Signal, Image and Video Processing manuscript

Statistical Estimation of the Structural Similarity Index for Image Quality Assessment

Felipe Osorio · Ronny Vallejos · Wilson Barraza · Silvia María Ojeda · Marcos Alejandro Landi

Received: 11 February 2021 / Accepted: 1 October 2021

Abstract The structural similarity (SSIM) index has been studied from different perspectives in the last decade. Most of the developments consider its parameters fixed. Because each of these parameters corresponds to the weight of a factor in the final SSIM coefficient, the usual assumption that all parameters are equal to one is questionable. In this article, a new estimation method is proposed from a statistical perspective. The approach we develop is a model-based estimation method so that, the usual assumption that all parameters are equal to one can be handled via approximate hypothesis-testing techniques that are properly developed in the context of regression. The method considers nonlinear models with multiplicative noise to explain the root mean square error (RMSE) as a function of the SSIM index. A numerical experiment based on a Monte Carlo simulation is carried out to test whether the parameters are all equal to one and to gain more insight into the performance of the estimates in practice. Our analysis showed that the assumption that the parameters are equal to one is not supported by the data and may lead to a misconception of the closeness between two images.

Keywords Hypothesis testing · Nonlinear models · Pseudo-likelihood · Structural similarity index.

This preprint has not undergone improvements or corrections. The Version of record of this article is published in *Signal, Image and Video Processing*, and is available online at https://doi.org/10.1007/s11760-021-02051-9.

F. Osorio and R. Vallejos
Departamento de Matemática, Universidad Técnica Federico Santa María. Avenida España 1680, Valparaíso, Chile.
E-mail: felipe.osorios@usm.cl, ronny.vallejos@usm.cl

W. Barraza
U-Planner, Avenida Tupungato 3850, Curauma, Valparaíso, Chile.
E-mail: wilson.barraza@u-planner.com

S.M. Ojeda
Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba. Medina Allende s/n. Ciudad Universitaria (X5000HUA), Córdoba, Argentina.
E-mail: sm.ojeda@unc.edu.ar

M.A. Landi
Instituto de Diversidad y Ecología Animal, CONICET y Facultad de Ciencias Exactas Físicas y Naturales, Universidad Nacional de Córdoba. Córdoba 5016, Argentina.
E-mail: marcoslandi1980@gmail.com

1 Introduction and Preliminaries

Image quality assessment (IQA) aims to quantitatively represent the human perception of quality. With the rapid proliferation of digital imaging and communication technologies, IQA has been becoming an important issue in numerous applications, such as image acquisition, transmission, compression, restoration, and enhancement. These indices are commonly designed to study the performance of algorithms for several different problems frequently address in image processing such as image compression, image restoration, among others. IQA is mainly divided into two categories: no reference IQA that refers to automatic quality assessment of an image using an algorithm such that the only input information the algorithm receives is the distorted image whose quality is being assessed [1]. Alternatively, full reference algorithms require not only the distorted image, but also the reference image. In this side of the spectrum, several measures have been suggested in the literature [2]. One of these coefficients is the SSIM index, which was first introduced by [3]. In this paper we consider the SSIM for several reasons. This coefficient is naturally an attractive index because it represents human vision better than the well-known mean

square error (MSE). The conventional metrics, such as the peak signal-to-noise ratio (PSNR) and the MSE operate directly on the intensity of the image, and they do not correlate well with the subjective fidelity ratings [4]. Moreover, studies conducted in [5] and [6] developed mathematical properties of the SSIM which highlight the coefficient from an optimization perspective. Because of its simplicity, the computational cost of the SSIM index is very low, and it involves simple factorization of three terms. The first two terms are functions of the sample means and variances, and the third term corresponds to the sample correlation coefficient between the images, i.e., the structural part of the index compares the means and variances and computes the linear association between two images. Precisely, we let $\mathbb{R}_+$ denote the nonnegative real line, and we let $\mathbb{R}_+^N$ denote the first orthant, i.e., the set of N-dimensional vectors with nonnegative components. An image is considered an element $\boldsymbol{x} \in \mathbb{R}_+^N$. If $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_+^N$ are two images, the SSIM index [7] is

$$\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) = l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma}, \tag{1}$$

where

$$l(\boldsymbol{x}, \boldsymbol{y}) = \frac{2\,\overline{x} \cdot \overline{y} + c_1}{\overline{x}^2 + \overline{y}^2 + c_1}, \qquad c(\boldsymbol{x}, \boldsymbol{y}) = \frac{2\,s_x \cdot s_y + c_2}{s_x^2 + s_y^2 + c_2},$$

$$s(\boldsymbol{x}, \boldsymbol{y}) = \frac{s_{xy} + c_3}{s_x \cdot s_y + c_3},$$

$\overline{x}, \overline{y}, s_x^2, s_y^2$, and s_{xy} represent the sample means of $\boldsymbol{x}$ and $\boldsymbol{y}$, the sample variances of $\boldsymbol{x}$ and $\boldsymbol{y}$, and the sample covariance between $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. The constants c_1, c_2 and c_3 are all positive. We emphasize that $l(\boldsymbol{x}, \boldsymbol{y})$, $c(\boldsymbol{x}, \boldsymbol{y})$ and $s(\boldsymbol{x}, \boldsymbol{y})$ represent the luminance, contrast and structure, respectively, between the images $\boldsymbol{x}$ and $\boldsymbol{y}$, and $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top$ is a parameter vector in which each component corresponds to a weight associated with the corresponding component of (1), and thus $\alpha > 0, \beta > 0, \gamma > 0$. The constants c_1, c_2 and c_3 characterize the saturation effects of the visual system to guarantee stability when the denominators are close to zero. In practice, the values of these constants are very small (see [3] for details).

