METHODOLOGY ARTICLE

Urban structure type mapping method using spatial metrics and remote sensing imagery classification

Luccas Z. Maselli[1] · **Rogério G. Negri[2]**

Abstract

Urban Structure Types (USTs) stand for areas with homogeneous appearance over the urban matrix. The use of spatial metrics rises as a convenient alternative to quantify the homogeneity of areas on a specific scale. Remote sensing imagery is largely used to assess and study the urban environment, and its classification is a way to recreate the Earth's surface digitally, both natural and urban spaces. This study proposes a method for city-scale UST mapping using remote sensing images as the unique source of information. Such a proposal comprehends the classification of images that express spatial metrics derived from previous land use and land cover (LULC) classification. We carried two case studies to assess the proposed method under different image resolutions and urban complexity conditions. For this purpose, Landsat-8 OLI and Sentinel-2 MSI images acquired from different cities in Brazil are submitted to the proposed method. An alternative object-based image classification method is included as a comparison baseline. The proposed method shows efficiency in the UST mapping process, which is highly influenced by the neighborhood size considered over the process. Also, it is verified that the proposed method is superior at a significance level of 5%.

Keywords Urban structure types · Spatial metrics · Image classification · Remote sensing · Urban mapping

Introduction

The urban environment is susceptible to changes in spatial dynamics, promoting impacts on the population and urban design. Such changes make this perceptible through its spatial growing patterns as well as the reshaping of its structure (Aljoufie et al. 2013). Furthermore, natural resources have been increasingly used as a consequence of urban growth, generating progressive environmental degradation. Therefore, methodologies for the study of urban space and

expansion dynamics are necessary, so that city planning becomes more assertive, efficient, and rapid (Pham et al. 2011).

Once simulating, foreseeing, and recreating the urban environment digitally is made possible, decisions about the real world can be taken in a digital sphere. With this purpose, remote sensing images have been used to assess urban settlements and population dynamics in various scales (Tomás et al. 2016). Studies that exemplify the digital representation of the urban environment were developed in Germany (Banzhaf and Hofer 2008) and Chile (Banzhaf et al. 2009), which analyzed Urban Structure Types (USTs) premised on the spatial distribution of land use and land cover (LULC) types. The UST concept is based on the subdivision of an area into minimal significant structures that has homogeneous appearance in the urban matrix and contains both built and open spaces (Böhm 1998).

According to Montanges et al. (2015), a UST is different from a LULC as it does not study specific objects such as vegetation, roofs, and pavements, but the spatial morphology on a specific scale. Also, the Local Climate Zones (LCZs), introduced by (Stewart and Oke 2012),

Communicated by: H. Babaie

✉ Luccas Z. Maselli
    maselliluccas@gmail.com

    Rogério G. Negri
    rogerio.negri@ict.unesp.br

[1] Federal University of São Carlos (UFSCar) - São Carlos, São Paulo, Brazil

[2] São Paulo State University (UNESP) - São José dos Campos, São Paulo, Brazil

differs from the UST since LCZs, applied in climate studies, are "regions of uniform surface cover, structure, material, and human activity" (Stewart and Oke 2012), while USTs deal only with the morphology of the urban space.

Previous studies showed the integration of three-dimensional and vector data with high-resolution images for classification into USTs (Berger et al. 2018) as well as the use of building geometries and its spatial distribution for UST characterization (Novack and Stilla 2017). Such studies make evident a tendency regarding the use of multiple data sources, different platforms, and low-automated methods.

An alternative for automating the process may be achieved through remote sensing image classification. This kind of application is frequently used to map the Earth's surface areas into different classes of interest (Mather 2004). UST mapping supported by image classification techniques is a topic that has already been addressed in previous studies (Wieland et al. 2016; Tam et al. 2018; Simanjuntak and Reckien 2019).

The use of spatial metrics combined with image classification can be a potential methodology for UST characterization. Spatial metrics are measures derived from maps that exhibit spatial heterogeneity on a particular scale (Herold et al. 2005). Some examples of spatial metrics are the *patch cover percentage*, *coefficient of variation of patch areas*, *patch density*, and *edge density*, calculated over each considered LULC classes (Herold et al. 2002; Herold et al. 2003). Among different alternatives, the use of images of spatial metrics derived from prior classification results rises as a convenient way to characterize USTs.

This study introduces a city-scale UST mapping method based on the concepts of spatial metrics and image classification with a Support Vector Machine (SVM). In contrast to previous studies, the proposed methodology adopts remote sensing imagery as its unique information source. This means that no additional data – such as vector data, spatial models, or even exchanges through processing platforms – are needed, allowing greater automation in the UST mapping process.

To assess our method and compare it against an alternative methodology, based on Wieland et al. (2016) and using the Random Forest (RF) classifier, we study two cases of UST mapping. These study cases are carried in urban areas of São José dos Campos and São Paulo cities, Brazil. For this, we employed images acquired from different satellites: Landsat-8 OLI and Sentinel-2 MSI.

