Analysis and Classification of SAR Textures using Information Theory

Eduarda T. C. Chagas, Alejandro C. Frery, *IEEE Senior Member,* Osvaldo A. Rosso, and Heitor S. Ramos

***Abstract*—We propose a new technique for texture analysis and classification based on the Bandt-Pompe symbolization, and we apply it to SAR data. It consists of (i) linearizing a 2-D patch of the image using the Hilbert-Peano curve, (ii) building an Ordinal Pattern Transition Graph that considers the data amplitude encoded into the weight of the edges; (iii) obtaining a probability distribution function derived from this graph; (iv) computing Information Theory descriptors (Permutation Entropy and Statistical Complexity) from this distribution, and using them as features to feed a classifier. The ordinal pattern graph we propose considers that the weight of the edges is related to the absolute difference of observations, which encodes the information about the data amplitude. This modification takes into account the scattering properties of the target and leads to the characterization of several types of textures. Experiments with data from Munich urban areas, Guatemala forest regions, and Cape Canaveral ocean samples show the effectiveness of our technique, which achieves satisfactory levels of separability. The two descriptors chosen in this work are simple and quick to calculate, and are used as input for a k-nearest neighbor classifier. Experiments show that this technique presents results similar to state-of-the-art techniques that employ a much larger number of features and, consequently, require a higher computational cost.**

***Index Terms*—Synthetic Aperture Radar (SAR), Time-series, Texture, Classification, Ordinal Patterns Transition Graphs.**

I. Introduction

TEXTURE is an elusive trait. When dealing with remotely sensed images, the texture of different patches carries relevant information that, more often than not, is hard to quantify and to transform into useful and parsimonious features. This may be due to the fact that texture, in this context, is a synesthesia phenomenon that triggers tactile responses from visual inputs. This paper presents a new way of extracting features from textures, both natural and resulting from anthropic processes, in SAR (Synthetic Aperture Radar) imagery.

SAR systems are a vital source of data because they provide high-resolution images in almost all weather and day-night conditions. They provide basilar information, complementary to that offered by sensors that operate in other regions of the electromagnetic spectrum, for a variety of Earth Observation applications.

SAR data are rich in information, although with challenging characteristics. Most notably, they do not follow the usual Gaussian additive model, and the signal-to-noise ratio is usually low.

Yue et al [1] provide a comprehensive account of how the physical properties of the target are translated into first- and second-order statistical properties of SAR intensity data.

There is general agreement that non-deterministic textures are encoded in the second-order features, i.e., in the spatial correlation structure. This notion often leads to using the covariance matrix and other measures that assume that a linear dependence, namely the Pearson correlation coefficient, suffices to characterize natural textures.

For the aforementioned reasons, this might not be the case when dealing with SAR imagery. Texture, in these images, is often visible only over large areas, and the multiplicative and non-Gaussian nature of speckle antagonizes with the additive assumption that underlies classical approaches.

SAR textures can be studied following two complementary approaches, namely analyzing the marginal properties of the data (first-order statistics), and observing their spatial structure [1], [2]. In this work, we focus on the second.

The most widely used approach to obtain textural features from SAR imagery is through co-occurrence matrices and Haralick's descriptors [3]. Radford et al. [4] used textural information derived from gray-level co-occurrence matrices, along with Random Forests, for geological mapping of remote and inaccessible localities; the authors obtained a classification accuracy of $\approx 90\,\%$, even when using limited training data ($\approx 0.15\,\%$ of the total data).

Hagensieker and Waske [5] evaluated the synergistic contribution of multi-temporal L-, C-, and X-band data to tropical land cover mapping, comparing classification outcomes of ALOS-2 [6], RADARSAT-2 [7], and TerraSAR-X [8] datasets for a study site in the Brazilian Amazon using a wrapper approach. The wrapper utilizes the gray-level co-occurrence matrix (GLCM) texture information and a Random Forest classifier to estimate scene importance.

Storie [9] proposed an open-source workflow for detecting and delineating the urban-rural boundary using Sentinel-1a SAR data. The author used a combination of GLCM information and a k-means classifier to produce a three-category map that distinguishes urban from rural areas. Other approaches include the Fourier power spectrum [10], and random fields [11].

E. T. C. Chagas and H. S. Ramos ar with Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil (e-mail: eduarda.chagas@dcc.ufmg.br, ramosh@dcc.ufmg.br).

A. C. Frery is with Laboratório de Computação Científica e Análise Numérica – LaCCAN, Universidade Federal de Alagoas, Brasil; (e-mail: acfrery@laccan.ufal.br)

O. A. Rosso is with Instituto de Física, Universidade Federal de Alagoas, Brasil (e-mail: oarosso@if.ufal.br)

Manuscript received XX YY, 20ZZ; revised WW UU, 20VV.

In our approach, we opt to analyze the 1-D signals resulting from the linearization of the image samples, using non-parametric time series analysis techniques. With this approach, we reduce the dimensionality of the data and map the spatial dependence of the pixels onto the temporal correlation of a time series. Next, we extract a probability distribution from the ordinal patterns of a new transition graph, herein proposed, which can discriminate different levels of amplitude, and we use well-known features borrowed from the Information Theory field to classify texture patches.

The main contribution of this work is the proposal of a new set of features for characterization and classification of SAR textures, based on a modification of the current ordinal patterns transition graph, incorporating relevant time-series information. Our proposal, hereinafter referred to as Weighted Amplitude Transition Graph (WATG), incorporates the absolute difference between the observations, weighting them in the calculation of the final value of their probabilities. Experiments performed on SAR image textures indicate that this approach preserves essential information about the dynamics of the system, presenting excellent results in the classification of regions, where it has reached satisfactory levels of separability.

The paper is structured as follows: Section II describes our proposed methodology. Section II-A presents the patch linearization process of the images. In the Section II-B we report the Bandt-Pompe symbolization process. Section II-C describes the ordinal patterns transition graphs. Section II-D shows our technique of ordinal amplitude transition graph weighting by amplitudes. In Section II-E we report the Information Theory descriptors used throughout this work. SectionřefMethods we summarize the weighted ordinal patterns methods used to compare with our proposal. Section IV describe the SAR image datasets, the analysis of ordinal pattern methods, experiments of sliding window selection, and a quantitative assessment. Finally, Section V concludes the paper.

II. Methodology

The methodology applied in this work is shown in Alg. 1. The main idea of the technique is as follows. After receiving the P texture patch, the algorithm linearizes the image using Hilbert-Peano curves, thus acquiring a time series. Then, the procedure calls the `WATG` subroutine over the time series and the ordinal pattern parameters to calculate the probability distribution. Finally, shannon permutation entropy and statistical complexity are calculated. This features obtained in the last step can be used either for characterization or classification.