To compute the SSIM index, there are currently routines available in several packages. For instance, a Matlab code is available from MathWorks[1], and a C++ code can be found at github[2]. Since the introduction of the SSIM index, several extensions have been published and discussed. They include the information fidelity index [8], the visual signal-to-noise ratio [9], a perceptual quality assessment for multi-exposure image fusion [10], and the regression SSIM measure [11]. A generalization of the SSIM index was studied by [12], which considers the codispersion coefficient instead the linear correlation between the two images making the coefficient to depend on a spatial lag, in a similar way as the variogram is described in spatial statistics. [13] introduced a new image similarity measure based on hypothesis testing to assess structural information change by evaluating the dependence of local blocks of the error signal on the images being compared.

In Equation (1), the parameters α, β, and γ are, in general, unknown. For simplicity, in several reports in the literature, these parameters are assumed to equal 1 (see [5] and [6]) regardless of the texture of the underlying images. The estimation of these parameters is addressed in a limited number of papers (see for instance, [14] and [15]). In this article, we look at the estimation of α, β, and γ from a statistical perspective, i.e., using a model-based approach, and we develop a hypothesis-testing procedure to test whether the null hypothesis

$$\begin{aligned} &H_0 : \alpha = \beta = \gamma = 1, \quad \text{against} \\ &H_1 : \alpha \neq 1, \text{or } \beta \neq 1, \text{or } \gamma \neq 1, \end{aligned} \tag{2}$$

should be rejected for a given significance value. That is, the alternative hypothesis H_1 represents the case in which at least one coefficient is different from one. In our approach, a regression model is considered: a nonlinear model with multiplicative error. The model is built to express the RMSE as a function of the SSIM index. Evidence regarding the relationship between MSE and SSIM has been provided by [16,17] and [18]. When the noise is multiplicative, the model becomes a heteroscedastic nonlinear model, and the estimation of the parameters is addressed by pseudo-likelihood maximization in the same fashion as in [19]. To accomplish this, an iterative estimation algorithm is explained in detail to obtain more insight into the estimation process for the suggested model.

Based on the proposed method for estimating the parameters of the SSIM index, we present a simulation experiment using three reference images with different textures, which are contaminated with multiplicative gamma noise (see [20]). We explore the performance of the estimates in the existing literature and conclude through the hypothesis-testing techniques that, in general, the assumption that $\alpha = \beta = \gamma = 1$ is not suitable in the context of remote-sensing data. In practice, we provide estimates based on the data (images) to determine the weights of the luminance, contrast, and correlation for the final SSIM coefficient.

The remainder of this paper is organized as follows. Section 2 considers that the parameters α, β and γ are

[1] URL: https://www.mathworks.com/matlabcentral/answers/9217-need-ssim-m-code

[2] URL: https://gist.github.com/Bibimaw/8873663

not fixed, and details a regression model approach to estimating the SSIM index and the estimation algorithm in the framework of nonlinear models with multiplicative noise. Then, in Section 3, we present a simulation study using three reference images to demonstrate that the assumption that all parameters equal 1 is not always appropriate and to report that the proposed methodology provides reasonable estimates. Finally, Section 4 summarizes this paper.

2 Methodology

This section establishes a nonlinear regression model to relate the SSIM and RMSE coefficients, and develops an estimation algorithm to implement in practice the hypothesis testing problem described in (2). We also approach the derivation of the score function associated with the statistical test, and the strategy to build routines in C, material that has been relegated to Appendix B in the supplementary material.

Throughout this article, we assume that the response variable is

$$Z = 1/\operatorname{RMSE}(\boldsymbol{x}, \boldsymbol{y}).$$

The estimation of functions such as the one described in Equation (1) is common in economic contexts, where they are called production functions [21]. In this type of model, it is possible to incorporate random noise in an additive or a multiplicative way as follows:

$$Z = l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma} + u, \tag{3}$$

$$Z = l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma} \cdot e^{u}, \tag{4}$$

where u is a normal random variable with distribution $\mathsf{N}(0, \sigma^2)$. For the additive model described in (3), the estimation of the parameters can be carried out using the well-known nonlinear least-squares estimator [22]. Because this work is dedicated to remote sensing images for which the assumption of a multiplicative noise is frequent [20] (see also [23]), the rest of the methodology is described for model (4).

