very low entropy and stronger anisotropy as here

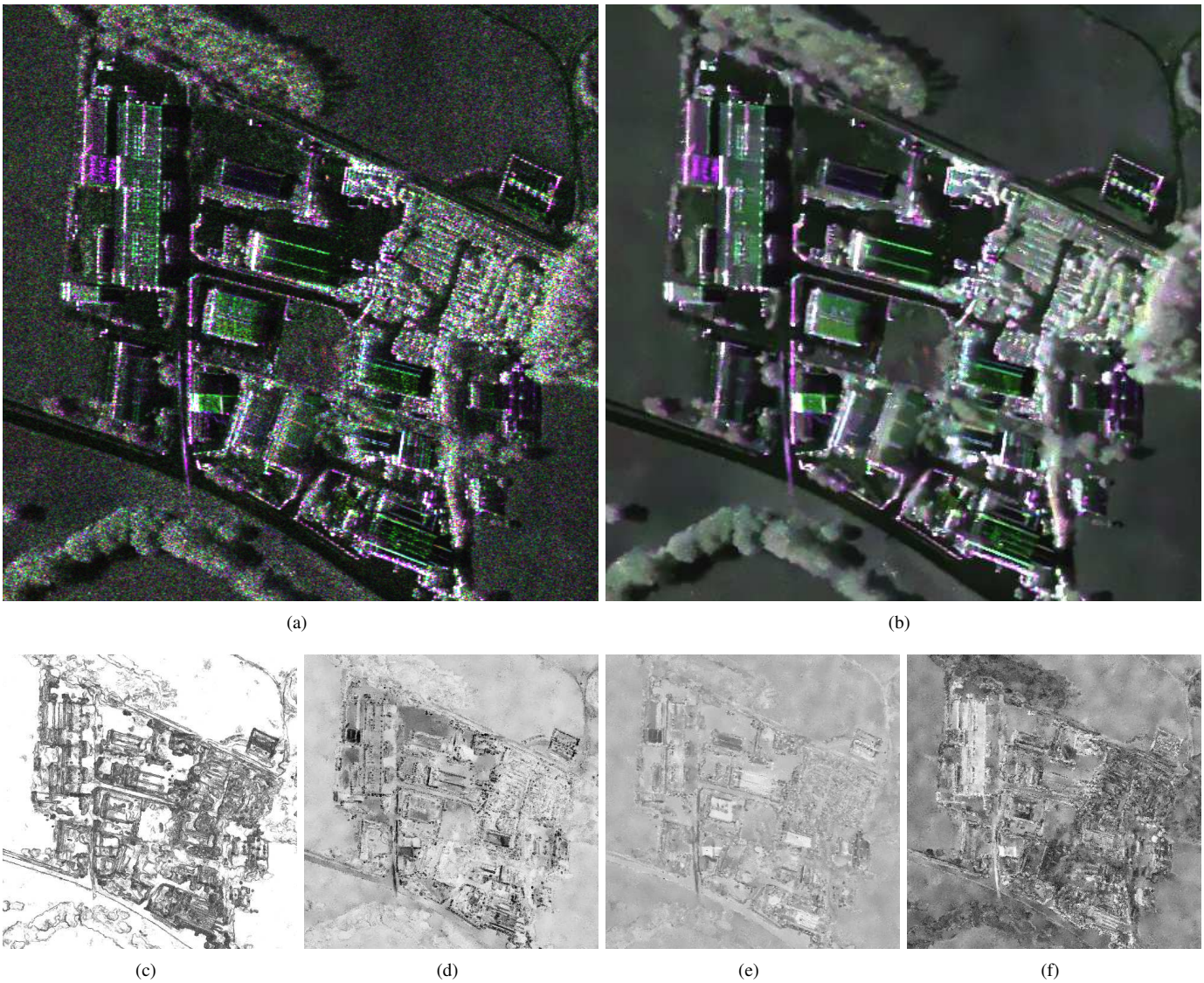


Fig. 12. (a) Polarimetric image of Kaufbeuren (Germany) sensed by S-band F-SAR ©DLR displayed using an RGB representation based on the Pauli basis. (b) Non-local denoising using our approach. (c) Map of the equivalent number of looks (values larger than 122 are displayed in white). Model-based decomposition [65], [66] computed from image (b): (d) entropy H (range $[0, 1]$), (e) α parameter (range $[0, \pi/2]$) and (f) anisotropy A (range $[0, 1]$).

only one scattering mechanism is dominating. The low values of entropy achieved on man-made structures give evidence that their energy is not spread out by the proposed non-local estimation technique.