The `WATG` function basically consists of three steps: (i) we use the embedding dimension D and time delay τ in `BPSymbolization` subroutine for generate the ordinal patterns of the given time series using Bandt-Pompe, (ii) with the symbols in hand, the `transitions` function calculates the sequence of alternations of the ordinal patterns, and in (iii) `weigthGraph` generates the desired incidence matrix of the graph using the weights of the amplitude differences between the time series elements. Finally, the probability distribution obtained by the proposed method is achieved by vectoring the final transition matrix. These steps are also depicted in Fig. 1, and are detailed in the following sections.

Algorithm 1 $H \times C$ point from a patch using WATG

Input: Patch of texture P, dimension D and time delay τ
Output: $H \times C$ feature

```
time.series ← hilbertCurve(P)
Probs ← WATG(time.series, D, τ)
H ← ShannonEntropy(Probs)
C ← StatisticalComplexity(Probs)
function WATG(time.series, D, τ)
    patterns ← BPSymbolization(time.series, D, τ)
    transitions ← transitions(patterns)
    graph ← weigthGraph(time.series, transitions)
    Probs ← as.vector(graph)
    return Probs
end function
```

A. Linearization of image patches

In this step, we perform a data dimensionality reduction by turning the 2-D patch into a times series, i.e., a 1-D signal. This could be accomplished by reading the data by lines, columns, or any transformation of 2-D indexes into a sequence of integers. In this work, we choose to use the Hilbert-Peano curve [12].

Nguyen et al. [13] firstly employed Space-filling curves, to map texture into a one-dimensional signal. When used as scanning methods of an image, such functions preserve relevant properties of pixel spatial correlation [12].

Carincotte et al. [14] used the Hilbert-Peano curve in the problem of change detection in pairs of SAR images. The authors noted that this transformation exploits the spatial locality and that its pseudo-randomness of direction changes works well for a large family of images, especially natural ones.

Assuming an image patch is supported by an $M \times N$ grid, we have the following definition.

Definition 1: An image scan is a bijective function $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ in the ordered pair set $\{(i,j) : 1 \leq i \leq M, 1 \leq j \leq N\}$, which denotes the points in the domain, for the closed range of integers $\{1, \ldots, MN\}$. A scan rule is $\{f^{-1}(1), \ldots, f^{-1}(MN)\}$.
This Definition imposes that each pixel is visited only once, and that all pixels are visited.

Space-filling curves, such as raster-1, raster-2, and Hilbert-Peano scanning techniques, stipulate a proper function f. Traditional Hilbert-Peano curves scan an array of pixels of dimension $2^k \times 2^k$, $k \in \mathbb{N}$, never keeping the same direction for more than three consecutive points, as shown in Fig. 2. Using the Hilbert-Peano curve, we reduce the data dimensionality by maintaining the spatial dependence information of the patch. In this work, we use patches of size 128×128.

B. Bandt-Pompe Symbolization

The representation of time series by ordinal patterns was introduced by Bandt and Pompe [15] as a transformation

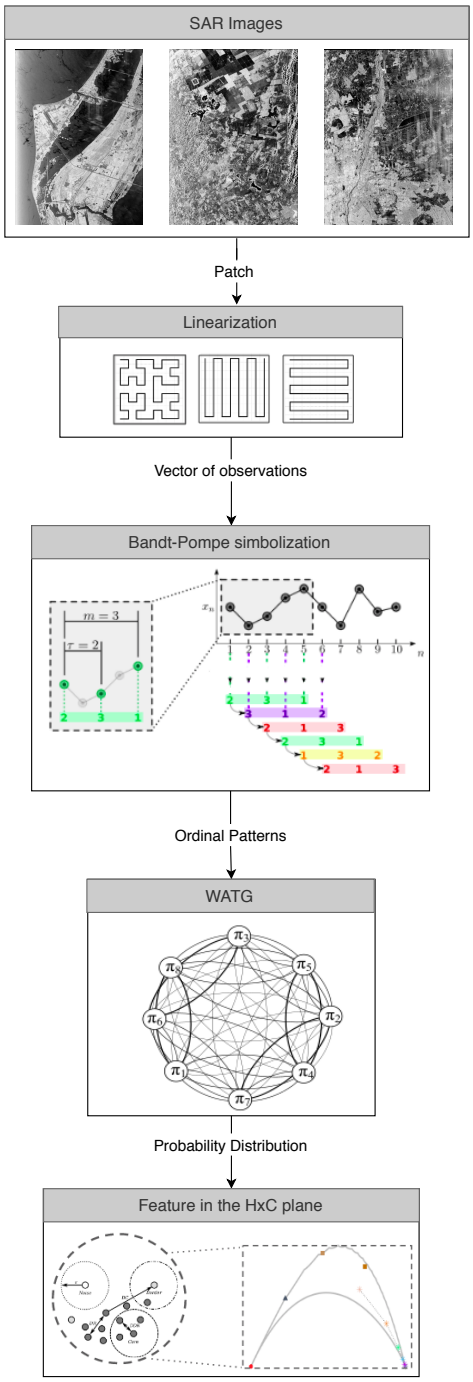

Figure 1. Outline of the methodology used for the classification of textures.

resistant to noise, and invariant to nonlinear monotonic transformations. Therefore, the first step of the `WATG` subroutine is to calculate the ordinal patterns of the time series by Bandt-Pompe symbolization.

Consider $\mathcal{X} \equiv \{x_t\}_{t=1}^{T}$ a real-valued time series of length T.

Let $\mathfrak{A}_D$ (with $D \geq 2$ and $D \in \mathbb{N}$) be the symmetric group of order $D!$ formed by all possible permutation of order D, and the symbol component vector $\boldsymbol{\pi}^{(D)} = (\pi_1, \pi_2, \ldots, \pi_D)$ so every element $\boldsymbol{\pi}^{(D)}$ is unique ($\pi_j \neq \pi_k$ for every $j \neq k$). Consider for the time series $\mathcal{X} \equiv \{x_t\}_{t=1}^{T}$ its time delay embedding representation, with embedding dimension $D \geq 2$ and time delay $\tau \geq 1$ ($\tau \in \mathbb{N}$, also called "embedding time," "time delay", or "delay"):

$$\mathbf{X}_t^{(D,\tau)} = (x_t, x_{t+\tau}, \ldots, x_{t+(D-1)\tau}), \tag{1}$$

for $t = 1, 2, \ldots, N$ with $N = T - (D-1)\tau$. Then the vector $\mathbf{X}_t^{(D,\tau)}$ can be mapped to a symbol vector $\boldsymbol{\pi}_t^D \in \mathfrak{A}_D$. This mapping is such that preserves the desired relation between the elements $x_t \in \mathbf{X}_t^{(D,\tau)}$, and all $t \in \{1, \ldots, T-(D-1)\tau\}$ that share this pattern (also called "motif") have to be mapped to the same $\boldsymbol{\pi}_t^D$.