This paper is organized as follows. Section "Theoretical background" presents the fundamental concepts regarding image classification, UST, and spatial metrics; "UST mapping framework based on spatial metrics classification" introduces the proposed method; the study cases and comparisons with alternative methods are presented in "Experiments"; and lastly, "Conclusions" summarizes the findings of this paper.

Theoretical background

A brief discussion on image classification

Remote sensing image classification has attracted the scientific community's attention as the derived results of this application prove to be useful in socioeconomic and environmental studies. Consequently, the development of more accurate classification methods is a constant challenge (Lu and Weng 2007).

Formally, a classifier is represented by a function $F : \mathcal{X} \to \mathcal{Y}$ that assigns elements from the attribute space $\mathcal{X}$ to a class in $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$, $c \in \mathbb{N}^*$, with class labels in $\mathcal{Y} = \{1, 2, \dots, c\}$. Under these conditions, for $\mathbf{x} \in \mathcal{X}$ and $y \in \mathcal{Y}$, $y = F(\mathbf{x})$ means that $\mathbf{x}$ corresponds to the class ω_y.

Considering $\mathcal{I}$ as an image defined on a support lattice $\mathcal{S} \subset \mathbb{N}^2$, the image classification consists of the application of F on the attribute vector $\mathbf{x} \in \mathcal{X}$ associated with a pixel $s \in \mathcal{S}$ of $\mathcal{I}$. By consequence, one can write $\mathcal{I}(s) = \mathbf{x}$ as a way to denote that the pixel s from $\mathcal{I}$ has attribute vectors $\mathbf{x}$, and $\mathcal{C}(s) = \omega_y$ means that s was associated with the class ω_y since $F(\mathbf{x}) = y$.

Different image classification methods proposed in the literature are distinct ways to model $F : \mathcal{X} \to \mathcal{Y}$ and apply it to classify $\mathcal{I}$. Supervised and unsupervised learning are examples of approaches for modeling F. The supervised approach uses available information in a training set $\mathcal{D} = \{(\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} : i = 1, 2, \dots, m\}$ composed by $m \in \mathbb{N}^*$ vectors whose associated classes are known.

Among several supervised classification methods, the SVM has received considerable attention given its solid theoretical foundation and notable characteristics, such as simple architecture, moderate computational complexity, and great generalization capability (Bruzzone and Persello 2009). According to Mountrakis et al. (2011), the SVM method has provided comparable and frequently better results concerning other classification methods.

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} : i = 1, 2, \dots, m\}$ a training set, with $\mathcal{Y} = \{+1, -1\}$, where $\mathbf{x}_i$ is assigned to ω_1 when $y_i = +1$, or to ω_2 when $y_i = -1$. The SVM method distinguishes ω_1 from ω_2 through the following largest margin discriminating function:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \tag{1}$$

where $\mathbf{w}$ represents an orthogonal vector to the hyperplane $f(\mathbf{x}) = 0$ and b is a scalar such that $|b| / \|\mathbf{w}\|$ express the distance between the hyperplane and the origin of the

attribute space. The notations $|\cdot|$, $\|\cdot\|$ and $\langle\cdot\rangle$ stands for the absolute value, vector norm, and inner product. The values for $\mathbf{w}$ and b are obtained by solving the following optimization problem (Theodoridis and Koutroumbas 2008):

$$\max_{\lambda} \left(\sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right)$$

$$\text{subjected to:} \begin{cases} 0 \leq \lambda_i \leq C, i = 1, \ldots, m \\ \sum_{i=1}^{m} \lambda_i y_i = 0 \end{cases} \quad (2)$$

where λ_i are Lagrange multipliers, and C is a parameter insert to deal with non-separable classes, acting as a misclassification penalty during the training stage.

The classification performance of the SVM method can be improved by embedding the input patterns into a more appropriate feature space with better separability. Kernel functions that substitute the inner product at Eq. 2 may be adopted for this purpose (Webb and Copsey 2011). The most usual kernel functions are:

Linear: $\quad K(x, y) = \langle x, y \rangle$
Polynomial: $\quad K(x, y) = (1 + \langle x, y \rangle)^p$
Radial Basis: $\quad K(x, y) = \exp\left(-\gamma \|x - y\|^2\right)$

where $p \in \mathbb{N}^*$ and $\gamma \in \mathbb{R}_+^*$ are parameters for polynomial and Radial Basis Function (RBF) kernel functions, respectively.

Moreover, accordingly to the previous formulation, the SVM is able to distinguish only two classes. In order to extend its application for non-binary classification problems it is adopted a multiclass strategy. Usually, such strategies comprehends a decomposition of the original problem into several binary sub-problems. Posteriorly, the results of each sub-problem are then combined as a multiclass classification result. "One-Against-All" (OAA) and "One-Against-One" (OAO) are examples of multiclass strategies based on binary decomposition (Webb 2002).

Introduced by Breiman (2001), the RF method is another example of a classifier frequently employed in recent remote sensing studies. The RF exploits the *ensemble learning* technique, combining the output of multiple decision trees through a major voting process, and producing a classification decision (Ananias and Negri 2021).

