UPC – ETSETB

INPT – ENSEEIHT – IRIT

Study of the bivariate Pearson system for the statistical description of optical and SAR images

A Master Thesis submitted by Víctor Manuel CABRERA SERRANO

Toulouse, April 2013

Supervised by:
Marie Chabert

Abstract

This report presents the work performed during my Erasmus Master Thesis in the study of the bivariate Pearson system, for the purpose of providing an accurate statistical description of remote sensing images acquired by heterogeneous sensors — specifically, optical and synthetic aperture radar (SAR) images—. This work is based on [1].

Contents

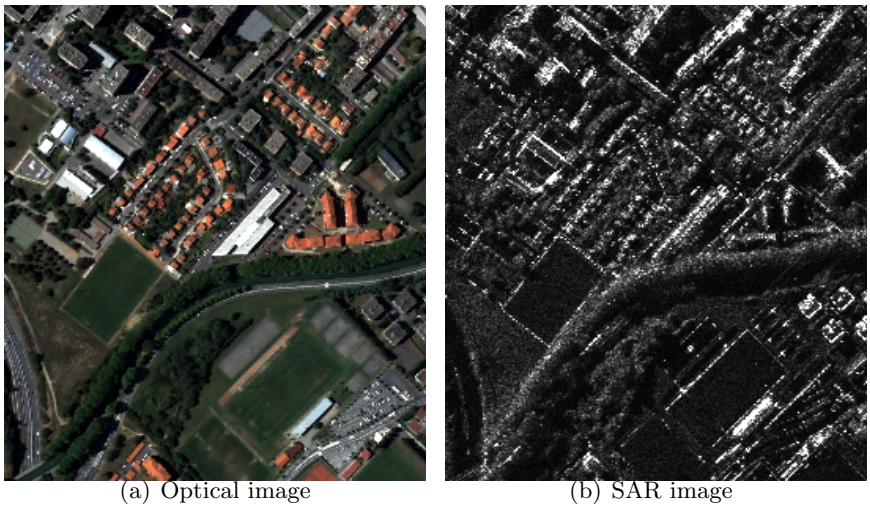
1 Introduction

Currently, remote sensing imagery has many applications in geoscience, for example, Earth observation satellites have an important role to combat the deforestation or to provide useful information after a natural disaster. These applications often combine images collected from several sensors to obtain heterogeneous data, which can be processed to extract important features of these images, for instance, correlation coefficient or mutual information that are very useful to change detection.

For example, SAR images can be acquired under any weather conditions, by day or by night; in contrast, they are corrupted by speckle noise. On the other hand, optical images have a better image quality, but under specifically weather conditions. For this reason, an appropriate and flexible statistical model is necessary to provide this resulting heterogeneous data, hence this project continues the study of accuracy of multivariate Pearson system properties (specifically, bivariate case) using the method of moments to estimate the parameters of the distributions which has been performed in [1].

The report is organized as follows: Section 2 introduces the Pearson system and the methods of moments. Section 3 presents several simulation results related to the goodness of fit of Pearson system. The results of applying the Kolmogorov–Smirnov test over the above results are presented in Section 4. Conclusion are reported in Section 5.

The real images processed in this project belong a suburban area of Toulouse (France) — these are provided by the CNES of Toulouse for the work performed in [1].



(a) Optical image (b) SAR image

Figure 1: Remote sensing images

2 The Pearson system and the method of moments

2.1 Introduction

This section introduces the univariate and multivariate Pearson system, focusing on the importance of the bivariate case for the statistical modeling of two remote sensing images. Also, the adequacy of the method of moments to estimate the parameters of the bivariate case is explained.

2.2 The Pearson system

As is explained in [2], Pearson distribution system is a good method to obtain random numbers from multivariate non-normal distributions, since it can represent wide class of distributions with various skewness and kurtosis (third- and fourth-order moment, respectively). For this reason, in [1] the Pearson system is defined as a good candidate to offer a highly flexible multivariate distribution for the heterogeneous data collected by several sensors.

2.2.1 Univariate case

The Pearson distribution system is defined by the following differential equation characterized by the probability density function p [2]:

$$-\frac{p'(x)}{p(x)} = \frac{b_0 + b_1 x}{c_0 + c_1 x + c_2 x^2} \tag{1}$$

and its types are classified following the expression:

$$\kappa = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)} \tag{2}$$

where β_1 and β_2 are the squared skewness ($E[X^3]^2$) and kurtosis ($E[X^4]$), respectively, being X a random variable with mean equal to zero and variance equal to one. According to the values of β_1 and β_2 the Pearson distribution system is divided into eight types:

- Type 0: Gaussian distribution ($\beta_1 = 0, \beta_2 = 3$)
- Type I: Beta distribution with non-zero skewness
- Type II: Beta distribution with zero skewness ($\beta_1 = 0, \beta_2 < 3$)
- Type III: Gamma distribution
- Type IV: non standard distribution
- Type V: inverse-gamma distribution
- Type VI: F-distribution
- Type VII: Student distribution ($\beta_1 = 0, \beta_2 > 3$)

As is mentioned previously, the Gaussian distribution can model residual noise in optical images and gamma distribution has been used to describe SAR images [1].

2.2.2 Multivariate case – Bivariate case

In [2] the random vector of multivariate non-normal distribution is defined as

$$\boldsymbol{X} = \boldsymbol{\nu}\boldsymbol{M}\boldsymbol{\xi}, \tag{3}$$

and the author remarks that when $\boldsymbol{\nu}$ is one, the distributions are represented by the pdf of the Pearson distribution system, then

$$\boldsymbol{X} = \boldsymbol{M}\boldsymbol{\xi} \tag{4}$$

being $\boldsymbol{\xi} = (\xi_1, ..., \xi_p)^T$ a random vector with independent components distributed according to univariate Pearson distributions, with $E(\xi_j) = 0$, $E(\xi_j^2) = 1$, $E(\xi_j^3) = \zeta_j$, $E(\xi_j^4) = \kappa_j$; and being $\boldsymbol{M}$ a $p \times p$ deterministic matrix called mixing matrix.

Focusing on the bivariate case $p = 2$ [1], the random vector $\boldsymbol{X}$ is defined as $\boldsymbol{X} = (X_1, X_2)^T$ where X_1 and X_2 represents the data (pixel values) of optical and SAR images, respectively. Consequently, $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)$, $\boldsymbol{\zeta} = (\zeta_1, \zeta_2)$ and the mixing matrix $\boldsymbol{M}$ is defined as

$$\boldsymbol{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

The bivariate pdf of $\boldsymbol{X}$ is defined as

$$p_{\boldsymbol{X}}(\boldsymbol{X}) = \frac{p_{\boldsymbol{\xi}}(\boldsymbol{M}^{-1}\boldsymbol{X})}{|\boldsymbol{M}|}, \tag{5}$$

since $\boldsymbol{\xi}$ is independent, it is possible to obtain its marginal pdfs (density of a Pearson type). On the contrast, as is explained in [1], the components of $\boldsymbol{X}$ are not necessarily marginally Pearson distributed.

2.3 The method of moments

The authors in [1] explain that the method which seems more appropriate to estimate the parameters of the bivariate Pearson distribution is the method of moments, since in the Pearson case the maximum likelihood method generates several analytical expressions depending on the values of β_1 and β_2.