We note that the mean and variance functions for model (4) are, respectively, given by

$$\mathrm{E}(Z) = l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma} e^{\sigma^2/2}$$

$$\operatorname{var}(Z) = [l(\boldsymbol{x}, \boldsymbol{y})^{\alpha} \cdot c(\boldsymbol{x}, \boldsymbol{y})^{\beta} \cdot s(\boldsymbol{x}, \boldsymbol{y})^{\gamma} \cdot e^{\sigma^2/2}]^2 (e^{\sigma^2} - 1),$$

which correspond to a heteroscedastic nonlinear regression model such as the one treated by [19] (see also [24]). To clarify some ideas, suppose that the images $\boldsymbol{x}, \boldsymbol{y}$ are divided into n subimages of size $k \times k$. That is, there are n available observations, and we define

$$f_i(\boldsymbol{\theta}) = \operatorname{SSIM}(\boldsymbol{x}_i, \boldsymbol{y}_i; \boldsymbol{\theta}), \quad \text{for } i = 1, \ldots, n.$$

Then, we estimate $\boldsymbol{\psi} = (\boldsymbol{\theta}^\top, \phi)^\top$ using the model

$$Z_i \sim \mathsf{N}(\phi f_i(\boldsymbol{\theta}), f_i^2(\boldsymbol{\theta}) g^2(\phi)), \quad i = 1, \ldots, n, \tag{5}$$

where $g^2(\phi) = \phi^2(\phi^2 - 1)$ with $\phi = e^{\sigma^2/2}$.

To facilitate the estimation of the scale parameter ϕ, the pseudo-likelihood associated with the model defined in (5) can be used based on an initial estimate for $\boldsymbol{\theta}$. Thus, an estimator based on the pseudo-likelihood $\widehat{\phi}_{\mathsf{PL}}$ maximizes the function

$$\ell_{\mathsf{PL}}(\phi) = -\frac{1}{2} \sum_{i=1}^{n} \left[\frac{r_i^2(\widehat{\boldsymbol{\theta}}_*, \phi)}{f_i^2(\widehat{\boldsymbol{\theta}}_*) g^2(\phi)} + \log f_i^2(\widehat{\boldsymbol{\theta}}_*) g^2(\phi) \right], \tag{6}$$

where $r_i(\boldsymbol{\theta}, \phi) = Z_i - \phi f_i(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\theta}}_*$ is the current estimate of the regression parameter $\boldsymbol{\theta}$. The function $\ell_{\mathsf{PL}}(\phi)$ can be maximized by a one-dimensional optimization procedure such as the method described in [25]. This method needs to be alternated with a substage in which $\widehat{\boldsymbol{\theta}}$ is updated as the solution of an unconstrained minimization problem. Algorithm 1 describes the stages necessary to obtain $\widehat{\boldsymbol{\psi}} = (\widehat{\boldsymbol{\theta}}^\top, \widehat{\phi})^\top$. To gain more insight into the convergence depicted in Step 6, we recommend following the suggestions given in Chapter 14 of [22].

Algorithm 1 Estimation of the SSIM index using pseudo-likelihood

Input: Images $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_+^N$, tolerance ($\tau$), maximum number of iterations ($\mathsf{maxiter}$), window size k.

Output: Estimated parameters: $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})^\top$ and $\widehat{\phi}$.

 # Preprocessing stage

1: Subdivide images $\boldsymbol{x}, \boldsymbol{y}$ into n windows of size $k \times k$, obtaining the dataset $(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_n, \boldsymbol{y}_n)$.

2: Compute $l(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $c(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $s(\boldsymbol{x}_i, \boldsymbol{y}_i)$ and the response $Z_i = 1/\operatorname{RMSE}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ for $i = 1, \ldots, n$.

 # Estimation stage

3: Using a preliminary estimator $\widehat{\boldsymbol{\theta}}_*$, compute $f_i(\widehat{\boldsymbol{\theta}}_*)$ and $r_i(\widehat{\boldsymbol{\theta}}_*, \phi)$, for $i = 1, \ldots, n$.

4: Update $\widehat{\phi}_{\mathsf{PL}}$ by maximizing $\ell_{\mathsf{PL}}(\phi)$ in relation to ϕ by using a one-dimensional optimization procedure.

5: Update $\widehat{\boldsymbol{\theta}}$ using the following iterative procedure:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \lambda_t \boldsymbol{\delta}_t, \qquad t = 0, 1, \ldots, \tag{7}$$

where λ_t denotes the step length and $\boldsymbol{\delta}_t$ is the search direction obtained using a quasi-Newton method [26]. Equation (9) provides the gradient information that is required for computing the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update (for $\phi = \widehat{\phi}_{\mathsf{PL}}$ held fixed). Upon completion of the inner cycle defined in (7), we set $\widehat{\boldsymbol{\theta}}$.