We finally illustrate the performance of NL-SAR to estimate an airborne polarimetric and interferometric image of the Amazonian forest sensed by the L-band TropiSAR system. This region is of particular interest for the understanding of forest ecosystems and the study of the biomass. Fig. 13 displays the obtained results with NL-SAR. Polarimetric components as well as the interferometric phase and coherence on both vegetation areas and man-made structures are restored while preserving the sharpest structures.

VI. Conclusion

This paper introduced a novel framework for radar image denoising. This framework implements several concepts:

1) *similarity* between patches defined for arbitrary SAR modalities (SAR, InSAR, PolSAR, PolInSAR) and any number of looks;
2) *learned kernels* to weight the similarities according to their distribution in a homogeneous area;
3) *multiple estimation* with various sets of parameters;
4) *local selection* of the most reliable estimate, defined as the estimate with least variance after a bias-variance trade-off.

Numerical experiments have shown that each of these elements is necessary to reach good performance on a variety of geometrical structures and SAR modalities. Results obtained in interferometry and polarimetry significantly improve on the state-of-the-art.

We believe that the strength of our method comes from the wide applicability and generality of the proposed framework. The method is designed to be robust to changes of SAR modality, number of looks or noise correlation. We provide the source code together with this paper describing our methodol-

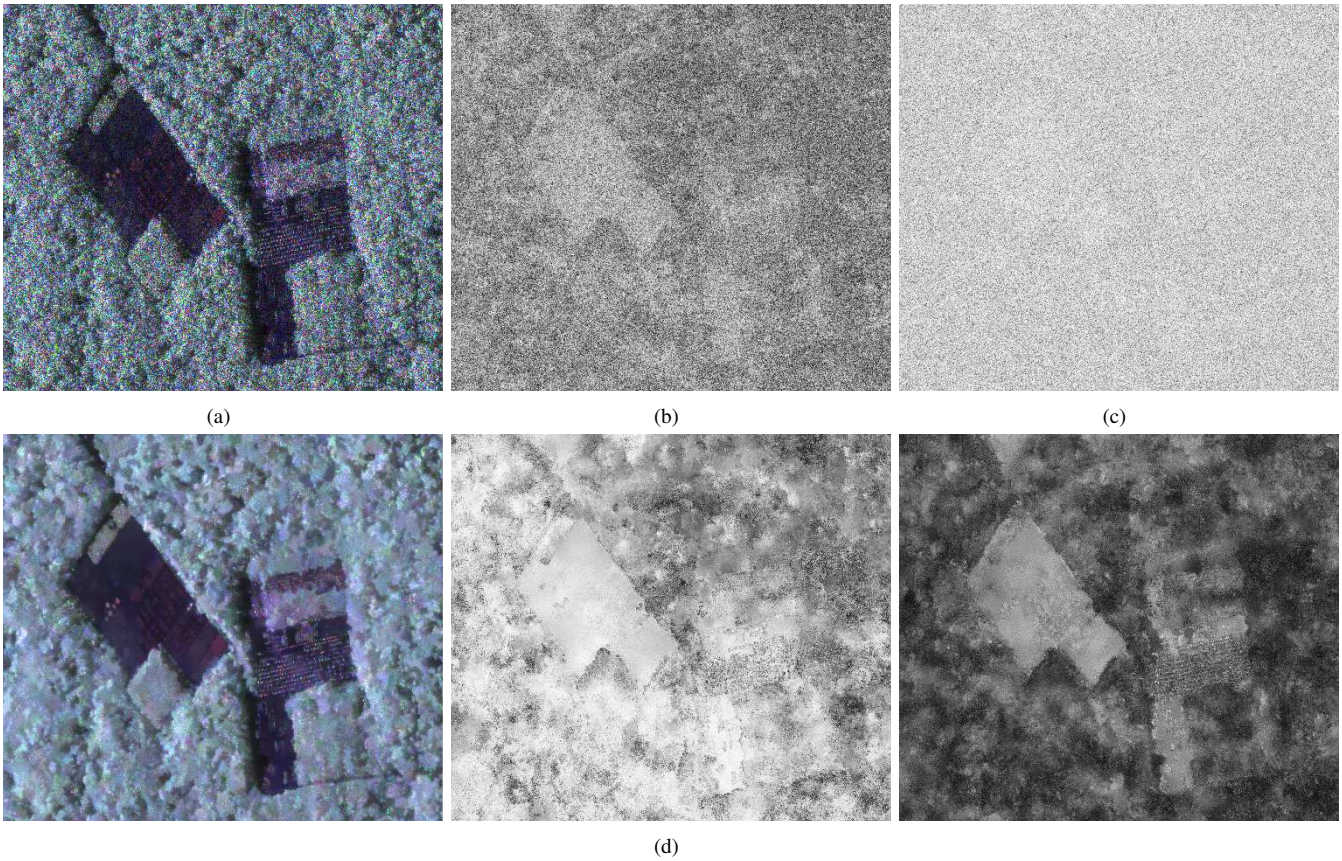


Fig. 13. Restoration of a polarimetric interferometric L-band TropiSAR image of Paracou (Guyane, France) ©ESA: (a,b,c) original polarimetric, phase and coherence images; (d) joint estimation with NL-SAR.

ogy. The code is reasonably fast and can be applied on large images using parallel architectures. A key feature that should ease the wide usage of the method is the fully automatic tuning of all parameters.

The very general framework that we propose leaves room to further improvements by combining more or different kinds of estimates, modifying the pre-estimation step, or using a better rule for estimator aggregation. Recent studies in additive Gaussian noise reduction have shown that similarities between overlapping patches should be weighted using a dedicated kernel [51]. This idea could be extended to our SAR denoising method by adapting the kernel ψ. Our method leaves isolated structures almost unchanged since no similar patches are found in the search