From a training set $\mathcal{D}$, several replications with the same cardinality of $\mathcal{D}$ are taken by bootstrapping process. Then, a decision tree is trained through each replica. The RF parameters, like the *maximum depth of trees*, *minimum number of samples in each node* to split, a *maximum number of trees* and *out-of-bag error* should be tuned before the training process. More details and discussions regarding those parameters are found in Breiman (2001).

Concerning the RF classification process, a vector $\mathbf{x}$ is assigned to a class in Ω that produces significant concordance among all individual trees. According to Belgiu and Drăguţ (2016), the RF method is a computationally efficient algorithm that does not overfit the final decision rule.

Urban structure types

USTs aim to describe land use arrangements in urban areas (Lehner and Blaschke 2019). Such a concept is sustained by the principle that cities are composed of several morphological elements, having an intrinsic metabolism with well-defined social and environmental patterns according to its activities and arrangements of build and open spaces (Pauleit and Duhme 2000). Furthermore, Hecht et al. (2013) states that USTs are determined as functions of buildings' predominance types and their patterns of spatial distribution.

As such, the UST rises as a convenient basis for effective urban-environmental planning. It allows us to recognize urban settlement groups with similar physic characteristics, which are essential information to define the urban development guidelines (Moon et al. 2009). Given a generalization scale, USTs consist of the aggregation of isolated objects inside the urban space on a block level, that is, concerning the elements into a spatial neighborhood. The LULC is the most generalist level for a city scale, and the *structural elements* the less generalist level, which is related to the building scale (Fig. 1).

Spatial metrics

Spatial metrics stand for measures derived from digital maps to quantify spatial heterogeneity at a specific scale and resolution (Herold et al. 2003). Such measures yield quantitative characterizations about spatial composition, habitat configuration, and land use. Moreover, spatial metrics
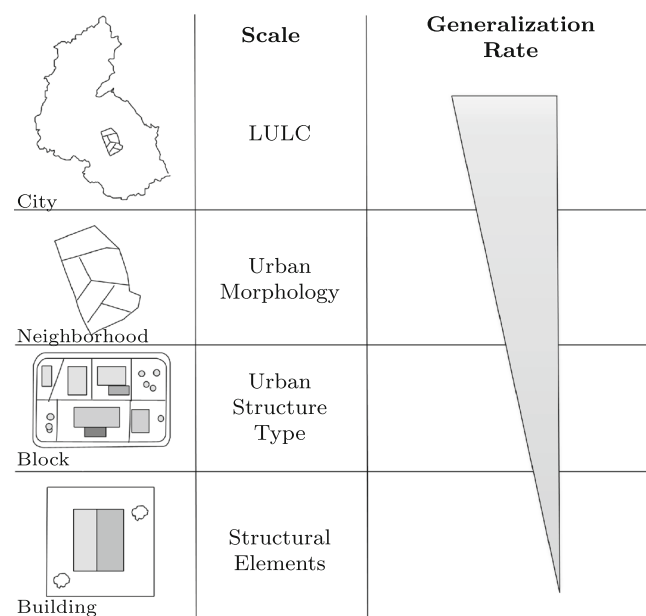

Fig. 1 Urban structure analysis according to the spatial scale. Adapted from Banzhaf and Hofer (2008)

on remote sensing data allow the generation of consistent and detailed information about the urban structure (Deng et al. 2009).

Among a plethora of proposals, four examples of spatial metrics that can be derived from remote sensing image classification are the following: *patch cover percentage*, *coefficient of variation of the patch areas*, *patch density* and *edge density of the patch*. Formalizations of such metrics as well as their components are presented to allow future methodological reproductions and applications of the proposed method.

Initially, we should define the spatial neighborhood concept:

$$\mathcal{V}_\rho(s) = \{s \in \mathcal{S} : d(s, t) < \rho; t \in \mathcal{S}\}, \tag{3}$$

where $d(\cdot, \cdot)$ is the maximum distance, which is $d(a, b) = \max\{|a_1 - b_1|, |a_2 - b_2|\}$, being $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$ elements from $\mathcal{S}$, and $|\cdot|$ the absolute value. ρ represents the neighborhood influence radius for s.

Once the spatial neighborhood is established, we define a patch as every set of spatially connected positions of a common class. Formally, for each position s and a given neighborhood influence radius ρ, a ω_y class patch is represented by the following:

$$M_j^{(y)}(s, \rho) = \{t \in \mathcal{V}_\rho(s) : \mathcal{C}(t) = \omega_y, \mathcal{C}(t) \\ = \mathcal{C}(r), \|t - r\|_2 \le 1\}. \tag{4}$$

where $\|\cdot\|_2$ is the Euclidean norm.

The *patch cover percentage* metric expresses the proportion of ω_y class areas in relation to the total area, given by the following:

$$P_y = \frac{A_y}{A}, \tag{5}$$

where $A_y = \# \bigcup_{j=1}^{m_y} M_j^{(y)}(s, \rho)$ is the area of the patches associated with the ω_y class accordingly to the amount of pixels related to this class, and $A = \# \bigcup_{k=1}^{c} \bigcup_{j=1}^{m_k} M_j^{(k)}(s, \rho)$ is the sum of the areas of all patches. Also, m_k is the number of patches of a certain class $\omega_k \in \Omega$.