6: Check the convergence of the algorithm. If it has been reached, let $\widehat{\boldsymbol{\psi}} = (\widehat{\boldsymbol{\theta}}^\top, \widehat{\phi})^\top$ be the parameter estimate with $\widehat{\phi} = \widehat{\phi}_{\mathsf{PL}}$. Otherwise, set $\widehat{\boldsymbol{\theta}}_* = \widehat{\boldsymbol{\theta}}$, and return to step 3.

The asymptotic distribution of the estimator obtained from the pseudo-likelihood method under very

general conditions with regard to the moments has been derived in [19]. Moreover, the asymptotic covariance matrix in this framework does not have a simple expression. However, based on the distributional assumption given in (5) and assuming suitable regularity conditions (see, for instance [27]), it follows that $\sqrt{n}(\widehat{\boldsymbol{\psi}}-\boldsymbol{\psi})$ has an asymptotic normal distribution with mean zero and covariance matrix $\boldsymbol{\mathcal{F}}_n^{-1}(\boldsymbol{\psi})$. The Fisher information matrix for the model given in (5) is deferred to the supplementary material. In this study, we avoid calculating the second-order information terms by considering an asymptotically equivalent test statistic proposed by [28]. Noting that the null hypothesis of interest $H_0 : \alpha = \beta = \gamma = 1$ can be written in a more compact way as $H_0 : \boldsymbol{\theta} = \mathbf{1}$, it follows that the gradient statistic is

$$T = \boldsymbol{U}^\top(\mathbf{1})(\widehat{\boldsymbol{\theta}} - \mathbf{1}), \tag{8}$$

where the score function $\boldsymbol{U}(\boldsymbol{\theta}) = \partial\ell(\boldsymbol{\psi})/\partial\boldsymbol{\theta}$ based on model (5), is given by:

$$\begin{aligned}\boldsymbol{U}(\boldsymbol{\theta}) = {} & \frac{\phi}{g^2(\phi)}\boldsymbol{F}^\top(\boldsymbol{\theta})\boldsymbol{W}^{-1}(\boldsymbol{\theta})(\boldsymbol{Z} - \phi\boldsymbol{f}(\boldsymbol{\theta})) \\ & - \frac{1}{2}\boldsymbol{H}^\top(\boldsymbol{\theta})\boldsymbol{V}^{-1}(\boldsymbol{\theta})\,\mathrm{vec}\left(\boldsymbol{W}(\boldsymbol{\theta}) - \frac{\boldsymbol{r}(\boldsymbol{\theta})\boldsymbol{r}^\top(\boldsymbol{\theta})}{g^2(\phi)}\right),\end{aligned} \tag{9}$$

with $\boldsymbol{H}(\boldsymbol{\theta}) = (\mathrm{vec}(\partial\boldsymbol{W}(\boldsymbol{\theta})/\partial\theta_1), \ldots, \mathrm{vec}(\partial\boldsymbol{W}(\boldsymbol{\theta})/\partial\theta_p))$, p is the the dimension of $\boldsymbol{\theta}$, $\mathrm{vec}(\cdot)$ denotes the vectorization operator, $\boldsymbol{V}(\boldsymbol{\theta}) = \boldsymbol{W}(\boldsymbol{\theta}) \otimes \boldsymbol{W}(\boldsymbol{\theta})$ and $\otimes$ indicates the Kronecker product. For the gradient statistic, $\boldsymbol{U}(\mathbf{1}) = \boldsymbol{U}(\boldsymbol{\theta})\big|_{\theta=1}$ corresponds to the score function given in (9) evaluated under H_0. Finally, the decision rule in the gradient test establishes that the null hypothesis is rejected if $T > \chi^2_{1-\alpha}(3)$, where $\chi^2_{1-\alpha}(3)$ is the lower quantile of order $100(1-\alpha)\%$ of the chi-square distribution with 3 degrees of freedom.

The estimation procedure and the hypothesis-testing method described in this section have been implemented in R and C, and the codes are available on github.[3] The computational strategy adopted in our C routines to evaluate the score function for the heteroscedastic regression model defined in (5) is described in the supplementary material.

3 Numerical Experiments

In this section, we evaluate the performance of the proposed methodology through a simulation study and the analysis of synthetic aperture radar (SAR) imagery provided publicly by ICEYE.[4] Statistical estimation and hypothesis testing about the parameters α, β, and γ will bring worth information leading us to consider non constant values for these parameters.