area. Noise variance could be reduced in this case by either ensuring some regularity as done in [56] or by using a dictionary of SAR structures learned off-line or from the noisy data, in the spirit of dictionary-based denoising [67]. Noise in very high resolution images is known to depart from Goodman's model, due to the predominance of some scatterers inside each resolution cell. Thermal noise components are also no longer negligible. The similarity criterion and the weighted maximum likelihood estimator could then be adapted to more accurate noise models.

The techniques introduced to weigh similarities and to combine different estimates could provide useful building blocks to design methods for fusion of several radar images (acquired at different dates and/or by different sensors), or to detect motion and change in radar images.

Given the efforts devoted to producing very high resolution images, we believe that resolution-preserving estimation of radar properties (radiometry, interferometric phases, polarimetric covariance matrices) deserves special attention and will gain increasing interest in the years to come.

Acknowledgments

The authors would like to thank the anonymous reviewers as well as the associate editor for their comments, criticisms and encouragements. We also thank the Centre National d'Études Spatiales, the Office Nationale d'Études et de Recherches Aérospatiales and the Délégation Générale pour l'Armement for providing the RAMSES data, the Airbus Defence and Space for the TerraSAR-X data, the German Aerospace Center (DLR) for the F-SAR data and the European Space Agency for the TropiSAR data.

Experiments presented in this paper were carried out using the PlaFRIM experimental testbed, being developed under the INRIA PlaFRIM development action with support from LABRI and IMB and other entities: Conseil Régional d'Aquitaine, FeDER, Université de Bordeaux and CNRS (see https://plafrim.bordeaux.inria.fr/).

Appendix A
Derivation of variance reduction

Assuming first that the weights $w(x, x')$ are constant w.r.t. $I(x)$ and $I(x')$ for all (x, x') and next that they select

pixel values $I(x')$ i.i.d. with $I(x)$ gives

$$\mathrm{Var}\left[\hat{I}^{\mathrm{NL}}(x)\right] = \frac{\sum_{x'} w(x,x')^2}{\left(\sum_{x'} w(x,x')\right)^2}\mathrm{Var}\left[I(x)\right]$$

which proves the approximation (19). Next, we use that

$$\begin{aligned}\hat{I}^{\mathrm{NLRB}}(x) &= (1-\alpha)\hat{I}^{\mathrm{NL}}(x) + \alpha I(x)\\ &=(1-\alpha)\frac{\sum_{x'\neq x} w(x,x')I(x')}{\sum_{x'} w(x,x')} + \left(\alpha + \frac{(1-\alpha)}{\sum_{x'} w(x,x')}\right) I(x)\ .\end{aligned}$$