The *coefficient of variation of the patch areas* expresses the percentage of variation of the areas concerning ω_y, which is the following:

$$CV_y(s, \rho) = \frac{\sigma\left(M_j^{(y)}(s, \rho)\right)}{\mu\left(M_j^{(y)}(s, \rho)\right)}; \ j = 1, 2, \ldots, m_y, \tag{6}$$

where, for ω_y and the neighborhood $\mathcal{V}_\rho(s)$, $\sigma\left(M_j^{(y)}(s, \rho)\right)$ and $\mu\left(M_j^{(y)}(s, \rho)\right)$ represent the standard deviation and the average area of the patches, respectively.

The *patch density* of the ω_y class quantifies the proportion between the number of ω_y patches and the area of all patches, given by the following:

$$D_y = \frac{m_y}{A}, \tag{7}$$

Lastly, the *edge density of the patch* regarding ω_y is the proportion between the length of edges for patches of class ω_y in relation to the area of all patches:

$$B_y = \frac{\sum_{j=1}^{m_y} b_j^{(y)}(s, \rho)}{A}, \tag{8}$$

where $b_j^{(y)}$ is the perimeter of a patch $M_j^{(y)}(s, \rho)$.

UST mapping framework based on spatial metrics classification

Figure 2 depicts the flowchart of the proposed UST mapping method. From an image with sufficient spatial resolution to identify the objects of interest, and a set of LULC samples collected over the study area (adequately partitioned between training and testing), an image classification process is carried out. To train the classification method, point-wise samples are further indicated to reduce the risk of defined samples with mixed information from multiple classes, once the imagery resolution usually does not allow a polygonal sample collection over small areas. We named the output result of this stage as "primary classification". The SVM method is used for this purpose, and different parameter configurations should be assessed to achieve the most accurate result.

Regarding the primary classification accuracy assessment, point-wise test samples are also indicated because the classified image remains on the same scale as the original input image, and polygonal samples may encompass more than one class.

Afterward, the obtained primary classification is submitted to the spatial metrics calculation. More precisely, the Eqs. 5 to 8 are applied on each pixel of the primary classification under a fixed spatial neighborhood of radius ρ. It is important to highlight that for a given ρ and according to the Eq. 3, a square-shaped spatial neighborhood with dimension $h \times h$, where $h = 2\rho + 1$, is defined.

From such a process comes an "image of metrics". This image has the same support (i.e., number of lines and columns) of the primary classification but with an attribute amount (i.e., bands) equivalent to four times the number of primary classes, since the four adopted spatial metrics are applied to each LULC class. The attribute values observed on the image of metrics correspond to the returns of spatial
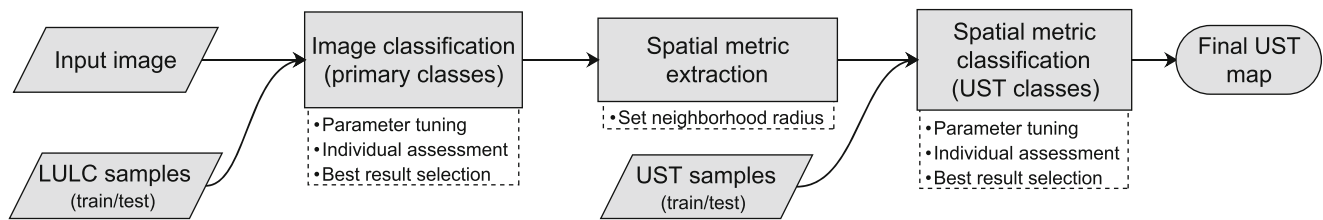
Fig. 2 Method structure flowchart

metrics for each pixel of the primary classification concerning its classes.

Posteriorly, taking the image of metrics as the input, a second classification process is carried out. A new sample set defined in terms of UST classes, again partitioned into training and testing, is adopted.

Additionally, since each pixel of the image of metrics expresses the spatial behavior over the analyzed area, considering its neighborhood, the use of spatially sparse point-wise observations as training samples is shown to be more convenient. Otherwise, the use of polygonal samples could encompass overlapping information from the pixels of its surroundings. Additionally, the local high variances shown by the spatial metrics may impair the classification process.

Similarly to the primary classification process, the SVM method was applied considering different parameter configurations, and the most accurate result was then selected. However, polygonal test samples were used to assess the UST classification accuracy. This choice follows the UST class definition: regions containing urban patterns in a city block level.

Lastly, the final mapping expresses the analyzed area in terms of UST, describing how the urban environment is organized according to its particular characteristics.

Experiments

In this section, we present two study cases regarding UST mapping using the framework proposed in "UST mapping framework based on spatial metrics classification". The following sections discuss the study areas and data used ("Study areas and data"), the experiment design ("Experiment design"), and finally, the results and respective analysis ("Results").