The method of moments estimates the unknown parameters $(\boldsymbol{M}, \boldsymbol{\zeta}, \boldsymbol{\kappa})$ minimizing the function defined by

$$\sum_{i=1}^{k} w_i (f_i(\boldsymbol{M}, \zeta_1, \zeta_2, \kappa_1, \kappa_2) - \widehat{f}_i)^2 \tag{6}$$

where f_i denotes the moments of $\boldsymbol{X}$ up the fourth order (listed below), $\widehat{f}_i$ denotes the empirical moments calculated from the real data and w_i represents the weights to modulate the function.

$$
\begin{aligned}
f_1 = E(X_1^3) &= m_{11}^3\zeta_1 + m_{12}^3\zeta_2, \\
f_2 = E(X_2^3) &= m_{21}^3\zeta_1 + m_{22}^3\zeta_2, \\
f_3 = E(X_1^2X_2^1) &= m_{11}^2m_{21}^2\zeta_1 + m_{12}^2m_{22}\zeta_2, \\
f_4 = E(X_1^1X_2^2) &= m_{11}m_{21}^2\zeta_1 + m_{12}m_{22}^2\zeta_2, \\
f_5 = E(X_1^4) &= m_{11}^4\kappa_1 + 6m_{11}^2m_{21}^2 + m_{12}^4\kappa_2, \\
f_6 = E(X_2^4) &= m_{21}^4\kappa_1 + 6m_{22}^2m_{12}^2 + m_{22}^4\kappa_2, \\
f_7 = E(X_1^3X_2^1) &= m_{11}^3m_{21}\kappa_1 + 3(m_{11}m_{21}^3 + m_{11}^2m_{22}m_{21}) + m_{21}^3m_{22}\kappa_2, \\
f_8 = E(X_1^2X_2^2) &= m_{11}^2m_{21}^2\kappa_1 + (m_{12}^4 + 4m_{11}m_{21}^2m_{22} + m_{11}^2m_{22}^2) + m_{21}^2m_{22}^2\kappa_2, \\
f_9 = E(X_1^1X_2^3) &= m_{11}m_{21}^3\kappa_1 + 3(m_{22}m_{12}^3 + m_{22}^2m_{11}m_{21}) + m_{21}m_{22}^3\kappa_2, \\
f_{10} = E(X_1^2) &= m_{11}^2 + m_{12}^2, \\
f_{11} = E(X_2^2) &= m_{22}^2 + m_{12}^2, \\
f_{12} = E(X_1^1X_2^1) &= m_{12}(m_{11} + m_{22}).
\end{aligned}
$$

To obtain the parameters, firstly, the mixing matrix $\boldsymbol{M}$ is estimated using the characteristic of the covariance of the vector $\boldsymbol{X}$ (since ξ_i are independent)

$$\boldsymbol{\Sigma} = \boldsymbol{M}\boldsymbol{M}^T \tag{7}$$

where $\boldsymbol{\Sigma}$ denotes the covariance. Then, the parameters $\zeta_1, \zeta_2, \kappa_1$ and κ_2 can be estimated from the mixing matrix and the empirical moments $\widehat{f}_i$ solving a linear system (see detailed in [1], Section 3).

3 Simulation results

3.1 Introduction

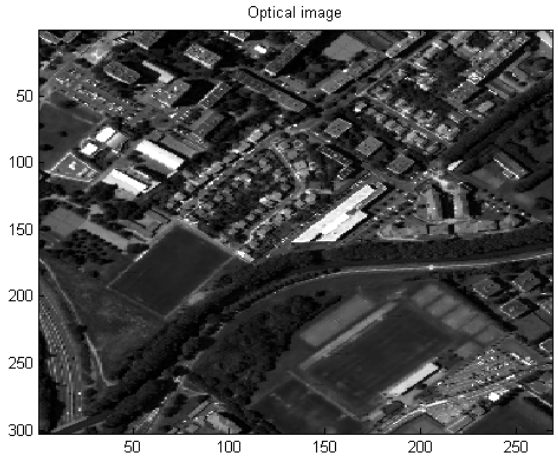
In this section, a set of simulations evaluates the level of agreement between real data and estimated Pearson model for both SAR and optical images. The simulations consider the whole and windowed image (homogeneous and heterogeneous areas) to obtain the results.

3.2 Whole image results

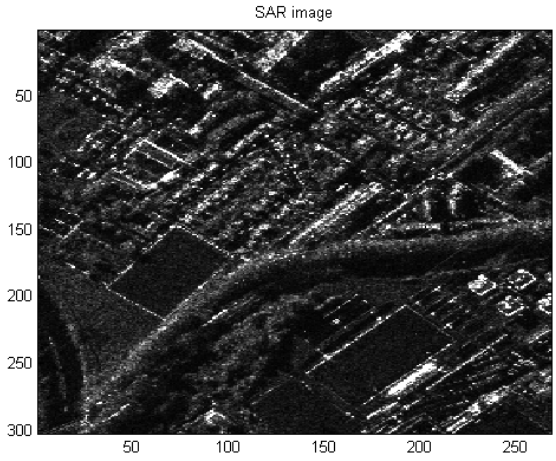
In the whole image case, the random variables X_1 and X_2 represent the whole pixel values from optical and SAR images, respectively. To generate the Pearson model, as is explained in Section 2, firstly, it is necessary to calculate the value of the mixing matrix $\boldsymbol{M}$ using the covariance of $\boldsymbol{X} = (X_1, X_2)^T$ (the pixel values are translated to the grey level and the variables are centered). Secondly, using the method of moments, the values of $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)$ and $\boldsymbol{\zeta} = (\zeta_1, \zeta_2)$ are obtained. Then, with these values of the third- and fourth-order moments, and with the mean and the variance values (highly close to zero and one, since the variables are centered), the independent components, according to univariate Pearson distribution, $\boldsymbol{\xi} = (\xi_1, \xi_2)$ are generated using a number of samples two order of magnitude greater than the real data (N_{pixels}·100). Finally, applying expression (4), the bivariate Pearson model to model the two remote sensing images are obtained.

In Figs. 2-c) and 2-d) are compared the empirical histograms from the real data of the two images (blue) with the estimated marginal Pearson model (red). In this case, as can been seen, the Pearson distributions are not in clearly good agreement with the histograms from the real images, specially, in SAR.

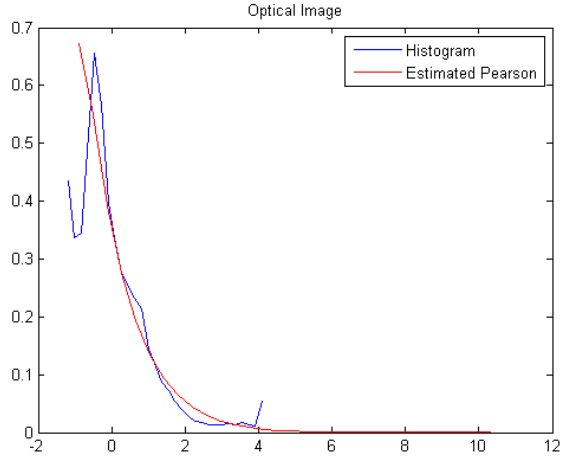
Figs. 3-a) and 3-b) display the logarithmic two-dimension histograms for optical and SAR images pair from real and estimated Pearson data, respectively. Therefore, in these Figures, the joint events between the two types of remote sensing images are showed for both cases (real and Pearson system).