3.1 Monte Carlo simulation study

3.1.1 Simulation setup

The images texmos2.S512, Lena, and Baboon used in this study (see Figure 1) belong to the USC-SIPI image database.[5] All images used in this work have been normalized and converted to a grayscale using the R package `SpatialPack` [29]. The advantages of these images is that they present a variety of different textures. We have contaminated each reference image $\boldsymbol{x}$ as

$$\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{w},$$

where $\boldsymbol{w}$ denotes a multilook intensity speckle noise. Motivated by [20], we will assume that each element of w follows a $\mathsf{Gamma}(L, L)$ distribution, with L also known as the equivalent (or effective) number of looks, which is related to the capture and processing of synthetic aperture radar (SAR) images. The multilook intends to mitigate the effect of interference due to the speckle noise. The higher the number of looks the lower the level of interference or contamination. In our experiment each reference images (Figure 1) were contaminated with multiplicative noise using a Gamma distribution, with different decreasing levels of contamination, that is, 1, 2, 4, 8, 16, and 32 looks. For each look, 1,000 images were constructed. Moreover, Lee, enhanced Lee, and Kuan filters were used to remove the multiplicative noise for each of the distorted images. A brief description of these filters is presented in Appendix D of the supplementary material. For all filters we have considered pixel window of size 3×3. Next, the parameter estimates of model (4) were obtained using the Algorithm 1, also we computed the gradient statistic to test the hypothesis $H_0 : \alpha = \beta = \gamma = 1$ and evaluated the SSIM under H_0 and H_1.

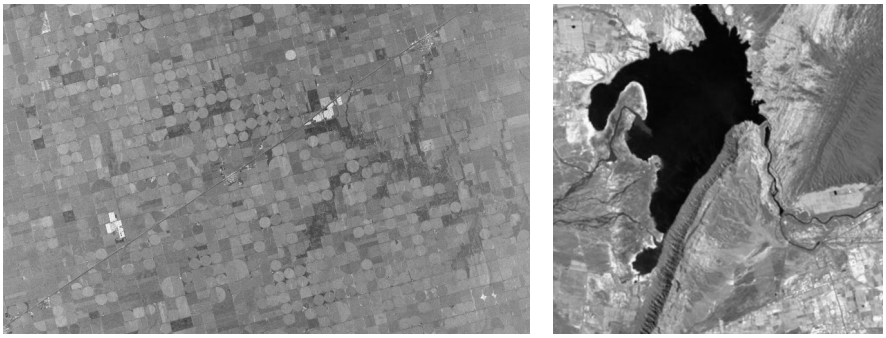
3.1.2 Simulation results

Table 1 shows the average of the estimates of $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top$ for the multiplicative model described in Equation (4). Results for Lena and Baboon reference images are deferred to the supplementary material for this article. The results are based on 1,000 simulation runs and $\boldsymbol{\theta}^{(0)} = (1, 1, 1)^\top$ was used as the initial estimate for SSIM estimation considering each distorted as well as each filtered image. The estimates allow us to question the assumption that $\alpha = \beta = \gamma = 1$. The results

[3] URL: https://github.com/faosorios/SSIM
[4] URL: https://www.iceye.com/downloads/datasets
[5] URL: http://sipi.usc.edu/database

(a) texmos2.S512 (b) Lena (c) Baboon

Fig. 1 Reference images from the USC-SIPI image database.



(a) Copeland (b) Dam

(c) Corpus Christi (d) Mississippi

Fig. 2 Images (a), (c), and (d) were obtained from the ICEYE SAR satellite constellation. Image (b) was taken from the school of mines database at Universidad Nacional de San Juan, Argentina.

depend on the type of image considered in the experiments, as well as on the filter used. In particular, for the texmos2.S512 the estimations of the parameters are equal or closed to 1 only when when no filter is used, and the number of looks is equal to 1 or 2. It is also observed that the performance of the filters is very similar, there being a high concordance between Kuan and Enhanced Lee filters. When the number of looks is greater than 2, the parameter estimates start varying from 1. The results looks slightly different for textured images, such as Lena and Baboon. Beyond these results, the major argument against the hypothesis $H_0 : \alpha = \beta = \gamma = 1$ is the percentage of rejection of the null hypothesis for most the cases in the simulations carried out (see Table 2 and Tables 3, 4 of the supplementary material). These results are in agreement with the studies performed by [15], where the authors provide some evidence supporting that, in general, H_0 is not true.

In addition, Table 3 report average values of the SSIM index (1) evaluated at $\boldsymbol{\theta} = \mathbf{1}$ and $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, or equivalently, under H_0 and H_1. The constants contained in the SSIM index have been chosen using the guidelines given in [3]. We note that every time that H_0 is rejected, the SSIM under H_0 yields an overestimation of

Table 1 Averages of parameter estimates for the SSIM index for Texmos2.S512 image. The results were obtained from 1,000 Monte Carlo simulations.

Number	No filter			Lee filter		
of looks	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$
1	1.000	1.000	1.000	1.178	1.233	1.166
2	1.000	1.000	1.000	1.247	1.326	1.241
4	1.037	1.051	1.037	1.308	1.400	1.304
8	1.082	1.113	1.082	1.431	1.540	1.429
16	1.160	1.214	1.160	1.613	1.748	1.612
32	1.297	1.386	1.297	1.825	1.958	1.824
Number	Enhanced Lee filter			Kuan filter		
of looks	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$
1	1.295	1.375	1.278	1.254	1.132	1.102
2	1.473	1.598	1.460	1.496	1.409	1.317
4	1.573	1.712	1.566	1.503	1.611	1.488
8	1.763	1.920	1.759	1.746	1.913	1.743
16	1.946	2.122	1.942	1.906	2.066	1.905
32	2.122	2.330	2.120	2.190	2.355	2.191

Table 2 Rejection percentages of $H_0 : \alpha = \beta = \gamma = 1$ for Texmos2.S512 image. The results were obtained from 1,000 Monte Carlo simulations.