We define the mapping $\mathbf{X}_t^{(D,\tau)} \mapsto \pi_t^D$ by ordering the observations $x_t \in \mathbf{X}_t^{(D,\tau)}$ in increasing order. Consider the time series $\mathcal{X} = (1.8, 1.2, 3.2, 4.8, 4.2, 4.5, 2.3, 3.7, 1.2, .5)$ depicted in Fig. 3. Assume we are using patterns of length $D = 5$ with unitary time lag $\tau = 1$. The code associated to $\mathbf{X}_3^{(5,1)} = (x_3, \ldots, x_7) = (3.2, 4.8, 4.2, 4.5, 2.3)$, shown in black, is formed by the indexes in $\boldsymbol{\pi}_3^5 = (1, 2, 3, 4, 5)$ which sort the elements of $\mathbf{X}_3^{(5,1)}$ in increasing order: 51342. With this, $\widetilde{\pi}_3^5 = 51342$, and we increase the counting related to this motif in the histogram of all possible patterns of size $D = 5$.

The dash-dot line in Fig. 3 illustrates $\mathbf{X}_1^{(5,2)}$, i.e. the sequence of length $D = 5$ starting at x_1 with lag $\tau = 2$. In this case, $\mathbf{X}_1^{(5,2)} = (1.8, 3.2, 4.2, 2.3, 1.2)$, and the corresponding motif is $\widetilde{\pi}_1^5 = 51423$.

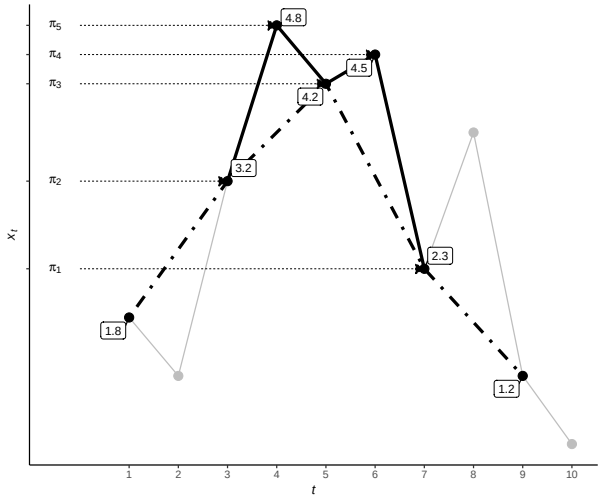

Figure 3. Illustration of the Bandt and Pompe coding

The classic approach to calculating the probability distribution of ordinal patterns is through the frequency histogram.

Denote Π the sequence of symbols obtained by a given series $\mathbf{X}_t^{(D,\tau)}$. The Bandt-Pompe probability distribution is the relative frequency of symbols in the series against the $D!$ possible patterns $\{\widetilde{\pi}_t^D\}_{t=1}^{D!}$:

$$p(\widetilde{\pi}_t^D) = \frac{\#\left\{\mathbf{X}_t^{(D,\tau)} \text{ is of type } \widetilde{\pi}_t^D\right\}}{T-(D-1)\tau}, \tag{2}$$

where $t \in \{1,\ldots,T-(D-1)\tau\}$. These probabilities meet the conditions $p(\widetilde{\pi}_t^D) \geq 0$ and $\sum_{i=1}^{D!} p(\widetilde{\pi}_t^D) = 1$, and are invariant before monotonic transformations of the time series values.

C. Ordinal Pattern Transition Graph

Alternatively, one may form an oriented graph with the transitions from $\widetilde{\pi}_t^D$ to $\widetilde{\pi}_{t+1}^D$. The Ordinal Pattern Transition Graph $G = (V, E)$ represents the transitions between two consecutive ordinal patterns over time t. The vertices are the patterns, and the edges the transitions between them: $V = \{v_{\widetilde{\pi}_t^D}\}$, and $E = \{(v_{\widetilde{\pi}_t^D}, v_{\widetilde{\pi}_{t+1}^D}) : v_{\widetilde{\pi}_t^D}, v_{\widetilde{\pi}_{t+1}^D} \in V\}$ [16].

The literature reports two approaches to compute the weight of edges. Some authors employ unweighted edges [17], [18] which represent only the existence of transitions, while others apply the frequency of transitions [19], [20]. The weights $\mathbb{W} = \{w_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} : v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D} \in V\}$ assigned to each edge describe the chance of transitions between the patterns $(v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D})$ The weights are calculated as the relative frequency of each transition, i.e., as:

$$w_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} = \frac{|\Pi_{\widetilde{\pi}_i^D, \widetilde{\pi}_j^D}|}{T-(D-1)\tau-1}, \tag{3}$$

where $|\Pi_{\widetilde{\pi}_i^D, \widetilde{\pi}_j^D}|$ is the number of transitions from pattern $\widetilde{\pi}_i^D$ to pattern $\widetilde{\pi}_j^D$, $\sum_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} w_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} = 1$, and the denominator is the number of transitions between sequential patterns in the series of motifs of length $T-(D-1)\tau$.

D. Weighted Ordinal Patterns Transition Graph

Our proposal for computing the probability distribution, henceforth referred to as Weighted Amplitude Transition Graph (WATG), takes into account the scattering properties of the target, leading to a good characterization of the textures. We propose a modification of the current ordinal pattern transition graph, incorporating the absolute difference between the observations that produced the patterns.

First, each $\mathcal{X}$ time series is scaled to $[0, 1]$, since we are interested in a metric able to compare datasets:

$$\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \longmapsto x_i, \tag{4}$$

where $x_{\min}$ and $x_{\max}$ are, respectively, the minimum and maximum values of the series. This transformation is relatively stable before contamination, e.g., if instead of $x_{\max}$ we observe $kx_{\max}$ with $k \geq 1$, the relative values are not altered. Nevertheless, other more resistant transformations as, for instance, z scores, might be considered.

Each $\mathbf{X}_t^{(D,\tau)}$ vector is associated with a weight β_t that measures the largest difference between its elements:

$$\beta_t = \max\{x_i - x_j\}, \tag{5}$$

where $x_i, x_j \in \mathbf{X}_t^{(D,\tau)}$.

We propose that the weight assigned to each edge is proportional to the amplitude difference observed in the transition:

$$w_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} = \sum_{i:\{\mathbf{X}_t^{(D,\tau)} \mapsto \widetilde{\pi}_i^D\}} \sum_{j:\{\mathbf{X}_t^{(D,\tau)} \mapsto \widetilde{\pi}_j^D\}} |\beta_i - \beta_j|. \tag{6}$$

Thus, the probability distribution taken from the weighted amplitude transition graph is given as follows:

$$\begin{cases} \lambda_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} = 1, \text{ if } (v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}) \in E, \\ \lambda_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} = 0, \text{ otherwise.} \end{cases}, \text{ and} \tag{7}$$

$$p(\widetilde{\pi}_i^D, \widetilde{\pi}_j^D) = \frac{\lambda_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} \cdot w_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}}}{\sum_{v_{\widetilde{\pi}_a^D}, v_{\widetilde{\pi}_b^D}} w_{v_{\widetilde{\pi}_a^D}, v_{\widetilde{\pi}_b^D}}}. \tag{8}$$

Note that the conditions $p(\widetilde{\pi}_i^D, \widetilde{\pi}_j^D) \geq 0$ and $\sum_{\widetilde{\pi}_i^D, \widetilde{\pi}_j^D} p(\widetilde{\pi}_i^D, \widetilde{\pi}_j^D) = 1$ are satisfied.