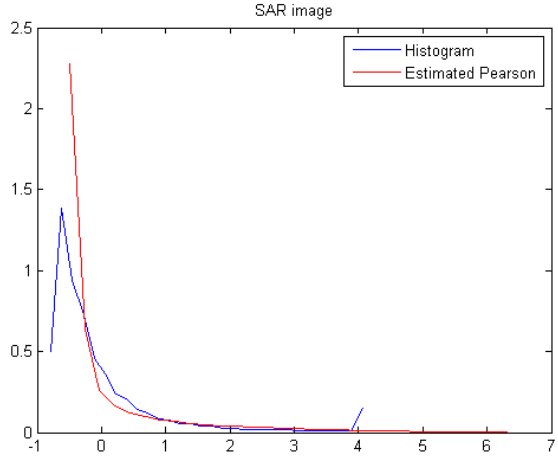
(a) Optical image



(b) SAR image

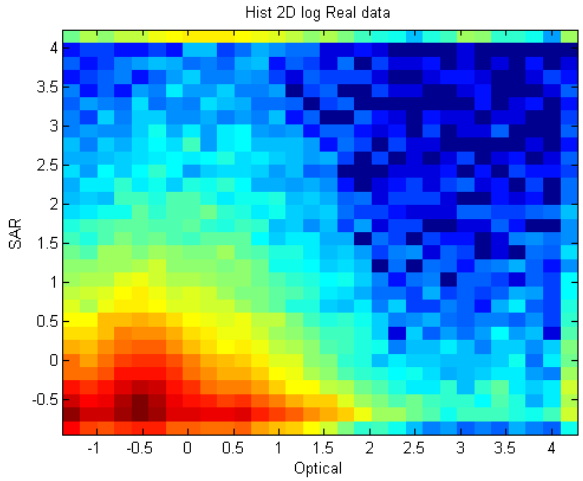


(c) Optical histogram 1D

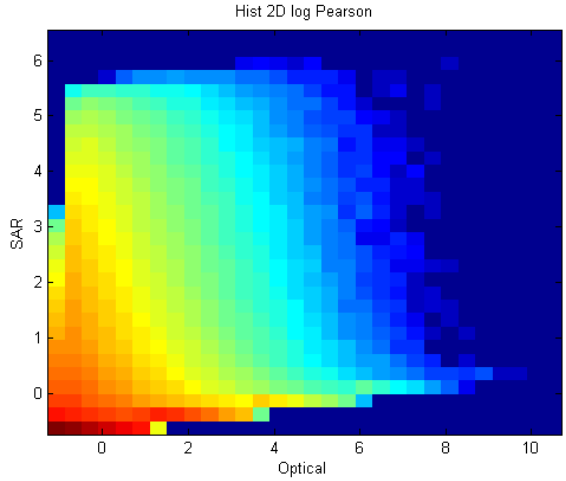


(d) SAR histogram 1D

Figure 2: Level of fit – One-dimensional histogram for whole image



(a) Real data histogram 2D



(b) Pearson model histogram 2D

Figure 3: Two-dimensional histogram (log) for whole image

3.3 Windowed image results

In the case of windowed images, the method to obtain the estimated Pearson model data is the same that in the sub-section above, with the difference that the values of X_1 and X_2 denote the pixel values of a specific area from optical and SAR images, respectively. That is to say, the data correspond to two sub-images and the parameters are estimated from these windowed regions data.

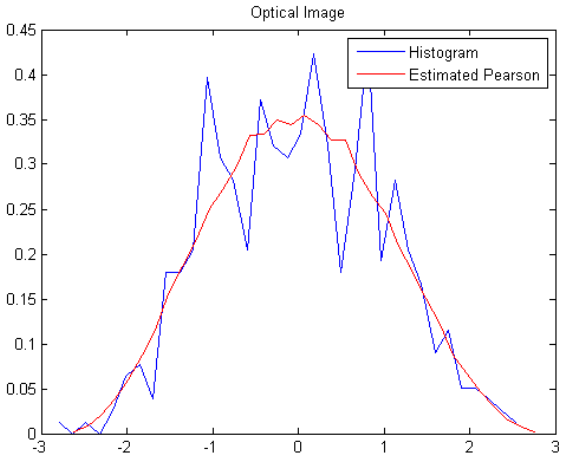
The aforementioned area consists of a square window of side 22 ($N_{pixels} = 22{\cdot}22$).

In the following Figures are showed the results for several representative areas of the images (homogeneous and heterogeneous types).

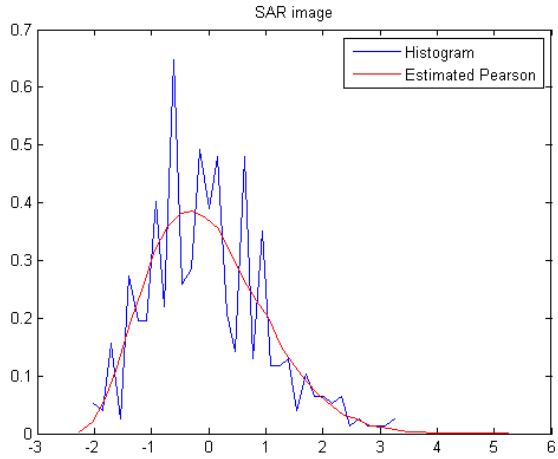
3.3.1 Homogeneous areas

(a) Optical image

(b) SAR image

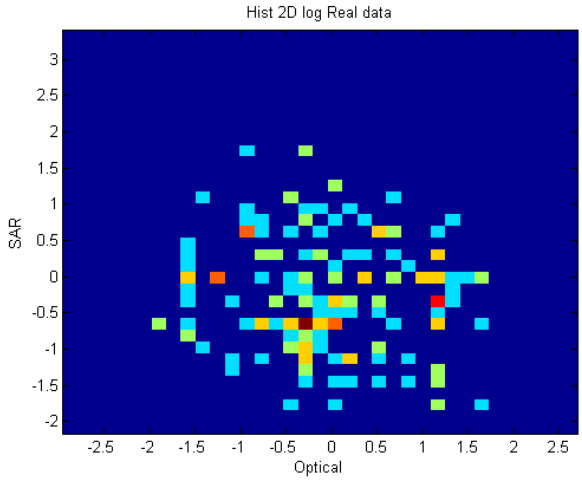


(c) Optical histogram 1D

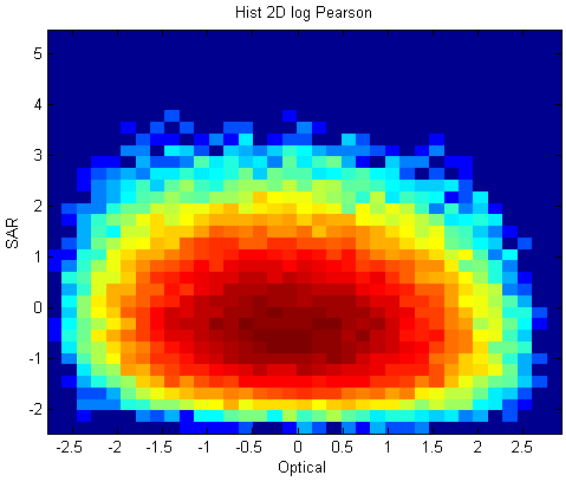


(d) SAR histogram 1D

Figure 4: Level of fit – One-dimensional histogram for Field 1 – homogeneous area

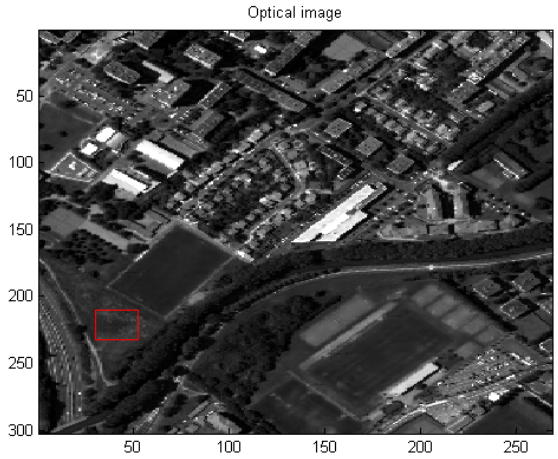


(a) Real data histogram 2D

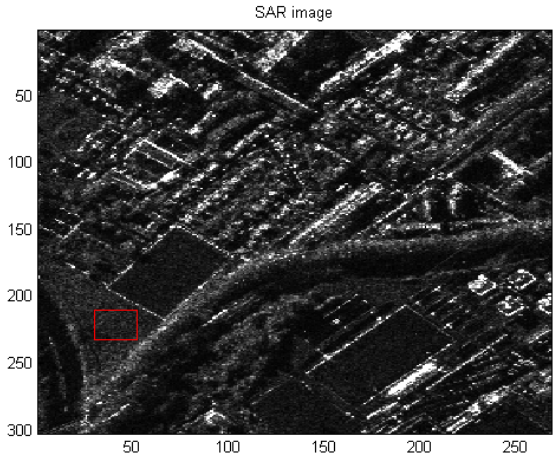


(b) Pearson model histogram 2D

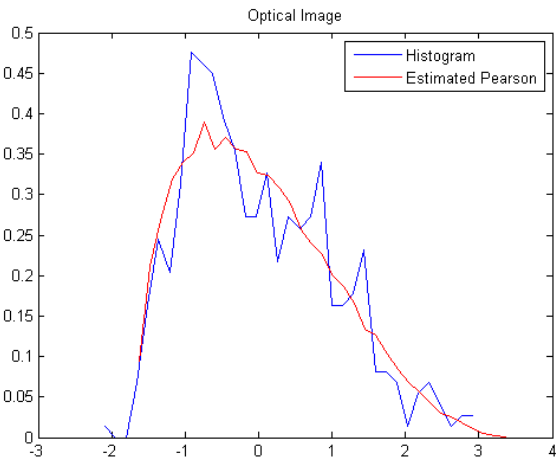
Figure 5: Two-dimensional histogram (log) for Field 1 – homogeneous area