Number	Filter			
of looks	None	Lee	Enhanced Lee	Kuan
1	0.0	100.0	100.0	100.0
2	100.0	100.0	100.0	100.0
4	100.0	100.0	100.0	100.0
8	100.0	100.0	100.0	100.0
16	100.0	100.0	100.0	100.0
32	100.0	100.0	100.0	100.0

Table 3 Averages of the SSIM index estimates for Texmos2.S512 image. The results were obtained from 1,000 Monte Carlo simulations.

Number	No filter		Lee filter	
of looks	Under H_0	Under H_1	Under H_0	Under H_1
1	0.551	0.551	0.758	0.722
2	0.690	0.690	0.857	0.824
4	0.802	0.796	0.917	0.893
8	0.882	0.873	0.953	0.934
16	0.933	0.923	0.974	0.958
32	0.963	0.953	0.986	0.974
Number	Enhanced Lee filter		Kuan filter	
of looks	Under H_0	Under H_1	Under H_0	Under H_1
1	0.772	0.715	0.795	0.771
2	0.883	0.830	0.888	0.851
4	0.940	0.907	0.940	0.911
8	0.969	0.945	0.968	0.945
16	0.984	0.968	0.983	0.968
32	0.991	0.981	0.991	0.980

the similarity between the images (see Table 3 and Tables 5 and 6 from the supplementary material), which can lead to a wrong impression of its closeness.

Figures 1 to 12 in the supplementary material display the empirical distribution function (CDF) of the SSIM coefficient under H_0 and H_1 for the images texmos2.S512, Lena and Baboon, considering several filters. For each reference image there is a big separation between the CDFs when H_0 is rejected. Indeed, the difference between these distributions has been tested using the Kolmogorov-Smirnov test, leading in each case to a significant difference at a level 0.05. Moreover, the distribution of H_0 has been placed in the right side of each graph indicating that the association between the reference and simulated images is larger than the actual value. From the processing of these images we also highlight the agreement between the SSIM coefficients when the Kuan and Enhanced Lee filters are used. The findings summarized here are not only valid for the images displayed in Figure 1. We carried out the same analysis for other images taken from the same database and the data analysis (not shown here) corroborates the findings discussed in this section.

To test the hypothesis $H_0 : \alpha = \beta = \gamma = 1$, we used the gradient test defined by (8). The rejection of H_0 was highly significant, despite the fact that in certain cases the similarity index is high. This provides evidence against the use of the SSIM index with the assumption $\alpha = \beta = \gamma = 1$.

3.2 Application using SAR imagery

The SAR imagery used in this study correspond to the Copeland, Kansas; Corpus Christi, Texas and the Mississippi River Delta in the state of Louisiana, USA and an image taken from Ullúm's dam in Argentina (see Figure 2). These images are characterized by presenting different landscapes from the structural point of view (for example hilly regions and inland valleys) along the coast and river valleys, agricultural lands and urban areas. This diversity of landscapes provides the ideal scenario to evaluate the methodology introduced in Section 2.

Figures 2 (a), (c) and (d) were obtained by the ICEYE SAR satellite constellation from 2019 to October 2020. Datasets are available in single look complex processing level with a ground range resolution of 3m. Figure 2 (b) is a Landsat image taken from the image database at Universidad Nacional de San Juan (UNSJ). This dataset was used in [30] for addressing the spatial association among several processes. In a preprocessing stage each image was converted to grayscale and rescaled it to the interval $[0, 1]$. In order to reduce

Table 4 Parameter estimates and SSIM index under H_0 and H_1 for SAR reference images.

Copeland						Corpus Christi				
Filter	Estimates			SSIM		Estimates			SSIM	
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	Under H_0	Under H_1	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	Under H_0	Under H_1
Lee	1.891	1.932	3.136	0.680	0.326	1.610	1.639	2.705	0.828	0.614
Enhanced Lee	1.691	1.722	2.987	0.652	0.306	1.600	1.626	2.512	0.818	0.615
Kuan	1.691	1.722	2.987	0.652	0.306	1.600	1.626	2.512	0.818	0.615
Dam						Mississippi				
Filter	Estimates			SSIM		Estimates			SSIM	
	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	Under H_0	Under H_1	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	Under H_0	Under H_1
Lee	1.000	1.000	1.000	0.999	0.999	1.591	1.623	2.670	0.890	0.740
Enhanced Lee	1.000	1.000	1.000	0.999	0.999	1.468	1.491	2.553	0.884	0.737
Kuan	1.000	1.000	1.000	0.999	0.999	1.468	1.491	2.553	0.884	0.737

the speckle noise, all the images were filtered using Lee, enhanced Lee and Kuan multiplicative filters with a moving window of size 3×3.