Study areas and data

The study areas comprehend two regions in Brazil (Fig. 3). The first one (Area 1) is a portion of São José dos Campos city, Brazil. An image acquired in September 2017 by the Landsat-8 OLI sensor was adopted for this area. This image has a spatial resolution of 30 m for the multispectral bands

and 15 m for the panchromatic band. In this case, it was used the following bands: blue, green, red, near-infrared, shortwave infrared (SWIR) 1, and SWIR 2. Also, it was used the panchromatic band for a pansharpening process.

The second study area (Area 2) comprehends a portion of São Paulo city, Brazil. In this case, it was employed an image acquired by Sentinel-2 MSI sensor in February 2021. Specifically, it were adopted the 10 m spatial resolution bands, regarding the visible (red, blue, and green) and near-infrared frequencies.

First, the Landsat-8 OLI multispectral bands were fused with the panchromatic band using the principal component analysis-based pansharpening method (Chavez and Kwarteng 1989), once it is a robust and well-known method designed to improve the spatial resolution of images (Pushparaj and Hegde 2017). This process generates multispectral bands with a spatial resolution of 15 m (Fig. 4a), yielding sufficient spatial information to define and distinguish the different LULC classes and USTs over the São José dos Campos study area. On the other hand, no additional image treatment was needed for the Sentinel-2 MSI image, once it has 10 m of spatial resolution (Fig. 4b), allowing the identification of the objects/targets over the study area.

LULC and UST samples (Fig. 5), required by the SVM method to perform the image classification processes, were collected on the fused Landsat-8 OLI image, and on the 10 m resolution bands acquired by Sentinel-2 MSI. The quantity of UST training samples was defined with a similar magnitude to the primary classification sample set. Reversely, since the test set designated to assess UST classifications comprises polygonal samples, its size tends to be much bigger than the sample set adopted to test the primary classifications. Table 1 summarizes the number of samples collected for the different classes, whether LULC or UST, used to train the SVM method and test the respective classification results. Also, the color key assigned to the classes, as presented in Table 1, remain the same for all the following figures and maps.