Assuming that α is constant w.r.t. $I(x)$ and $I(x')$ for all (x,x') prove eq. (20) since the inverse noise reduction ratio $\mathrm{Var}\left[\hat{I}^{\mathrm{NLRB}}(x)\right]/\mathrm{Var}\left[I(x)\right]$ is given by

$$\begin{aligned}&(1-\alpha)^2\frac{\sum_{x'\neq x} w(x,x')^2}{(\sum_{x'} w(x,x'))^2} + \left(\alpha + \frac{(1-\alpha)}{\sum_{x'} w(x,x')}\right)^2\\ =&\frac{(1-\alpha)^2}{\hat{L}^{\mathrm{NL}}(x)} + \alpha^2 + \frac{2\alpha(1-\alpha)}{\sum_{x'} w(x,x')}\end{aligned}$$

where we have used that $w(x,x) = 1$ and

$$\frac{\sum_{x'\neq x} w(x,x')^2}{(\sum_{x'} w(x,x'))^2} = \left(\frac{1}{\hat{L}^{\mathrm{NL}}(x)} - \frac{1}{(\sum_{x'} w(x,x'))^2}\right)\ .$$

REFERENCES

[1] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, no. 2, pp. 165–168, 1980.

[2] V. Frost, J. Stiles, K. Shanmugan, and J. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 4, pp. 157–166, 1982.

[3] D. Kuan, A. Sawchuk, T. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, no. 2, pp. 165–177, 1985.

[4] ——, "Adaptive restoration of images with speckle," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 35, no. 3, pp. 373–383, 1987.

[5] A. Lopes, E. Nezry, R. Touzi, and H. Laur, "Structure detection and statistical adaptive speckle filtering in SAR images," *International Journal of Remote Sensing*, vol. 14, no. 9, pp. 1735–1758, 1993.

[6] R. Touzi, "A review of speckle filtering in the context of estimation theory," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2392–2404, 2002.

[7] H. Xie, L. Pierce, and F. Ulaby, "SAR speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2196–2212, 2002.

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

[9] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1773–1784, 2003.

[10] M. Bhuiyan, M. Ahmad, and M. Swamy, "Spatially adaptive wavelet-based method using the Cauchy prior for denoising the SAR images," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 4, pp. 500–507, 2007.

[11] T. Bianchi, F. Argenti, and L. Alparone, "Segmentation-Based MAP Despeckling of SAR Images in the Undecimated Wavelet Domain," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 9, pp. 2728–2742, 2008.

[12] S. Durand, J. Fadili, and M. Nikolova, "Multiplicative noise removal using L1 fidelity on frame coefficients," *Journal of Mathematical Imaging and Vision*, vol. 36, no. 3, pp. 201–226, 2010.

[13] G. Aubert and J.-F. Aujol, "A variational approach to removing multiplicative noise," *SIAM Journal on Applied Mathematics*, vol. 68, no. 4, pp. 925–946, 2008.

[14] L. Denis, F. Tupin, J. Darbon, and M. Sigelle, "SAR image regularization with fast approximate discrete minimization," *IEEE Trans. on Image Processing*, vol. 18, no. 7, pp. 1588–1600, 2009.

[15] G. Steidl and T. Teuber, "Removing multiplicative noise by douglas-rachford splitting methods," *Journal of Mathematical Imaging and Vision*, vol. 36, no. 2, pp. 168–184, 2010.

[16] J. M. Bioucas-Dias and M. A. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization," *IEEE Trans. on Image Processing*, vol. 19, no. 7, pp. 1720–1730, 2010.

[17] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse problems*, vol. 19, no. 6, pp. 165–187, 2003.

[18] L. Denis, F. Tupin, J. Darbon, and M. Sigelle, "Joint regularization of phase and amplitude of InSAR data: Application to 3-D reconstruction," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3774–3785, 2009.

[19] J.-S. Lee, "Speckle analysis and smoothing of synthetic aperture radar images," *Computer Graphics and Image Processing*, vol. 17, no. 1, pp. 24–32, September 1981.