Results reported in Table 4 allow us to question the working assumption that $\alpha = \beta = \gamma = 1$. Indeed, for the SAR images Copeland, Corpus Christi and Mississippi the null hypothesis $H_0 : \alpha = \beta = \gamma = 1$ is rejected. It should be stressed that the rejection is highly significant for each of the filters used. Additionally, in all cases where H_0 is rejected, the SSIM evaluated at the estimators for α, β and γ is lower than their counterparts when we assume the working assumption. This may give a misleading idea about their degree of similarity. Finally, these results allow us to confirm our previous findings reported in the simulation study.

4 Concluding Remarks and Discussion

This paper has presented a novel method for estimating the SSIM index based on a nonlinear statistical model. To the best of our knowledge this is the first attempt to present parameter estimates of α, β and γ from a statistical perspective. A hypothesis-testing approach has been developed and efficiently implemented to test whether the parameters of the SSIM index are all equal to one. The methodology is accompanied by suitable routines developed in R and C that maintain a low computational complexity. The Monte Carlo experiment provided evidence that is in agreement with the results of other approaches found in the literature, and highlighted both, the use of the estimated SSIM in practice, and the overestimation of the SSIM that can occur when all parameters are assumed to be equal to one. Our findings detailed in the Monte Carlo simulation study are confirmed using real SAR images. In summary, we suggest to practitioners use the weighted version of the SSIM in which the parameters of the index are estimated via Algorithm 1, instead of assuming that $\alpha = \beta = \gamma = 1$. Indeed, the traditional assumption may lead to misjudge the similarity between two images.

Further exploration with remote-sensing data is needed to characterize the range of the estimates. This is one way to explore how the variable parameters can be characterized for a set group of images with similar characteristics. A study in this respect could be implemented following the guidelines given in [20].

Several related problems arise from the methodology suggested in this article. For instance, the definition of a similarity index based on the SSIM index for a multiplicative model of the form $Z = X \cdot Y$, where X models the terrain backscatter and Y models the speckle noise [20], and improvement of the estimation procedure following the guidelines of the algorithm proposed by [31] appear to be challenging problems to be addressed in future research.

Another relevant problem is the estimation of the SSIM in those situations where an additive model is appropriate. In such a case, a way to relate the SSIM and the MSE that we plan to study is through the use of a likelihood-based approach. Defining

$$Z_i = \log(\mathrm{MSE}(\boldsymbol{x}_i, \boldsymbol{y}_i)),$$
$$\boldsymbol{u}_i = [\log(l(\boldsymbol{x}_i, \boldsymbol{y}_i)), \log(c(\boldsymbol{x}_i, \boldsymbol{y}_i)), \log(s(\boldsymbol{x}_i, \boldsymbol{y}_i))]^\top,$$

and $\boldsymbol{\theta} = (\alpha, \beta, \gamma)^\top$. Then, the model of interest is

$$Z_i = \boldsymbol{u}_i^\top \boldsymbol{\theta} + \epsilon_i, \quad i = 1, \ldots, n. \tag{10}$$

Let us assume that D_i represents the mismeasured response and let $f(z|\boldsymbol{u}_i, \boldsymbol{\omega})$ and $f(d_i|z)$ denote the densities of Z_i and $D_i|Z_i = z$, respectively, where $\boldsymbol{\omega} =$

$(\boldsymbol{\beta}^\top, \sigma^2)^\top$. Then, D_i has a marginal density given by

$$f(d_i|\boldsymbol{x}_i, \boldsymbol{\omega}) = \int f(d_i|z) f(z|\boldsymbol{x}_i, \boldsymbol{\omega}) \, \mathrm{d}\, z. \tag{11}$$

Consequently, the likelihood function of $\boldsymbol{w}$ is

$$L(\boldsymbol{\omega}) = \prod_{i=1}^{n} f(d_i|\boldsymbol{x}_i, \boldsymbol{\omega}). \tag{12}$$

The MLE of $\boldsymbol{\omega}$ can be obtained using the EM algorithm. Full details regarding the computational aspects associated with the maximum likelihood estimation under normal models with constant variance and asymptotic results for the MLE estimates in a calibration context can be found in [32]. The SSIM estimation using the likelihood (12) from a general nonlinear modeling perspective with error in the response appear to be a challenging problem to address in further research.

Acknowledgements The authors acknowledge the suggestions and comments from the Associate Editor and anonymous referees that led to a significant improvement of the manuscript. Felipe Osorio and Ronny Vallejos acknowledge financial support from CONICYT through the MATH-AMSUD program, grant 20-MATH-03, from UTFSM, grants PI_LI_19_11 and P_LIR_2020_20. Ronny Vallejos was also partially funded by the Advanced Center for Electrical and Electronic Engineering (AC3E), grant FB-0008. Silvia M. Ojeda and Marcos Landi were supported by Secretaría de Ciencia y Tecnología (SeCyT) from Universidad Nacional de Córdoba. Proyecto Consolidar 2018-2121 (P.I.D. No. 33620180100055CB) y CIEM (Córdoba), CONICET.