About the Area 1 (Landsat-8 OLI image – São José dos Campos), seven LULC classes were considered to perform the primary classification: ceramic roof, concrete roof, water, bare soil, asphalt, vegetation, and pasture. Such classes were chosen concerning the possibility of describing the USTs in
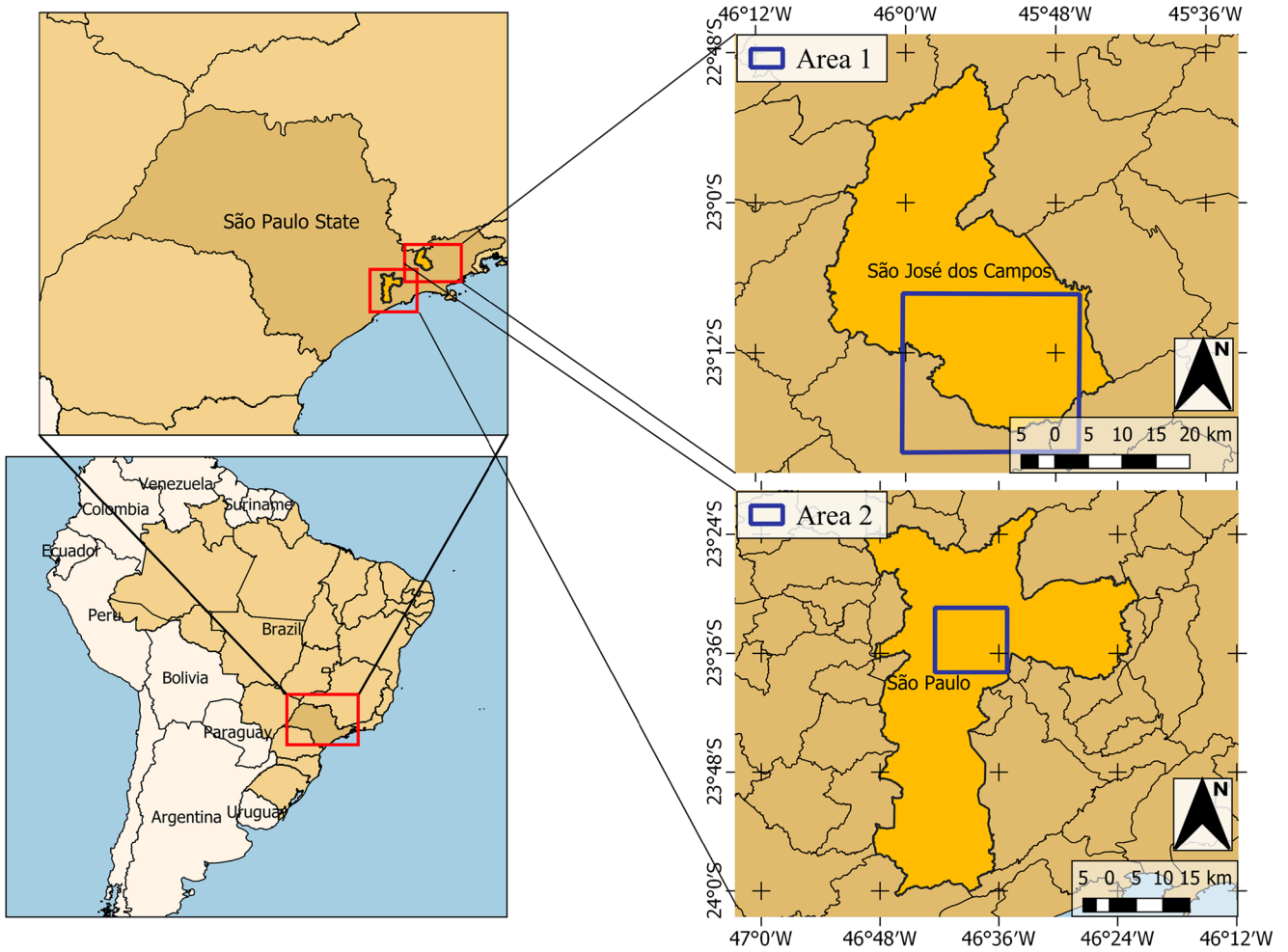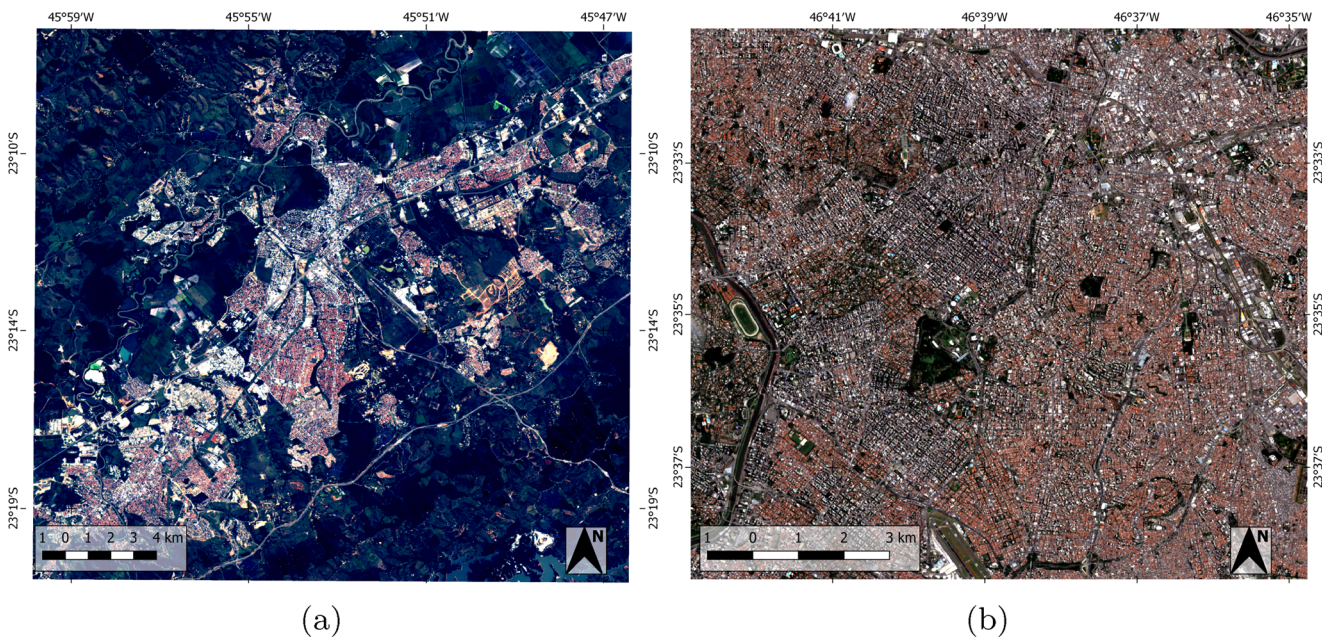
Fig. 3 Study areas location