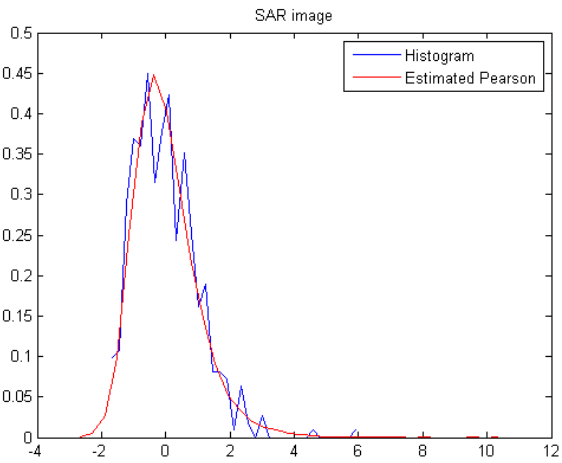
(a) Optical image



(b) SAR image

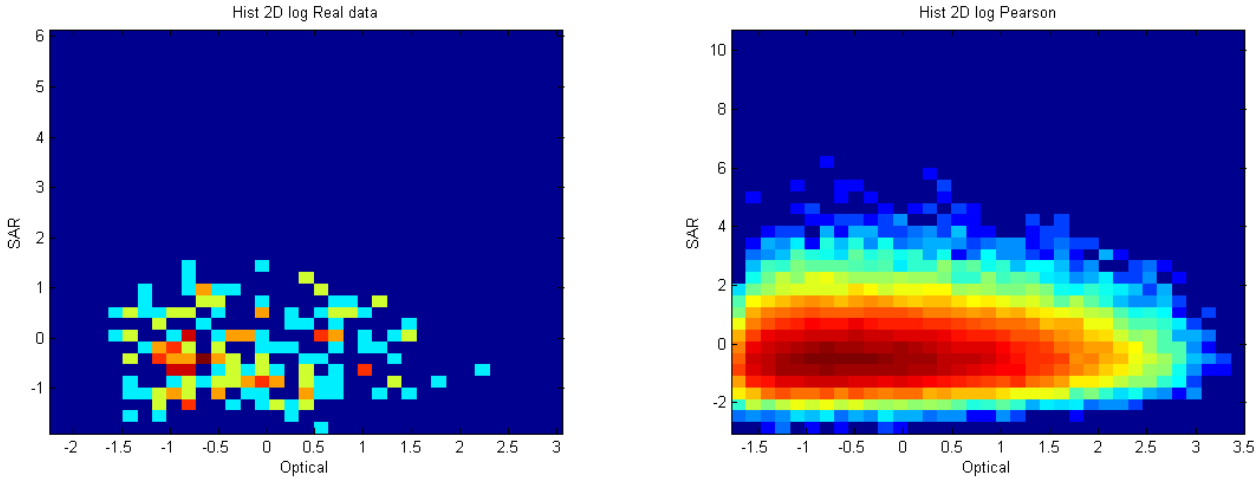


(c) Optical histogram 1D



(d) SAR histogram 1D

Figure 6: Level of fit – One-dimensional histogram for Field 2 – nearly homogeneous area



(a) Real data histogram 2D

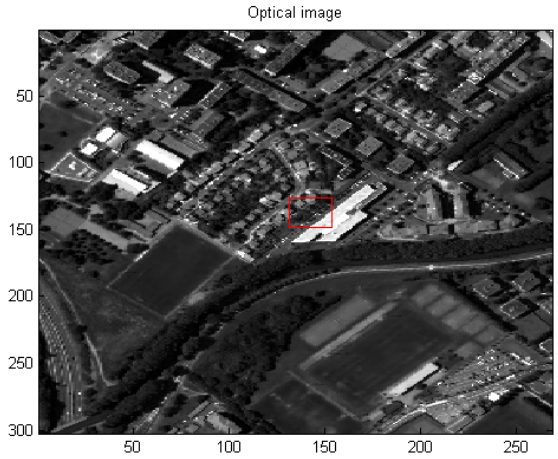
(b) Pearson model histogram 2D

Figure 7: Two-dimensional histogram (log) for Field 2 – nearly homogeneous area

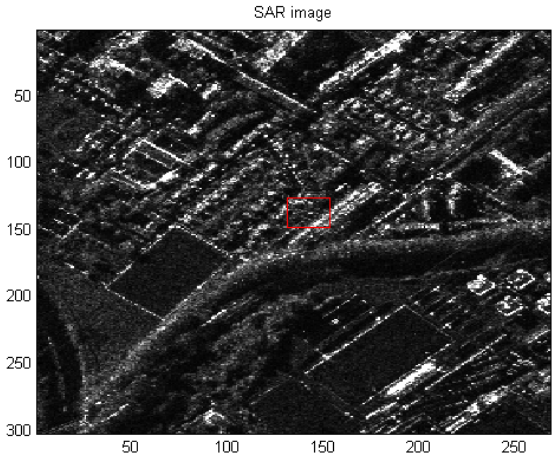
In Figs. 4-c) and 4-d) can be seen the good level of agreement between the histograms for the real images and the marginal Pearson distributions for field homogeneous area (Field 1). Moreover, as has been mentioned previously, in this case is clear to see that a homogeneous region from a optical image can be modeled by the Gaussian distribution and gamma distribution has been used to describe the homogeneous region of a SAR image.

Figs. 6-c) and 6-d) shows the level of fit for an area which is not clearly homogeneous (Field 2). Although, for this region the marginals of two images does not follow a Gaussian and gamma distribution, respectively, (specifically, the optical one); the marginal Pearson models and the real data of the images are in good agreement. So, the Pearson system also seems appropriate for this case.