Thus, series with uniform amplitudes have edges with probability of occurrence well distributed along the graph, while those with large peaks have edges with probability of occurrence much higher than the others.

E. Information-Theoretic Descriptors

We chose two Information Theory descriptors: Shannon Entropy and Statistical Complexity. Computing these quantities is the last step of the algorithm, i.e., obtaining the point in the $H \times C$ plane.

Entropy measures the disorder or unpredictability of a system characterized by a probability measure $\mathbb{P}$.

Let $\mathbb{P} = \{p_{(\widetilde{\pi}_1^D, \widetilde{\pi}_1^D)}, p_{(\widetilde{\pi}_1^D, \widetilde{\pi}_2^D)}, \ldots, p_{(\widetilde{\pi}_{D!}^D, \widetilde{\pi}_{D!}^D)}\} = \{p_1, \ldots, p_{D!^2}\}$ be the probability function obtained from the time series weighted amplitude transition graph $\mathbb{X}$. Its normalized Shannon entropy is given by:

$$H(\mathbb{P}) = -\frac{1}{2\log D!} \sum_{\ell=1}^{D!^2} p_\ell \log p_\ell. \tag{9}$$

The ability of the entropy to capture system properties is limited, so it is necessary to use it in conjunction with other descriptors to obtain a more complete analysis. Other interesting measures are the distances between $\mathbb{P}$ and a probability measure that describes a non-informative process, typically the uniform distribution.

The Jensen-Shannon distance to the uniform distribution $\mathbb{U} = (\frac{1}{D!^2}, \ldots, \frac{1}{D!^2})$ is a measure of how similar the underlying dynamics is to a non-informative process. It is calculated as:

$$Q'(\mathbb{P}, \mathbb{U}) = \sum_{\ell=1}^{D!^2} \left(p_\ell \log \frac{p_\ell}{u_\ell} + u_\ell \log \frac{u_\ell}{p_\ell} \right). \tag{10}$$

This quantity is also called "disequilibrium." The normalized disequilibrium is $Q = Q'/\max\{Q'\}$.

Conversely to entropy, the statistical complexity seeks to find interaction and dependence structures among the elements of a given series, being an extremely important factor in the study of dynamic systems.

The Statistical Complexity is then defined as [21]:

$$C(\mathbb{P}, \mathbb{U}) = H(\mathbb{P})Q(\mathbb{P}, \mathbb{U}). \tag{11}$$

In our analysis, each time series can then be described by a point $(H(\mathbb{P}), C(\mathbb{P}, \mathbb{U}))$. The set of all pairs $(H(\mathbb{P}), C(\mathbb{P}, \mathbb{U}))$ for any time series described by patterns of length D lies in a compact subset of $\mathbb{R}^2$: the Entropy-Complexity plane. Martín et al [22] obtained explicit expressions for the boundaries of this closed manifold, which depend only on the dimension of the probability space considered, that is, $D!$ for the traditional Bandt-Pompe method and our case $D! \times D!$.

Through such a tool it is possible to discover the nature of the series, determining if it corresponds to a chaotic (or other deterministic dynamics) or stochastic sequences.

III. WEIGHTED ORDINAL PATTERNS METHODS

Recent works proposed using weights in the calculation of relative frequencies for ordinal patterns. They all aim at incorporating the information coded in the amplitude of the observations back into the Permutation Entropy. We summarize in the following those that we used for comparison with our proposal.

A. Weighted Permutation Entropy

The Weighted Permutation Entropy (WPE) was proposed by Fadlallah et al. [23].

Denote $\overline{X}_t^{(D,\tau)}$ the arithmetic mean:

$$\overline{X}_t^{(D,\tau)} = \frac{1}{D} \sum_{k=1}^{D} x_{t+(k-1)}. \tag{12}$$

The weight w_t is the sample variance of each vector $X_t^{(D,\tau)}$:

$$w_t = \frac{1}{D} \sum_{k=1}^{D} \left[x_{t+(k-1)} - \overline{X}_t^{(D,\tau)} \right]^2. \tag{13}$$

Then, the probability distribution is given from the weighted relative frequencies:

$$p(\widetilde{\pi}_t^D) = \frac{\sum_{i:\{\mathbf{X}_i^{(D,\tau)} \mapsto \widetilde{\pi}_t^D\}} w_i}{\sum_{i=1}^{T-(D-1)\tau} w_i}. \tag{14}$$

B. Fine-Grained Permutation Entropy

The Fine-Grained Permutation Entropy (FGPE) was introduced in Ref. [24].

Let β_t be the difference series:

$$\beta_t = \left\{ |x_{t+1} - x_t|, \ldots, |x_{t+(D-1)} - x_{t+(D-2)}| \right\}. \tag{15}$$

The weight w_t quantifies such differences:

$$w_t = \left\lfloor \frac{\max\{\beta_t\}}{\alpha s(\beta_t)} \right\rfloor, \tag{16}$$

where s is the sample standard deviation, α is a user-defined parameter, and $\lfloor \rfloor$ is the floor function. Then, w_t is added as a symbol at the end of the corresponding pattern, leading to an update of Π:

$$\pi'^D_t = \{ \widetilde{\pi}_t^D \cup w_t \}. \tag{17}$$

Finally, the probability distribution is calculated as:

$$p(\pi'^D_t) = \frac{\# \left\{ \mathbf{X}_t^{(D,\tau)} \text{ is of type } \pi'^D_t \right\}}{T - (D-1)\tau}. \tag{18}$$

C. Amplitude-Aware Permutation Entropy

The Amplitude-Aware Permutation Entropy (AAPE) was proposed in Ref. [25]. It consists of weighting the amplitude of ordinal patterns by both the mean and the differences of the elements. For this, only an additional parameter $A \in [0, 1]$ is required, being:

$$w_t = \frac{A \cdot |x_t|}{D} + \sum_{k=1}^{D-1} \left(\frac{A \cdot |x_{t+k}|}{D} + \frac{(1-A) \cdot |x_{t+k} - x_{t+k-1}|}{D-1} \right) \tag{19}$$

The probability distribution is given from the weighted relative frequencies:

$$p(\widetilde{\pi}_t^D) = \frac{\sum_{i:\{\mathbf{X}_i^{(D,\tau)} \mapsto \widetilde{\pi}_t^D\}} w_i}{\sum_{i=1}^{T-(D-1)\tau} w_i}. \tag{20}$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we describe the dataset, the classification process, and the results of the experiments. To assess the performance of the technique here proposed, we first analyze the impact of its parameters and then compare its results in the classification with other methods.