Fig. 4 (a) Area 1 - Landsat-8 OLI and (b) Area 2 - Sentinel-2 MSI images in natural color composition
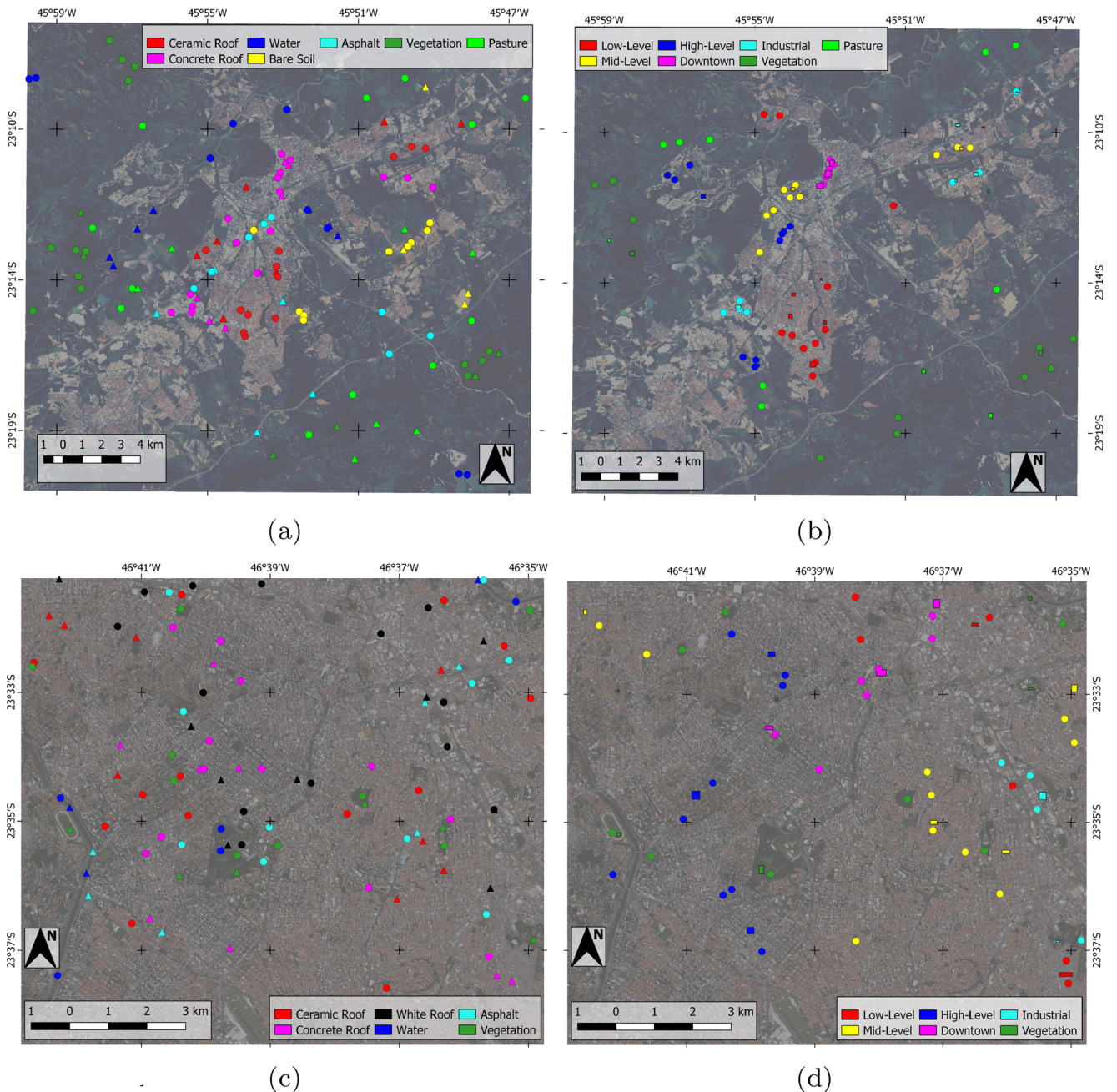
Fig. 5 Spatial distribution of training samples (●), point-shaped testing samples for primary classes (▲), and region-shaped testing samples for UST classes (■), where (a) Area 1 - primary samples, (b) Area 1 - UST samples, (c) Area 2 - primary samples, and (d) Area 2 - UST samples

the study area. Conversely, for Area 2 (Sentinel-2 MSI – São Paulo), it was considered almost the same primary classes of Area 1, except for including the "white roof" class and excluding both "bare soil" and "pasture" classes due to their absence.

Regarding the final mapping, seven USTs were selected in consonance with Wieland et al. (2016). Such USTs include three residential patterns (low-, mid-, and high-level), two service patterns (downtown and industrial), and

two rural patterns (vegetation and pasture). As previously mentioned, as Area 2 does not include "pasture" as a primary class, consequently, the respective UST class is not defined. The residential patterns differ from one another in terms of building sizes and open green spaces. The service patterns are described by the sizes and shapes of the buildings, usually with concrete roofs. In turn, the vegetation aspect and its concentration are the key elements to differentiate rural patterns.

Table 1 Training and testing samples of LULC primary and USTs classes for study Areas 1 and 2

LULC		Samples		UST		Samples	
Class		Train	Test	Class		Train	Test
Area 1							
Ceramic roof	■	12	14	Low-level	■	11	820
Concrete roof	■	16	15	Mid-level	■	10	256
Water	■	9	11	High-level	■	9	779
Bare soil	■	9	6	Downtown	■	5	1172
Asphalt	■	8	10	Industrial	■	8	595
Vegetation	■	14	9	Vegetation	■	11	460
Pasture	■	12	8	Pasture	■	8	467
Area 2							
Ceramic roof	■	13	8	Low-level	■	6	449
Concrete roof	■	12	8	Mid-level	■	10	697
White roof	■	13	9	High-level	■	9	912
Water	■	5	3	Downtown	■	7	859
Asphalt	■	10	6	Industrial	■	4	363
Vegetation	■	9	7	Vegetation	■	8	401

Experiment design

As already stated, a primary classification is initially obtained with the application of the SVM method, trained with samples of LULC classes (Table 1) with regards to the respective study area. To achieve accurate classification results, different parameter configurations for the SVM method are tested. Such configurations regard distinct penalty values ($C \in \{1, 10, 100, 1000, 10000\}$) under the linear, RBF (parameters $\gamma \in \{0, 05; 0, 1; 0, 25; 0, 5; 1, 0; 1, 5; 2, 0; 3, 0\}$), and polynomial (parameters $p = \{2, 3, 4, 5\}$) kernel functions using the One-Against-All (OAA) or One-Against-One (OAO) multiclass strategies.