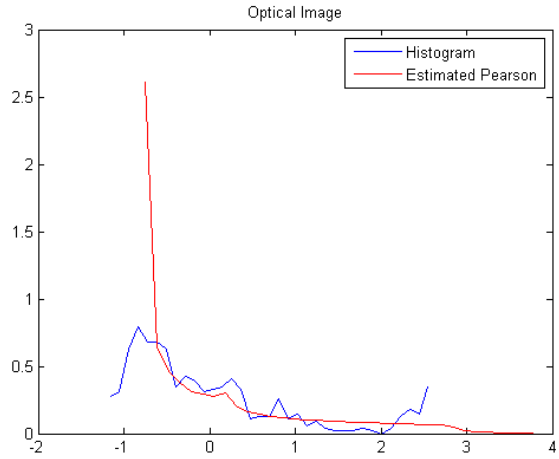
3.3.2 Heterogeneous areas



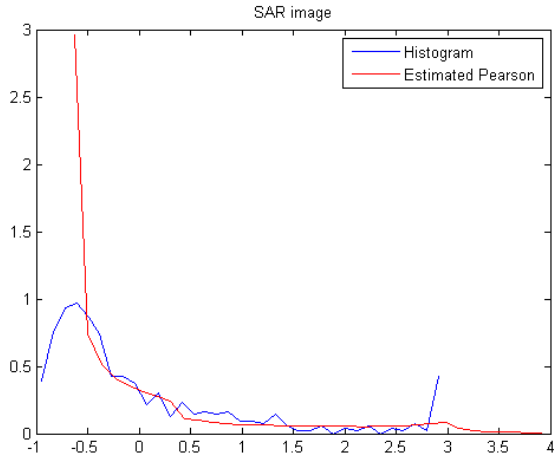
(a) Optical image



(b) SAR image

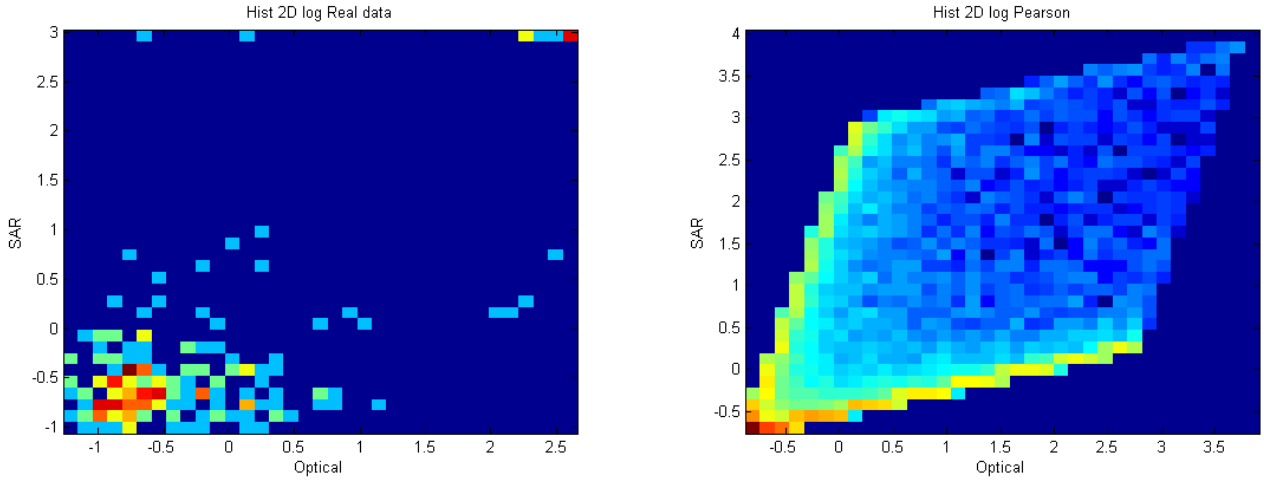


(c) Optical histogram 1D



(d) SAR histogram 1D

Figure 8: Level of fit – One-dimensional histogram for Buildings – heterogeneous area



(a) Real data histogram 2D

(b) Pearson model histogram 2D

Figure 9: Two-dimensional histogram (log) for Buildings – heterogeneous area

In Figs. 8-c) and 8-d) are displayed the histograms from the real data and the marginal distributions of estimated Pearson model for a representative strongly heterogeneous area corresponding to the buildings zone. For this case, it is clearly to see that the Pearson distribution system can not model the heterogeneous area.

It is important to mentioning that for some heterogenous areas belonging to the top of the both real images, the obtained values of $\boldsymbol{\zeta}$ and $\boldsymbol{\kappa}$ are not included in the range of Pearson distribution system, since $\boldsymbol{\kappa} >> (\boldsymbol{\zeta})^2 + 1$.

In Fig. 9-b) it is possible to perceive clearly the linear combination between the optical and SAR Pearson model.

4 Goodness of fit test: Kolmogorov–Smirnov results

4.1 Introduction

This section reports the results of applying Kolmogorov–Smirnov test over the set of simulations of Section 3.3. The purpose is to continue studying the level of agreement between real data and Pearson model to get more conclusive results.

An explanation of Kolmogorov–Smirnov test is introduced firstly, next, the test results in one-dimensional (univariate) and two-dimensional (bivariate) case are presented.

4.2 The Kolmogorov–Smirnov test

The Kolmogorov–Smirnov statistic is one of the most important test of goodness of fit, which is based on the calculation of the distance between the empirical distribution function of a random sample and some specified distribution function.

Let F denote the distribution function of a sample $x_1, \ldots, x_n$ of a random variable X, and F_0 is some specified hypothesized distribution function. Considering the null hypothesis as $H_0 : F = F_0$, and on the other hand, $H_1 : F \neq F_0$; and the Kolmogorov–Smirnov statistic as

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|, \tag{8}$$

being F_n the empirical distribution function of X. The test decides to accept the null hypothesis, i.e., two distributions are good fit, when the $p - value$ (possibility of obtaining this result if the null hypothesis is true), obtained from D_n, is greater than the significance level α of the test; on the contrary, the null hypothesis is rejected. Therefore, the null hypothesis should be accepted in the case that the real data and the Pearson model distribution are alike.

4.3 One-dimensional case results

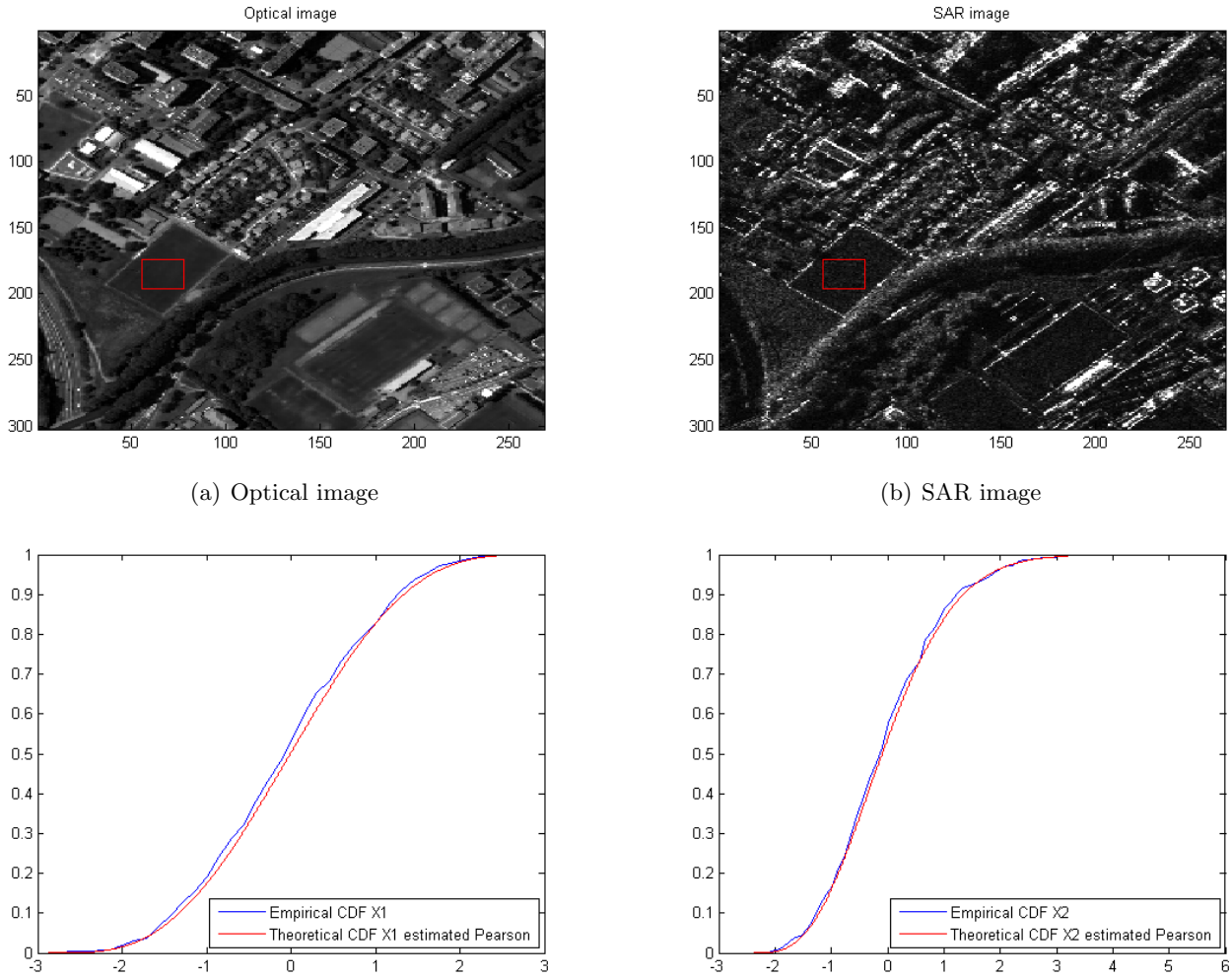
In one-dimension, Kolmogorov–Smirnov test follows the above definition (8), being samples $x_1, \ldots, x_n$ the data vector belonging to a windowed image of Section 3.3, i.e., the random variable X represents the subdata of optical (X_1) or SAR (X_2) windowed image. Thus, in the same windowed area, the test is applied separately over each type of remote sensing image, considering X_1 or X_2 univariate, and evaluating Kolmogorov–Smirnov for the marginal distributions of each one.

To effectuated the test, the empirical cumulative distribution $F_n(x)$ have been generated directly from the subdata vector, in contrast, the theoretical cumulative distribution $F(x)$ have been achieved through the Pearson model data, using the method of moments explained in Section 2.3 to estimate the parameters, and then, generating the samples. Despite the theoretical distribution function is also obtained using a empirical form, we can consider it as theoretical since it has been generated using a order of magnitude twice greater that the real data.