A. Image Dataset

The proposed method was evaluated based on three quad-polarimetric L-band SAR images from the NASA Jet Propulsion Laboratory's (JPL's) uninhabited aerial vehicle synthetic aperture radar (UAVSAR). We used the HH backscatter magnitudes of:

- Sierra del Lacandón National Park, forest region in Guatemala (acquired on April 10, 2015)[1];
- Cape Canaveral Ocean Regions (acquired on September 22, 2016);
- Urban area of the city of Munich, Germany (acquired on June 5, 2015)[2].

We manually selected 160 samples of size 128×128 to compose the dataset used in the experiments. It is organized as follows: 40 samples from Guatemalan forest regions; 80 samples from the oceanic regions of Cape Canaveral, divided into two types with different contrast; and 40 samples of urban regions of the city of Munich. Fig. 4 shows examples of each.

We randomly split the samples in training (85 %) and test (15 %) sets. We used the first set to train a k-nearest neighbor classifier algorithm with tenfold cross-validation. The test samples were not used in the training stage.

B. Analysis of ordinal patterns methods

Fig. 5 shows examples of the ocean, forest and urban samples as sequences values, after the linearization process. Data resulting from remote sensing have a peculiar feature that justifies the application in this article: The variation in the magnitude of the targets' backscatter and, consequently,

[1]https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=Lacand_30202_15043_006_150410_L090_CX_01#dados

[2]https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=munich_19417_15088_002_150605_L090_CX_01#data

in the intensity of the image pixels, depends on the intrinsic properties of the regions under analysis. Urban targets usually exhibit the strongest variation, followed by forests, and finally, water bodies.

By adding information related to amplitude, the proposed method is able to increase, compared to traditional methods, the granularity of information captured by ordinal patterns. Fig. 7 shows the evolution of the descriptive power of these techniques.

The Bandt-Pompe symbolization was the first method based on ordinal patterns proposed in the literature. As shown in Fig. 7 left, it provides a limited separation of the textures. Transition graphs (Fig. 7 center) improve the spread of the features, but with some amount of confusion. Our proposal, shown in Fig. 7 right, produces well-separated features.

In this way, we were able to obtain, for this experiment, a perfect characterization and, consequently, the high descriptive power of the regions.

As already described in Section II-D, our proposal weights the edges in terms of the difference of amplitudes. As expected, the greatest impact is observed on the transition graphs obtained from urban areas.

The urban area time series shown in Fig. 5 has the largest dynamic range. Fig. 6 shows how this information alters the weights of the transition graph. Notice, in particular, that $(v_{\widetilde{\pi}^3_{123}}, v_{\widetilde{\pi}^3_{123}})$ almost doubled, while $(v_{\widetilde{\pi}^3_{312}}, v_{\widetilde{\pi}^3_{231}})$ and $(v_{\widetilde{\pi}^3_{213}}, v_{\widetilde{\pi}^3_{132}})$ became negligible.

We highlight the impact of the weighting on the probability distribution in the two extreme cases observed:

- If the time series presents a low amplitude variation and intensity peaks between, then the transitions of ordinal patterns that represent the latter have larger weights. This contributes so that the probability distribution becomes less uniform among the symbols, since it will be more concentrated in these edges. This will also cause a drop in the Entropy, when compared to the traditional method.
- In time series that show a uniform amplitude variation, the weights are well distributed between their edges, giving rise to a more random probability distribution, thus obtaining larger entropy.

C. Experiments on sliding window selection

In this section, we analyze the parameters of the proposed method and its impact on the classification of textures. McCullough et al. [17] report that inadequate values may hinder important characteristics of the phenomenon under analysis.

The two parameters of the transition graph are the dimension D, and the delay τ. In the experiments below, we present the results in the classification using different values of these parameters.

The performance of the classification method based on ordinal patterns is sensitive to window size, the embedding dimension, and the delay. In techniques based in Bandt-Pompe symbolization, for a fixed time series, as the size of embedding dimension decreases, more ordinal patterns are produced. Therefore, we acquire a greater granularity of information about the dynamics of the system and, consequently, we capture more spatial dependencies between the elements.

We used the ROC curve shown in Fig. 8 for different values of $D \in \{3, 4, 5, 6\}$ and $\tau \in \{1, 2, 3, 4, 5\}$ to select the best configuration.

As we can see in Fig. 8, the configuration that extracted most information from the time series and, thus, that presented the best results in the experiments, is $D = 3$ and $\tau = 1$. The technique, thus, shows its best performance choosing the parameters with the lowest computational cost.

Figure 9 shows the points in the $H \times C$ produced by the same samples with all the parameters mentioned above. The spatial distribution of the points changes with the parameters, and certain configurations promote better separation.

We obtain the lowest Entropy and the largest Statistical Complexity in urban areas with $D = 3$ and $\tau = 1$. The Statistical Complexity discriminates best ocean and forest regions.

Fig. 9 shows the WATG points dimension values D and delays τ. This figure shows that discrimination ability decreases with increasing τ. Larger values of delay dilute the spatial dependence, as neighboring points in the sample tend to be more distant in the image. For this reason, we will use $\tau = 1$.

Considering $\tau = 1$ (first column of Fig. 9), we also notice that $D = 3$ produces the best separation among classes. Increasing D also increases the Statistical Complexity; this is noticeable for the Forest class. The other effect of considering larger values of D is an increased Entropy of Ocean and, with it, an undesirable overlap with Urban samples.

D. Quantitative Evaluation

We present a comparison between our proposal and other methods for characterization and texture classification. We use the following eight methods: Gray-level co-occurrence matrices (GLCM) [26], Gabor filters [27], Bandt-Pompe probability distribution [15], Weighted Permutation Entropy (WPE) [23], Fine-Grained Permutation Entropy (FGPE) [24], Amplitude-Aware Permutation Entropy (AAPE) [25] and Ordinal patterns transition graphs [16]. Table I summarizes the number of features each method produces, as well as its performance.

As in [28], we computed four statistics from co-occurrence matrices: contrast, correlation, energy, and homogeneity. Likewise, we implemented the Gabor filters in five scales and eight orientations; using the Energy, we obtained an 80-dimensional feature vector for each patch.

We used the k-nearest neighbors algorithm for classification with Euclidean distance, selecting the value of k with the automatic grid search method of the Caret R package [29]. In the following, we used $k = 20$. For validation, we used 10-fold cross-validation. More details about the classifier and the sampling can be seen in [30].

Table I presents the results of classifying the 160 samples. We assess the effectiveness of each approach using the following metrics: recall or True Positive Rate (TPR), precision or Positive Predictive Value (PPV), Overall Accuracy (OA), and F1-score.