[20] J. Lee, M. Grunes, and G. De Grandi, "Polarimetric SAR speckle filtering and its implication for classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, no. 5 Part 2, pp. 2363–2373, 1999.

[21] J. Lee, S. Cloude, K. Papathanassiou, M. Grunes, and I. Woodhouse, "Speckle filtering and coherence estimation of polarimetric SAR interferometry data for forest applications," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 41, no. 10 Part 1, pp. 2254–2263, 2003.

[22] G. Vasile, E. Trouvé, J. Lee, and V. Buzuloiu, "Intensity-driven adaptive-neighborhood technique for polarimetric and interferometric SAR parameters estimation," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1609–1621, 2006.

[23] J. Lee, "Digital image smoothing and the sigma filter," *Computer Vision, Graphics, and Image Processing*, vol. 24, pp. 255–269, 1983.

[24] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Int. Conf. on Computer Vision (ICCV)*, 1998, pp. 839–846.

[25] O. D'Hondt, S. Guillaso, and O. Hellwich, "Iterative bilateral filtering of polarimetric SAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, 2013.

[26] J.-S. Lee, J.-H. Wen, T. L. Ainsworth, K.-S. Chen, and A. J. Chen, "Improved sigma filter for speckle filtering of SAR imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 202–213, 2009.

[27] A. Buades, B. Coll, and J. Morel, "A Review of Image Denoising Algorithms, with a New One," *Multiscale Modeling and Simulation*, vol. 4, no. 2, p. 490, 2005.

[28] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *International Journal of Computer Vision*, vol. 86, no. 1, pp. 1–32, 2010.

[29] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. on Image Processing*, vol. 16, no. 8, p. 2080, 2007.

[30] C.-A. Deledalle, L. Denis, G. Poggi, F. Tupin, and L. Verdoliva, "Exploiting patch similarity for SAR image processing: the nonlocal paradigm," *IEEE Signal Processing Magazine*, Jul. 2014.

[31] X. Yang and D. A. Clausi, "Structure-preserving speckle reduction of SAR images using nonlocal means filters," in *Int. Conf. on Image Processing (ICIP)*. IEEE, 2009, pp. 2985–2988.

[32] M. Mäkitalo, A. Foi, D. Fevralev, and V. Lukin, "Denoising of single-look SAR images based on variance stabilization and nonlocal filters," in *Proc. Int. Conf. Math. Meth. Electromagn. Th. (MMET)*, 2010.

[33] C. Kervrann, P. Pérez, and J. Boulanger, "Bayesian non-local means, image redundancy and adaptive estimation for image representation and applications," in *SIAM Conf. on Imaging Science*, San Diego, CA, July 2008.

[34] H. Zhong, Y. Li, and L. Jiao, "SAR image despeckling using bayesian nonlocal means filter with sigma preselection," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 809–813, 2011.

[35] C. Deledalle, L. Denis, and F. Tupin, "Iterative Weighted Maximum Likelihood Denoising with Probabilistic Patch-Based Weights," *IEEE Trans. on Image Processing*, vol. 18, no. 12, pp. 2661–2672, Dec. 2009.

[36] ——, "NL-InSAR: Non-Local Interferogram Estimation," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 49, no. 4, 2011.

[37] C. Deledalle, F. Tupin, and L. Denis, "Polarimetric SAR estimation based on non-local means," *Int. Geosc. and Remote Sensing Symp. (IGARSS)*, 2010.

[38] J. Chen, Y. Chen, W. An, Y. Cui, and J. Yang, "Nonlocal Filtering for Polarimetric SAR Data: A Pretest Approach," *IEEE Trans. on Geoscience and Remote Sensing*, no. 99, pp. 1–11, 2011.

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

[40] T. Teuber and A. Lang, "A new similarity measure for nonlocal filtering in the presence of multiplicative noise," *Computational Statistics & Data Analysis*, 2012.

[41] C.-A. Deledalle, L. Denis, and F. Tupin, "How to compare noisy patches? patch similarity beyond gaussian noise," *International Journal of Computer Vision*, vol. 99, no. 1, pp. 86–102, 2012.

[42] L. Torres, S. J. Sant'Anna, C. da Costa Freitas, and A. C. Frery, "Speckle reduction in polarimetric SAR imagery with stochastic distances and nonlocal means," *Pattern Recognition*, 2013.