Figs. 10-c) and 10-d) shows the results of one-dimension Kolmogorov–Smirnov test[1] for the field homogeneous area (named previously Field 1, $size = 22{\cdot}22$). For both SAR and optical images, the empirical and theoretical cumulative distribution functions are clearly in good agreement, and the test parameters confirm that the null hypothesis is accepted ($H = 0$) and the $p - value$

[1] These one-dimension test results are achieved using the Kolmogorov–Smirnov function of MATLAB.

has a consistent value (see Table 1). For this region, we can observe that the good fit is stronger in the SAR image that in the optical one.



(a) Optical image

(b) SAR image

(c) Optical CDF comparison

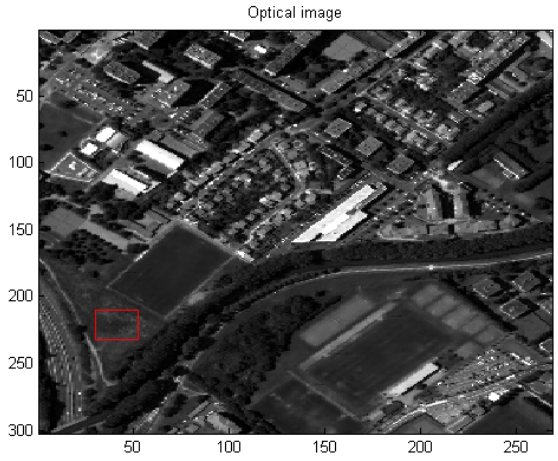
(d) SAR CDF comparison

Figure 10: One-dimensional Kolmogorov–Smirnov test for Field 1 – homogeneous area

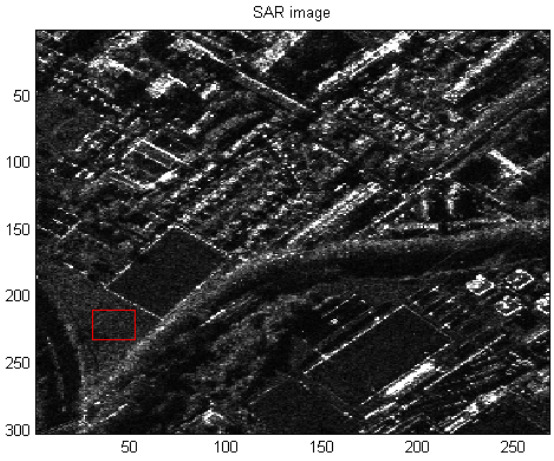
Image	Hypothesis (H)	$p-value$	CDF maximum distance
Optical	0	0.3378	0.0425
SAR	0	0.5456	0.036

Table 1: Parameters of One-dimensional Kolmogorov–Smirnov test for Field 1 – homogeneous area, at the significance level $\alpha = 0.05$

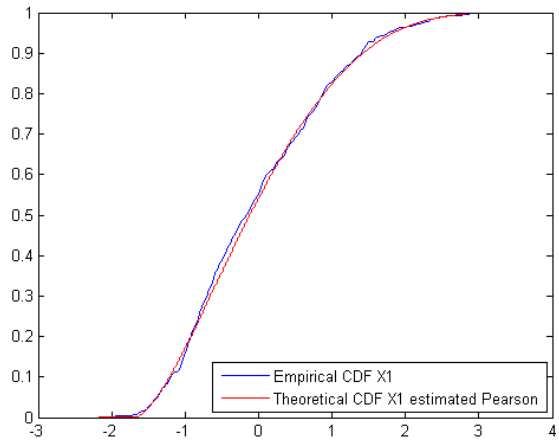
In the case of the nearly homogeneous area (Field 2, $size = 22{\cdot}22$) showed in Figs. 11-a) and 11-b, the results of Kolmogorov–Smirnov test prove that for this region the null hypothesis is also accepted for both images. Thus, we can get the confirmation that the Pearson system is acceptable to model this real subdata as was suspected in Section 3.3.1 (see Figs. 11-c) and 11-d), and Table 2).



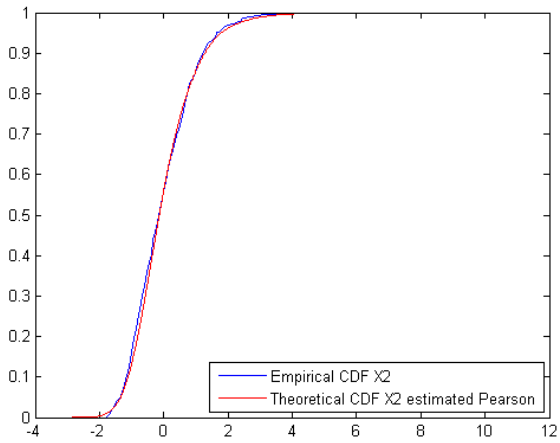
(a) Optical image



(b) SAR image



(c) Optical CDF comparison



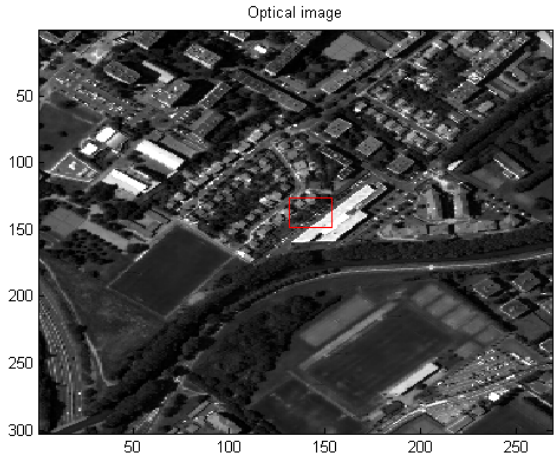
(d) SAR CDF comparison

Figure 11: One-dimensional Kolmogorov–Smirnov test for Field 2 – nearly homogeneous area

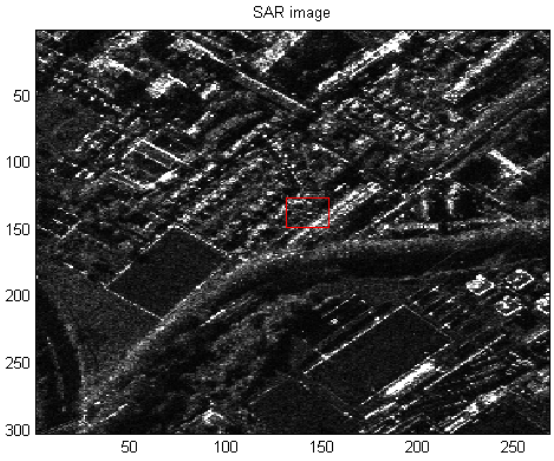
Image	Hypothesis (H)	$p-value$	CDF maximum distance
Optical	0	0.6035	0.0344
SAR	0	0.2466	0.0461

Table 2: Parameters of One-dimensional Kolmogorov–Smirnov test for Field 2 – nearly homogeneous area, at the significance level $\alpha = 0.05$

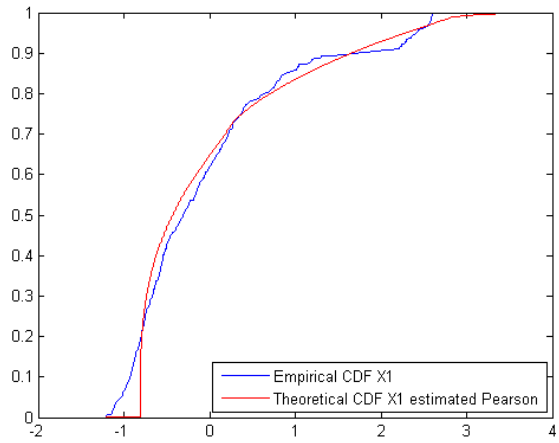
Figs. 12-c) and 12-d) shows the test results for the Buildings heterogeneous area ($size = 22{\cdot}22$). In this case, as expected, the test confirms that the Pearson model does not fit with the real data. As can see in Table 3, the null hypothesis is rejected ($H = 1$), the $p-value$ has an extremely small value for both images, and the maximum distance between cumulative distribution functions is an order of magnitude greater than in the above cases.