References

1. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE T. Image Process. 21, 4695-4708 (2012).
2. Wang, K., Yong, B., Gu, X., Xiao, P., Zhang, X.: Spectral similarity measure using frequency spectrum for hyperspectral image classification. IEEE Geosci. Remote S. 12, 130-134 (2015).
3. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Proc. Let. 9, 81-84 (2002).
4. Zhang, L., Zhang, L., Mou, X., ZHang, D.: FSIM: A Feature Similarity Index for Image Quality Assessment. IEEE Transactions on Image Processing 20, 2378-2386 (2011).
5. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index IEEE T. Image Process. 21, 1488-1499 (2012).
6. Vallejos, R., Mancilla, D., Acosta, J.: Image similarity assessment based on coefficients of spatial association. J. Math. Imaging Vis. 56, 77-98 (2016).
7. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE T. Image Process. 13, 600-612 (2004).
8. Sheikh, H.R., Bovik, C.: Image information and visual quality. IEEE T. Image Process. 15, 430-444 (2006).
9. Chandler, D.M., Hemami, S.S.: VSNR: A wevelet-based visual sinal-to-noise-ratio for natural images. IEEE T. Image Process. 16, 2284-2298 (2007).
10. Ma, K, Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. IEEE T. Image Process. 24, 3345-3356 (2015).
11. Wang, Y.K., Li, L., Zhou, X.Y., Cui, T.J.: Supervised automatic detection of UWB ground-penetrating radar targets using the regression SSIM measure. IEEE Geosci. Remote S. 13, 621-625 (2016).
12. Ojeda, S., Vallejos, R., Lamberti, P.: Measure of similarity between images based on the codispersion coefficient. Journal Electron. Imaging 21, 023019 (2012).
13. Wang, H., Maldonado, D., Silwal, S.: A nonparametric-test-based structural similarity measure for digital images. Comput. Stat. Data An. 55, 2925-2936 (2011).
14. Rehuman, A., Wang, Z.: Reduced-reference SSIM estimation. Proceedings of 2010 IEEE 17th International Conference on Image Processing. Hong Kong, 26-29 (2010).
15. Wang, Z., Li, L., Wu, S., Xia, Y., Wan, Z., Cai, C.: A new image quality assessment algorithm based on SSIM and multiple regressions. Int. J. Sig. Process. 8, 221-230 (2015).
16. Dosselmann, R., Yang, X.D.: A comprehensive assessment of the structural similarity index. SIViP 5, 81–91 (2011).
17. Yeo, C., Tan, H.L., Tan, Y.H.: On rate distortion optimization using SSIM. IEEE T. Circ. Syst. Vid. 23, 1170-1181 (2013).
18. Kim, S., Pak, D., Lee, S.: SSIM-based distortion metric for film grain noise in HEVC. SIViP 12, 489-496 (2018).
19. Davidian, M. Carroll, R.J.: Variance function estimation. J. Am. Stat. Assoc. 82, 1079-1091 (1987).
20. Frery, A.C., Müller, H.J., Yanasse, C.C.F., Sant'Anna, S.J.S.: A model for extremely heterogeneous clutter. IEEE T. Geoscie. Remote 35, 648-659 (1997).
21. Goldfeld, S.M., Quandt, R.E.: Nonlinear Methods in Econometrics. North-Holland Publishing Company, Amsterdam (1972).
22. Seber,G.A.F., Wild, C.J.: Nonlinear Regression. Wiley, New York (1988).
23. Cribari-Neto, F., Frery, A.C., Silva, M.F.: Improved estimation of clutter properties in speckled imagery. Comput. Stat. Data An. 40, 801-824 (2002).
24. Carroll, R.J., Ruppert, D.: Transformation and Weighting in Regression. Chapman and Hall, New York (1988).
25. Brent, R.P.: Algorithms for Minimization without Derivatives. Dover, New York (1973).
26. Nash, J.C.: Compact Numerical Methods for Computers. Linear Algebra and Function Minimization. Adam Hilger, Bristol (1990).
27. Gorieroux, C., Monfort, A.: Statistics and Econometrics Models: General concepts, estimation, prediction, and algortihms. Cambridge University Press, Cambridge (1995).
28. Terrell, G.R.: The gradient statistic. Comp. Sci. Stat. 34, 206-215 (2002).
29. Osorio, F., Vallejos, R.: SpatialPack: Tools for assessment the association between two spatial processes. R package version 0.3-8196. URL: CRAN.R-project.org/package=SpatialPack (2020).
30. Vallejos, R., Osorio, F., Bevilacqua, M. Spatial Relationships Between Two Georeferenced Variables: With Applications in R. Springer, Cham (2020).
31. Loader, C., Pilla, R.S.: Iteratively reweighted generalized least squares for estimation and testing with correlated data: An inference function framework. J. Computat. Graph. Stat. 16, 925-945 (2007).
32. Buonaccorsi, J.P.: Measurement error in the response in the general linear model. J. Am. Stat. Assoc. 91, 633-642 (1996).