[43] C.-A. Deledalle, "Image denoising beyond additive gaussian noise - patch-based estimators and their application to SAR imagery," Ph.D. dissertation, Telecom ParisTech, 2011.

[44] N. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Annals of Mathematical Statistics*, pp. 152–177, 1963.

[45] S. Kay, *Fundamentals of statistical signal processing Volume 2: Detection theory*. Prentice-Hall, 1998.

[46] K. Conradsen, A. A. Nielsen, J. Schou, and H. Skriver, "A test statistic in the complex Wishart distribution and its application to change detection in polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 4–19, 2003.

[47] C. Kervrann and J. Boulanger, "Local Adaptivity to Variable Smoothness for Exemplar-Based Image Regularization and Representation," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 45–69, 2008.

[48] A. Buades, B. Coll, and J. Morel, "A Non-Local Algorithm for Image Denoising," in *Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2005, pp. 60–65.

[49] J. Salmon, "On two parameters for denoising with non-local means," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 269–272, 2010.

[50] A. Buades, B. Coll, and J.-M. Morel, "Non-Local Means Denoising," *Image Processing On Line*, vol. 2011, 2011.

[51] Y. Wu, B. Tracey, P. Natarajan, and J. P. Noonan, "Probabilistic non-local means," *IEEE Signal Processing Letters*, vol. 20, 2013.

[52] A. Buades, B. Coll, and J. Morel, "Nonlocal Image and Movie Denoising," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 123–139, 2008.

[53] J. Polzehl and V. Spokoiny, "Propagation-separation approach for local likelihood estimation," *Probability Theory and Related Fields*, vol. 135, no. 3, pp. 335–362, 2006.

[54] C. Louchet and L. Moisan, "Total variation as a local filter," *SIAM Journal on Imaging Sciences,*, vol. 4, 2011.

[55] C.-A. Deledalle, V. Duval, and J. Salmon, "Non-local Methods with Shape-Adaptive Patches (NLM-SAP)," *J. Mathematical Imaging and Vision*, pp. 1–18, 2011.

[56] C. Sutour, C.-A. Deledalle, and J.-F. Aujol, "Adaptive regularization of the NL-means: Application to image and video denoising," *IEEE Trans. on Image Processing (to appear)*, 2014.

[57] F. Bunea, A. B. Tsybakov, and M. H. Wegkamp, "Aggregation for gaussian regression," *The Annals of Statistics*, vol. 35, no. 4, pp. 1674–1697, 2007.

[58] D. Van De Ville and M. Kocher, "SURE-Based Non-Local Means," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 973–976, November 2009.

[59] ——, "Nonlocal means with dimensionality reduction and SURE-based parameter selection," *IEEE Trans. on Image Processing*, vol. 20, no. 9, pp. 2683–2690, 2011.

[60] V. Duval, J.-F. Aujol, and Y. Gousseau, "A bias-variance approach for the nonlocal means," *SIAM Journal on Imaging Sciences*, vol. 4, no. 2, pp. 760–788, 2011.

[61] J. Salmon and Y. Strozecki, "From patches to pixels in non-local methods: Weighted-average reprojection," in *Int. Conf. on Image Processing (ICIP)*. IEEE, 2010, pp. 1929–1932.

[62] J. Darbon, A. Cunha, T. Chan, S. Osher, and G. Jensen, "Fast nonlocal filtering applied to electron cryomicroscopy," in *Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI)*. IEEE, 2008, pp. 1331–1334.

[63] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, "Additional illustrations of NL-SAR method for resolution-preserving (Pol)(In)SAR denoising," Mar. 2014. [Online]. Available: http://hal.archives-ouvertes.fr/hal-00955194

[64] G. Di Martino, M. Poderico, G. Poggi, D. Riccio, and L. Verdoliva, "Benchmarking framework for SAR despeckling," *to appear in IEEE Trans. on Geoscience and Remote Sensing*, 2013.

[65] S. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 68–78, 1997.

[66] L. Ferro-Famil, E. Pottier, and J.-S. Lee, "Unsupervised classification of multifrequency and fully polarimetric sar images based on the h/a/alpha-wishart classifier," *Geoscience and Remote Sensing, IEEE Trans. on*, vol. 39, no. 11, pp. 2332–2342, 2001.

[67] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, p. 4311, 2006.