GLCM produced the worst overall accuracy results: OA $= 66.6\,\%$ and F1-score $= 80.0\,\%$. On the other hand, AAPE (with $\alpha = 0.5$) produced the worst F1-score results: OA $=$

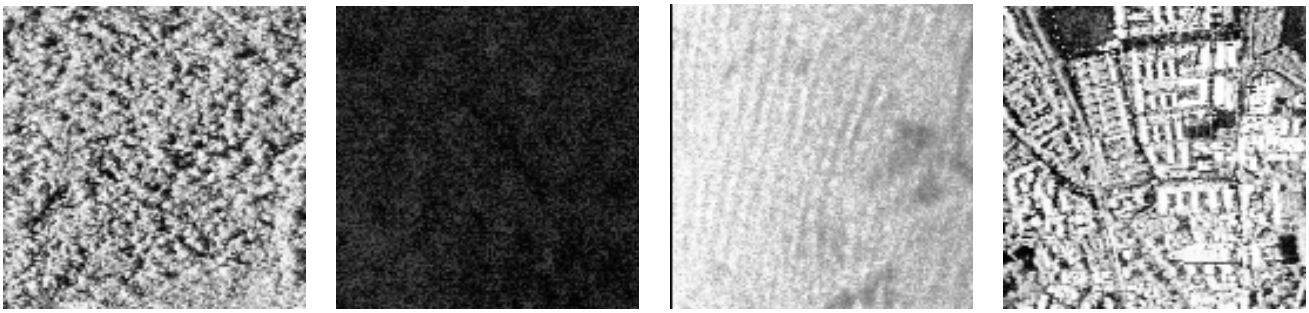

Figure 4. Types of regions analyzed: Guatemala forest, Canaveral ocean type 1, ocean type 2, and Munich urban area

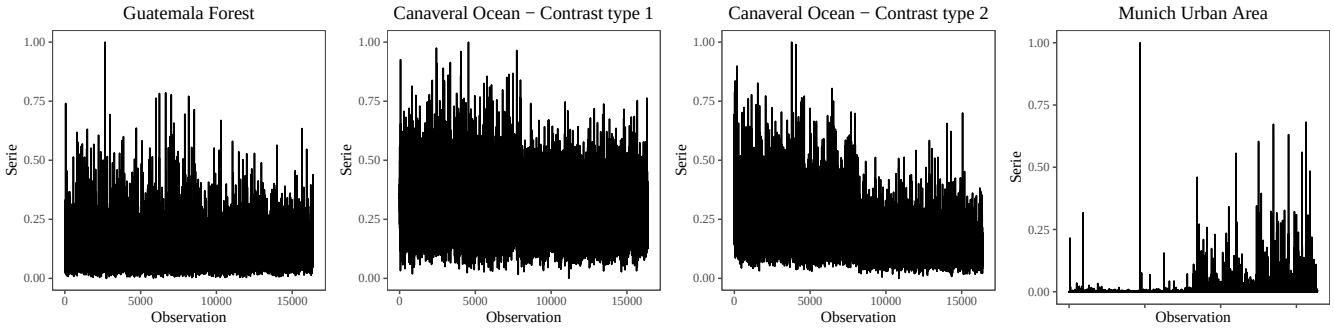

Figure 5. Amplitude of different types of regions

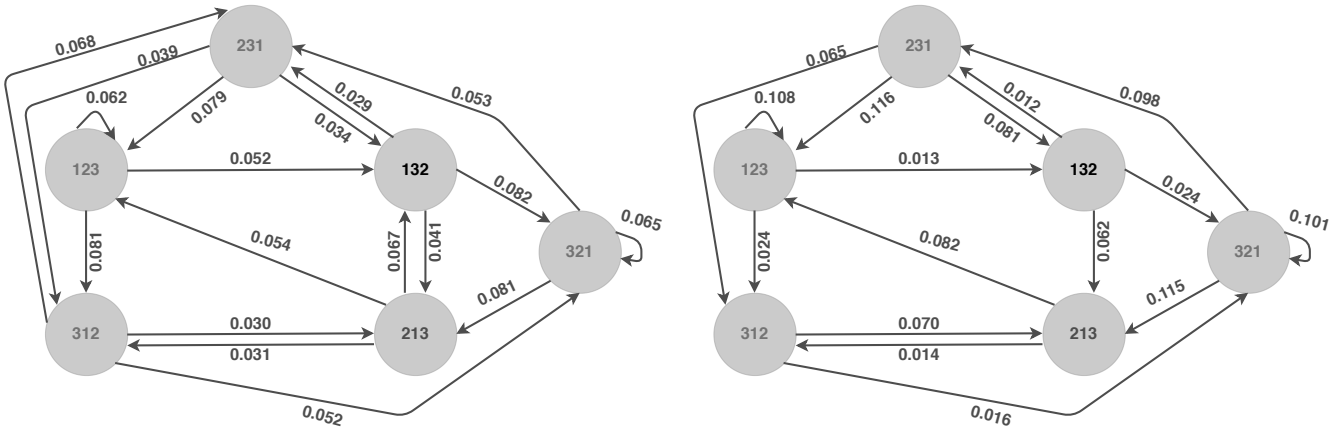

Figure 6. Example of the difference in weights at the edges in a sample of Munich urban regions in the transition graph and in the weighted graph of ordinal patterns transition to dimension 3 and delay 1

$75.0\,\%$ and F1-score $= 40.0\,\%$. Transitions graphs improve these classification results, leading to OA $= 83.3\,\%$ and F1-score $= 60.0\,\%$. Although better (OA $= 95.8\,\%$ and F1-score $= 90.9\,\%$), FGPE (with $A = 1$) is not able to have high descriptive power in the proposed dataset. Both Gabor filters and the proposed method have the highest success rate OA $= 100\,\%$ and F1-score $= 100\,\%$. Thus, WATG attains this performance with 2 features instead of the 80 employed by Gabor filters. This reduction implies less computational power requirement and avoids the curse of dimensionality [31].

V. CONCLUSION

We presented and assessed a new method of analysis and classification of SAR image textures. This method consists of three steps: (1) linearization, (2) computing the Weighted Ordinal Pattern Transition Graph, and (3) obtaining Information Theory descriptors.

The method consists in finding a probability distribution function of ordinal patterns which is sensitive to amplitude information. We then applied the k-NN algorithm to classify the descriptors.

Experiments using patches from UAVSAR images showed that the proposal performs better than GLCM, Bandt-Pompe, Transition Graphs, and other techniques of amplitude information analysis in ordinal patterns and that it provides the same quality of results obtained with Gabor filters. However, while Gabor filters employ 80 features, our proposal requires only two.