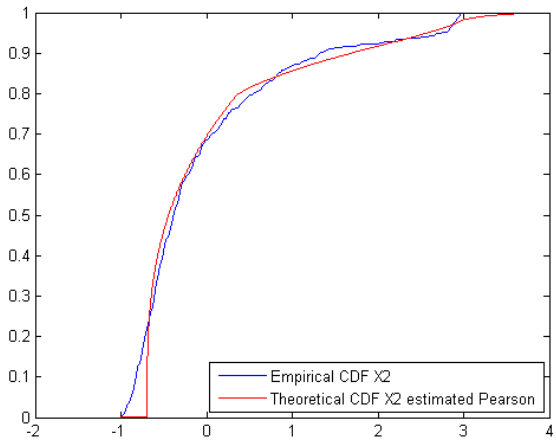
(a) Optical image



(b) SAR image



(c) Optical CDF comparison



(d) SAR CDF comparison

Figure 12: One-dimensional Kolmogorov–Smirnov test for Buildings – heterogeneous area

Image	Hypothesis (H)	$p-value$	CDF maximum distance
Optical	1	$4.367exp(\text{-}15)$	0.1859
SAR	1	$6.8618exp(\text{-}19)$	0.2086

Table 3: Parameters of One-dimensional Kolmogorov–Smirnov test for Buildings – heterogeneous area, at the significance level $\alpha = 0.05$

The application of one-dimensional Kolmogorov–Smirnov test confirms the results obtained in Section 3.3. As a consequence, it is possible to ensure that the Pearson model are in good agreement with the real subdata acquired from homogeneous regions of the images. In contrast, the model is not acceptable for heterogeneous regions.

4.4 Two-dimensional case results

Two-dimension Kolmogorov–Smirnov test presents a more serious difficulty that in the one-dimensional case, since the multivariate statistics are not distribution free[2] (see [3] [4]).

Now, the real data is a bivariate random variable which is constituted by the data from both SAR and optical images:

$$\boldsymbol{X} = (X_1, X_2), \tag{9}$$

as a consequence, it is not possible to apply directly the above definition of Kolmogorov–Smirnov test (8).

In [3] the authors present a method to use the test in the multivariate case. They explain it is possible considering the multivariate Kolmogorov–Smirnov statistic as distribution free using the theorem due to Rosenblatt. Then, a brief explanation of this method is presented.

Given the previous statistic definition (8), as a result of its distribution free property we can define it as

$$D_n = \sup_{0 \leq u \leq 1} |G_n(u) - u|, \tag{10}$$

being $G_n(u)$ the empirical distribution function of the uniform 0-1 transformed sample $u_i = F_o(x_i)$, for $i = 1, \ldots, n$.

The authors describe that the distribution free property is the result of that any continuos random variable X can be transformed to a uniform random variable Y using $Y = F(X)$, being F the distribution function of X. It can be extended for a continuous multivariate random variable using the Rosenblatt theorem (see detailed [3]).

Consequently, applying this theorem for the two remote sensing images case, the joint density of (9) can be defined as

$$f_{\boldsymbol{X}}(x_1, x_2) = f_1(x_1) \cdot f_2(x_2|x_1), \tag{11}$$

and using the above distribution function transformation

$$Y_1 = F_1(X_1) \tag{12}$$

$$Y_2 = F_2(X_2|X_1), \tag{13}$$

where Y_1 and Y_2 are independent and identically distributed (i.i.d.) uniform 0-1.

Now, the bivariate statistic can be defined as

$$D_n = \sup_{\boldsymbol{y}} |G_n(\boldsymbol{y}) - y_1 \cdot y_2|, \tag{14}$$

being G_n the empirical distribution function of $\boldsymbol{y} = F(\boldsymbol{x})$.

The authors also explain that it is much simpler compute the statistic taking the supremum on the set of transformed sample points (named A) to avoid complexly computation, being also distribution free, so

$$\widehat{D_n} = \sup_{\boldsymbol{y} \in \boldsymbol{A}} |G_n(\boldsymbol{y}) - y_1 \cdot y_2|, \tag{15}$$

Finally, they compute a procedure to obtain the maximum distance in several set of points and they choose the absolute maximum between them (see detailed [3], Section 3).

[2] Distribution free statistic: to compute the statistic is not necessary to know the form or the parameters of the distribution.

The method is used to obtain the cumulative distribution functions ($F_1(X_1)$ and $F_2(X_2|X_1)$) for the remote sensing image of this project is based, firstly, on the bivariate pdf of the random vector $\boldsymbol{X}$ (5) — presented in Section 2.2.2 of this report. In this way, the bivariate Kolomogorov–Smirnov test is applied over the Pearson model data.

Secondly, the marginal pdf for both cases are obtained

$$f_{X_1}(X_1) = \int p_{\boldsymbol{X}}(X_1, X_2) dX_1 \tag{16}$$

$$f_{X_2}(X_2 \mid X_1) = \frac{p_{\boldsymbol{X}}(\boldsymbol{X})}{f_{X_1}(X_1)}, \tag{17}$$

finally, the cumulative distribution functions F_{X_1} and F_{X_2} are obtained from (16) and (17) respectively.

As is explained in [4], to obtain an independent statistic of the ordering chosen, it is necessary to define that the statistic should be the largest difference between empirical and theoretical cumulative distributions considering all possible combinations. For this reason, the procedure explained in [3] to obtain the maximum statistic D_n is computed twice, using each time the cdfs in one of the following two orders:

1. D_{n1}: $F_{X_1}(X_1)$ and $F_{X_2}(X_2 \mid X_1)$
2. D_{n2}: $F_{X_2}(X_2)$ and $F_{X_1}(X_1 \mid X_2)$,

and choosing the maximum

$$D_n = max(D_{n1}, D_{n2}). \tag{18}$$

Given (18), its $p - value$ is obtained to compare with the significance level α at the test to accept or to reject the null hypothesis H_0. The formula applied to get the $p - value$ from D_n is based on $P[D_n > \varepsilon]$ [5].

The bivariate Kolomogorov–Smirnov test results are presented below. The test is applied over several regions, with different sizes and covering homogeneous and heterogeneous areas, using a sliding window ($size = 13{\cdot}13$) through them. Fig. 13 shows the bivariate test results for a field homogeneous area. The red pixels indicate the center pixel of each sliding window through the region where the null hypothesis is accepted ($H_0 = 0$).

Figure 13: Two-dimensional Kolmogorov–Smirnov test for Field homogeneous area ($size = 20{\cdot}30$), at the significance level $\alpha = 0.05$

Figs. 14, 15, 16 and 17 below show the bivariate Kolmogorov–Smirnov test results for a several areas. In these results is easy to see that homogeneous areas have an important number of red pixels (in the Field area significantly more than in the Trees area), i.e. the bivariate test accepts the null hypothesis in most cases, while in the heterogeneous areas the red pixels are almost nonexistent or they belong to homogeneous subareas within the heterogeneous areas. As a consequence, for the bivariate case also, the Pearson system provides an appropriate model only for the homogeneous regions data from optical and SAR images.

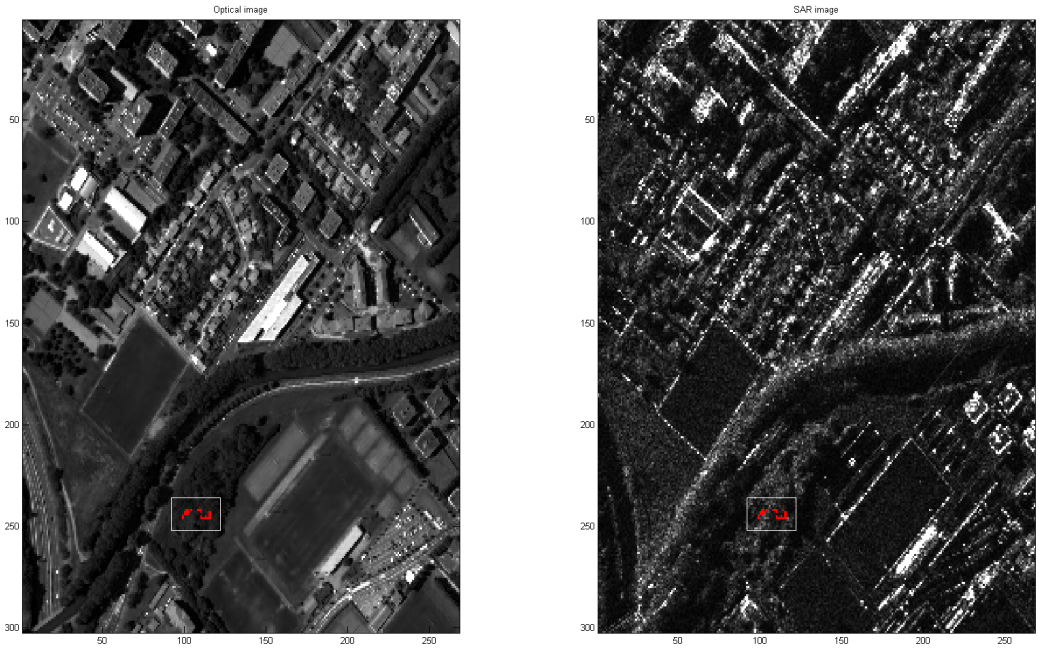


Figure 14: Two-dimensional Kolmogorov–Smirnov test for Trees 1 homogeneous area ($size = 16{\cdot}30$), at the significance level $\alpha = 0.05$

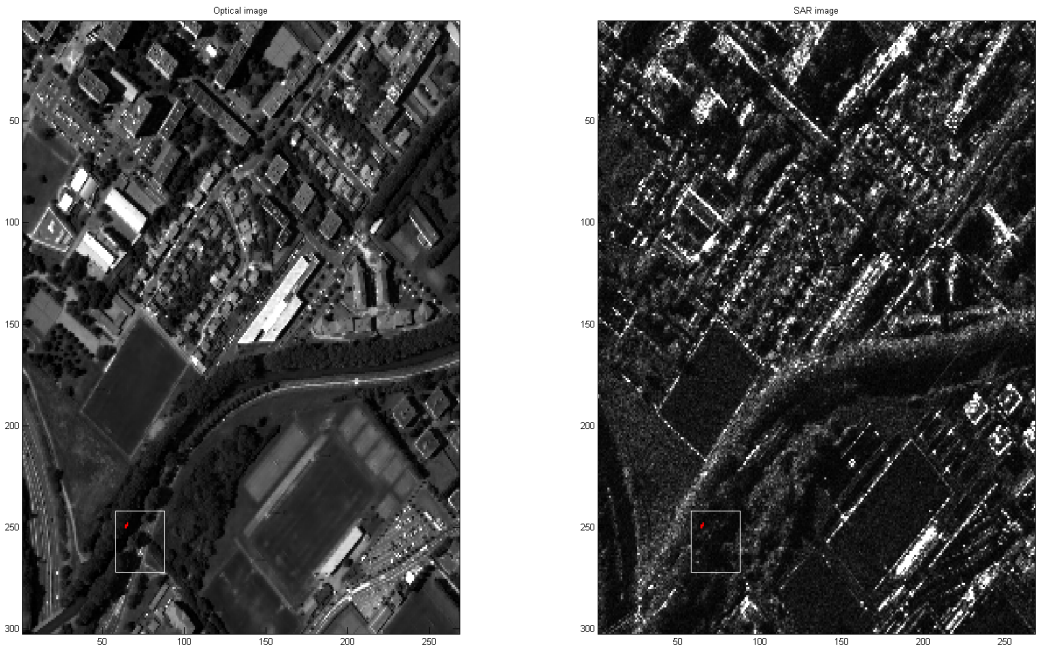


Figure 15: Two-dimensional Kolmogorov–Smirnov test for Trees 2 heterogeneous area ($size = 30{\cdot}30$), at the significance level $\alpha = 0.05$

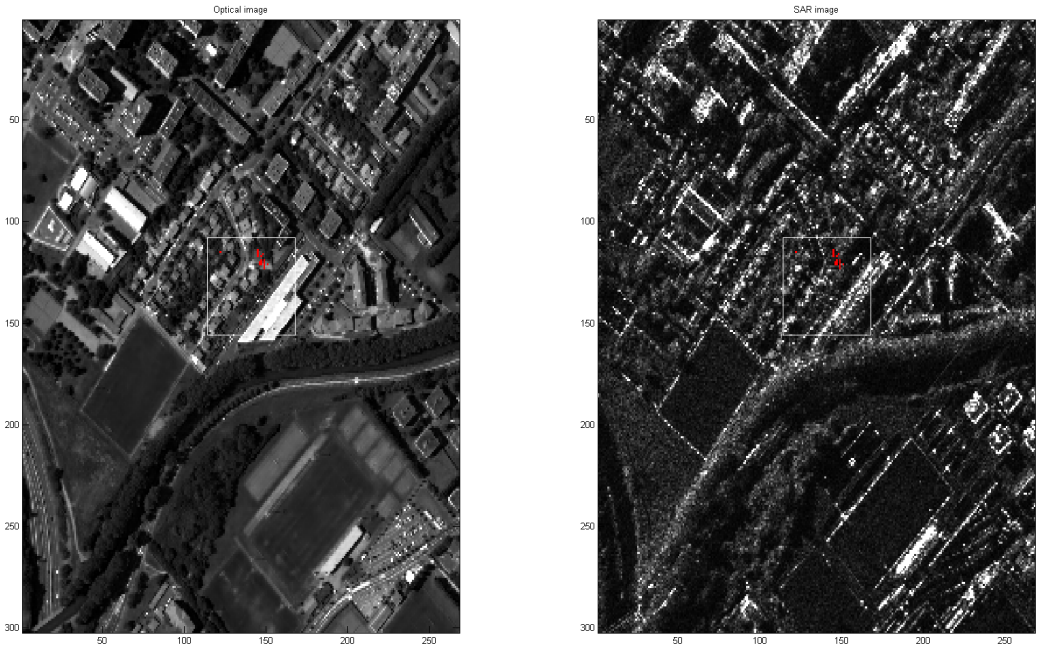


Figure 16: Two-dimensional Kolmogorov–Smirnov test for Buildings 1 heterogeneous area ($size = 48{\cdot}54$), at the significance level $\alpha = 0.05$

Figure 17: Two-dimensional Kolmogorov–Smirnov test for Buildings 2 heterogeneous area ($size = 40{\cdot}40$), at the significance level $\alpha = 0.05$

5 Conclusion

This report collects the study and the results of the adequacy of the bivariate Pearson system for the statistical description of optical and SAR remote sensing images. Based on the results obtained in Section 3 and Section 4, the bivariate Pearson distribution, whose parameters are estimated using the method of moments, can offer an appropriate statistical description for homogeneous areas of optical and SAR images.

6 Acknowledgments

I would like to thank my supervisor, Prof. Marie Chabert, for allow me participate in this interesting project, for her help and her knowledge transfer. Besides, I would like to thank the students of Project Long 2012 of INPT – ENSEEIHT, specially L. Cui, for their programs and J. Prendes for his contributions.

On the other hand, I would also like to thank the UPC – ETSETB, the INPT – ENSEEIHT – IRIT and other institutions which have allowed me to do my Master Thesis (Erasmus modality) in Toulouse. It has been for me an enriching academic and personal experience.

7 References

[1] M. Chabert and J. Y Tourneret. Bivariate Pearson distributions for remote sensing images. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 4038–4041, July 2011.

[2] Y. Nagahara. A method of simulating multivariate nonnormal distributions by the Pearson distribution system and estimation. In *Computational Statistics and Data Analysis, 47*, pages 1–29, 2004.

[3] A. Justel, D. Pea, and R. Zamar. A multivariate Kolmogorov-Smirnov test of goodness of fit. In *Statistics and Probability Letters, 35*, pages 251–259, 1997.

[4] J.A. Peacock. Two-dimensional goodness-of-fit testing in astronomy. In *Mon. Not. R. astr. Soc. 202*, pages 615–627, 1983.

[5] Vladimir N.Vapnik. In *The Nature of Statistical Learning Theory, Springer Ed.*, page 87, 1999.