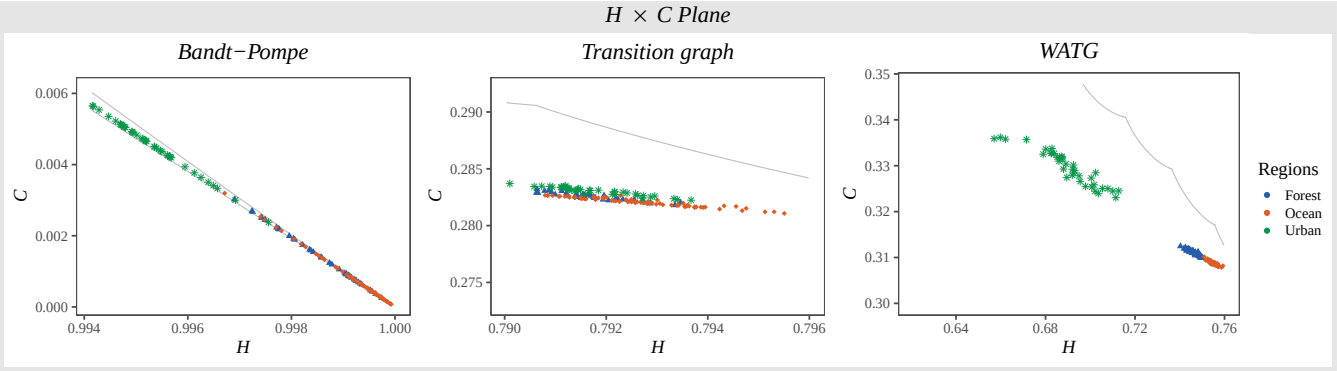

Figure 7. Location of Guatemala (forest), Cape Canaveral (ocean), and Munich (urban) descriptors in the $H \times C$ plane for $D = 3$ and $\tau = 1$. The continuous curves are the maximum and minimum values of C as a function of H.

Table I. Experimental results using k-NN

Method	# features	TPR			PPV			OA	F1-Score
		Urban	Forest	Ocean	Urban	Forest	Ocean		
GLCM	4	1.000	0.833	1.000	1.000	1.000	0.923	0.666	0.800
Gabor	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Bandt-Pompe	2	0.833	0.666	0.833	1.000	0.571	0.833	0.792	0.615
Transition Graph	2	1.000	0.666	0.916	0.857	1.000	0.846	0.875	0.800
WPE	2	0.833	0.666	0.666	0.833	0.500	0.800	0.708	0.571
FGPE (α = 1)	2	1.000	0.833	0.833	1.000	0.714	0.909	0.875	0.769
FGPE (α = 0.5)	2	1.000	0.833	1.000	1.000	1.000	0.923	0.958	0.909
AAPE (A = 1)	2	1.000	0.500	0.750	1.000	0.500	0.750	0.750	0.500
AAPE (A = 0.5)	2	1.000	0.333	0.833	1.000	0.500	0.714	0.750	0.400
WATG	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

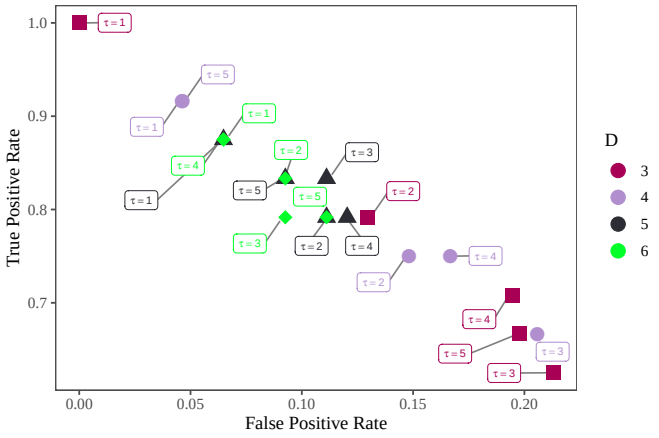

Figure 8. Evaluation of the sliding window parameters using ROC curve

As a result, in addition to perfectly separating urban areas from the others analyzed by entropy values, we are still able to differentiate oceanic and forest areas through their different values of statistical complexity, which informs us of the degree of spatial dependence between their elements.

Since the proposed method has a lower number of descriptive features than its direct competitor (Gabor filters), we were able to obtain some additional advantages. First, by reducing the dimension of the features from 8 to 2, we can accurately visualize the differences between the classes of regions analyzed. In addition, for machine learning algorithms, the smaller the number of dimensions, the faster the process is, and the less storage space is required. Finally, we managed to avoid overfitting, a recurring problem in data with high dimensionality.

VI. SOURCE CODE AVAILABILITY

The text, source code, and data used in this study are available at the *SAR-WATG* repository https://github.com/EduardaChagas/SAR-WATG.

REFERENCES

[1] D.-X. Yue, F. Xu, A. C. Frery, and Y.-Q. Jin, "A generalized Gaussian coherent scatterer model for correlated SAR texture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2947–2964, Apr. 2020.

[2] F. N. Numbisi, F. Van Coillie, and R. De Wulf, "Multi-date Sentinel SAR image textures discriminate perennial agroforests in a tropical forest-savannah transition landscape," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 42, no. 1, 2018.

[3] Q. Yu, M. Xing, X. Liu, L. Wang, K. Luo, and X. Quan, "Detection of land use type using multitemporal SAR images," in *IGARSS 2019 – IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 1534–1537.

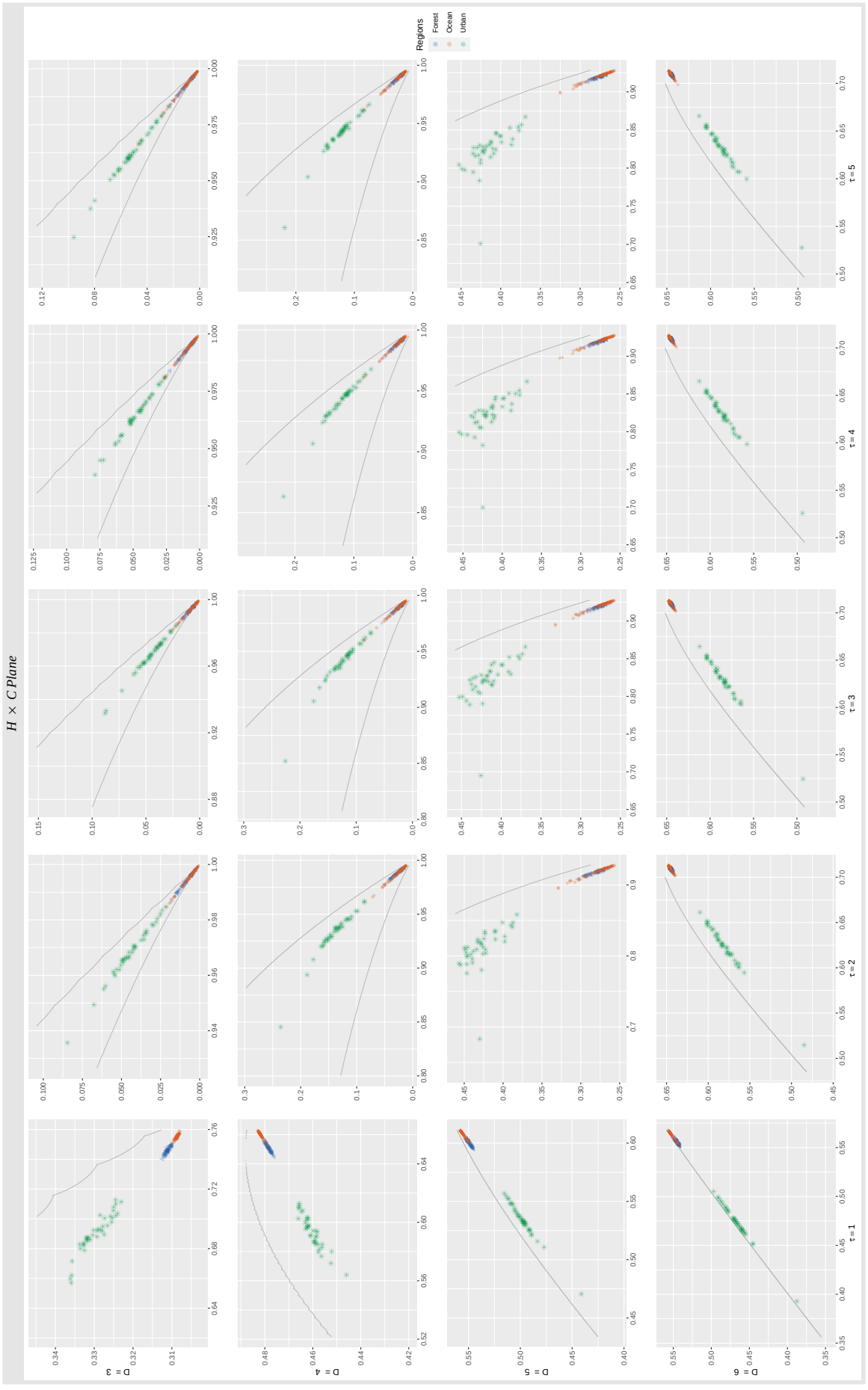

Figure 9. Characterization resulting in $H \times C$ Plane from the application of the Hilbert-Peano curve in WATG on textures of different regions: Guatemala (forest), Cape Canaveral (ocean) and Munich (urban). The continuous curves correspond to the maximum and minimum values of C as a function of H.

[4] D. D. Radford, M. J. Cracknell, M. J. Roach, and G. V. Cumming, "Geological mapping in Western Tasmania using radar and random forests," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3075–3087, 2018.
[5] R. Hagensieker and B. Waske, "Evaluation of multi-frequency SAR images for tropical land cover mapping," *Remote Sensing*, vol. 10, no. 2, p. 257, 2018.
[6] Y. Kankaku, S. Suzuki, and Y. Osawa, "Alos-2 mission and development status," in *IGARSS 2013 – IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2013, pp. 2396–2399.
[7] L. Morena, K. James, and J. Beck, "An introduction to the RADARSAT-2 mission," *Canadian Journal of Remote Sensing*, vol. 30, no. 3, pp. 221–234, 2004.
[8] H. Breit, T. Fritz, U. Balss, M. Lachaise, A. Niedermeier, and M. Vonavka, "TerraSAR-X SAR processing and products," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 727–740, 2009.
[9] C. D. Storie, "Urban boundary mapping using Sentinel-1A SAR data," in *IGARSS 2018 – IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 2960–2963.
[10] J. B. Florindo and O. M. Bruno, "Fractal descriptors based on Fourier spectrum applied to texture analysis," *Physica A*, vol. 391, no. 20, pp. 4909–4922, 2012.
[11] T. Zhu, F. Li, G. Heygster, and S. Zhang, "Antarctic sea-ice classification based on conditional random fields from RADARSAT-2 dual-polarization satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2451–2467, 2016.
[12] J.-H. Lee and Y.-C. Hsueh, "Texture classification method using multiple space filling curves," *Pattern Recognit. Lett.*, vol. 15, no. 12, pp. 1241–1244, 1994.
[13] P. Nguyen and J. Quinqueton, "Space filling curves and texture analysis," in *IEEE Intl. Conf. Pattern Recognition*, 1982, pp. 282–285.
[14] C. Carincotte, S. Derrode, and S. Bourennane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 432–441, Feb 2006.
[15] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, p. 174102, 2002.
[16] J. Borges, H. Ramos, R. Mini, O. A. Rosso, A. C. Frery, and A. A. F. Loureiro, "Learning and distinguishing time series dynamics via ordinal patterns transition graphs," *Appl. Math. Comput.*, vol. 362, p. UNSP 124554, 2019.
[17] M. McCullough, M. Small, T. Stemler, and H. H.-C. Iu, "Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 5, p. 053101, 2015.
[18] C. W. Kulp, J. M. Chobot, H. R. Freitas, and G. D. Sprechini, "Using ordinal partition transition networks to analyze ECG data," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, no. 7, p. 073114, 2016.
[19] T. Sorrentino, C. Quintero-Quiroz, A. Aragoneses, M. Torrent, and C. Masoller, "Effects of periodic forcing on the temporally correlated spikes of a semiconductor laser with feedback," *Opt. Express*, vol. 23, 2015.
[20] J. Zhang, J. Zhou, M. Tang, H. Guo, M. Small, and Y. Zou, "Constructing ordinal partition transition networks from multivariate time series," in *Scientific Reports*, 2017.
[21] P. W. Lamberti, M. T. Martín, A. Plastino, and O. A. Rosso, "Intensive entropic non-triviality measure," *Physica A*, vol. 334, no. 1–2, pp. 119–131, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437103010963
[22] M. T. Martín, A. Plastino, and O. A. Rosso, "Generalized statistical complexity measures: Geometrical and analytical properties," *Physica A*, vol. 369, no. 2, pp. 439–462, 2006.
[23] B. H. Fadlallah, B. Chen, A. Keil, and J. C. Príncipe, "Weighted-permutation entropy: a complexity measure for time series incorporating amplitude information." *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, vol. 87 2, p. 022911, 2013.
[24] L. Xiao-Feng and W. Yue, "Fine-grained permutation entropy as a measure of natural complexity for time series," *Chin. Phys. B*, vol. 18, no. 7, p. 2690, 2009.
[25] H. Azami and J. Escudero, "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation," *Comput. Methods Programs Biomed.*, vol. 128, pp. 40–51, 2016.
[26] A. Kourgli, M. Ouarzeddine, Y. Oukil, and A. Belhadj-Aissa, "Texture modelling for land cover classification of fully polarimetric SAR images," *International Journal of Image and Data Fusion*, vol. 3, no. 2, pp. 129–148, 2012.
[27] T. P. Weldon, W. E. Higgins, and D. F. Dunn, "Efficient Gabor filter design for texture segmentation," *Pattern Recognit.*, vol. 29, no. 12, pp. 2005–2015, 1996.
[28] D. Guan, D. Xiang, X. Tang, L. Wang, and G. Kuang, "Covariance of textural features: A new feature descriptor for SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3932–3942, 2019.
[29] M. Kuhn, "Building predictive models in R using the caret package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008.
[30] T. M. Mitchell, *Machine Learning*. McGraw-hill New York, 1997.
[31] N. Altman and M. Krzywinski, "The curse(s) of dimensionality," *Nat. Methods*, vol. 15, no. 6, pp. 399–400, May 2018.

ACKNOWLEDGEMENTS

This work was partially funded by the Coordination for the Improvement of Higher Education Personnel (CAPES) and National Council for Scientific and Technological Development (CNPq).