The classification results obtained by each parameter configuration are evaluated in terms of kappa coefficient (Congalton and Green 2009), computed based on the test samples (Table 1). Afterward, the most accurate result observed is selected as the primary classification. Consequently, each spatial metric (Eqs. 5 to 8) is computed considering diverse neighborhood influence radii ρ. Different ρ ranges were used for each study image, specifically $\{1, 2, \ldots, 20\}$ for Area 1, and $\{15, 16, \ldots, 24\}$ for Area 2. The divergence of radii ranges between study areas results from the higher spatial resolution of the Sentinel-2 MSI sensor (Area 2), which demands bigger neighborhood radius values to encompass sufficient spatial information.

Each image of metrics generated from a given ρ value is classified using the SVM method and trained using the selected UST samples. All the different parameter configurations considered in the primary classification process are also evaluated for UST classification. Furthermore, the

kappa coefficient was used to evaluate the results. A final UST classification is selected according to the higher kappa value observed, considering all the adopted ρ values.

In Wieland et al. (2016), a UST mapping method is proposed through the SVM method and object-based classification concepts. Such a method is incorporated in the following experiments as a comparison baseline. Additionally, to provide UST mappings by a distinct classification method, the RF was adopted in alternative to SVM. For such purpose, the Orfeo Toolbox 7.1.0 (OTB) was used to carry out all classification steps (for more OTB details, see Grizonnet et al. (2017)). First, a segmentation using the Large-Scale Mean Shift algorithm (Fukunaga and Hostetler 1975) was carried out for the classification inputs. The segmentation's *minimum area* values are determined to ensure a dimensional equivalence with the neighborhood sizes regarded by the spatial metrics. For this, it was adopted the values of $h \times h$, used for build the spatial windows created for each ρ value in the spatial metrics calculation step. The object-based classification approach was then trained by the segment-shaped samples, selected by the same location as the UST point-shaped samples used in the proposed method. As aforementioned, the RF (Ho 1998) was used as the object-based classification method once, according to Huang et al. (2015), it could provide better results in urban studies when compared to SVM. The parameter configuration was based on the variation of the *maximum depth of trees* ($\{3, 5, 7, 9, 11\}$) and *minimum number of samples in each node* ($\{1, 2\}$), while other RF parameters were fixed at their default values, such as *maximum number of trees* (100) and *out-of-bag error* (0.01).

Finally, the significance of the best results from the proposed method are compared according their different ρ values. Also, they are compared against the best result from the alternative approach. The statistical test derived from the kappa coefficients (Congalton and Green 2009) is applied with 5% significance.

The experiments were run on a computer with an Intel Core i7 processor and 16 GB of RAM running the Debian Linux version 8.1 operating system. The programming platform was the IDL (Interactive Data Language), version 7.1. The code of the proposed framework is available for free at https://github.com/luccasmaselli/svmust.

Results

Area 1 – Landsat-8 OLI

Following the experiment design, the primary classifications were generated for the Landsat-8 image. Figure 6 shows the kappa values assigned to the different parameter configurations. The higher kappa value observed is equivalent to 0.952, obtained using the polynomial kernel function with $p = 3$, $C = 10^4$, and the OAA multiclass strategy.

Based on the selected primary classification, the considered spatial metrics were applied under different ρ values to verify the neighborhood radius influence on the final result.

As this case study considers four spatial metrics and seven primary classes, the generated images of metrics have 28 features.

Subsequently, the UST classifications were carried out. Figure 7 shows the kappa values achieved for different parameter configurations. In this case, the higher kappa value observed is 0.872, whose assigned parameter configuration is $C = 10^4$, with the polynomial kernel function of $p = 4$, the OAA multiclass strategy, and $\rho = 20$. The increasing trend of kappa values, given a kernel function and a multiclass strategy, appears when the classification results are ploted in ascending order in terms of neighborhood influence radius ρ. Such behavior implies that ρ plays a strong influence on the results of the proposed method.

Regarding the object-based image classification process, assumed as an alternative method for UST mapping, the most accurate result is assigned to a kappa value of 0.696, achieved by the parameter configuration of *maximum depth of trees* of 11 and *minimum number of samples in each node* of 2 and a segmentation generated by *minimum area* around 840 pixels (equivalent to $\rho = 14$). Figure 7 also summarizes the kappa values achieved by the alternative proposal, separated by classification methods and ordered in terms of *minimum area* value.

Figure 8a presents the best result achieved for the primary classification. Likewise, Fig. 8b and c present the best
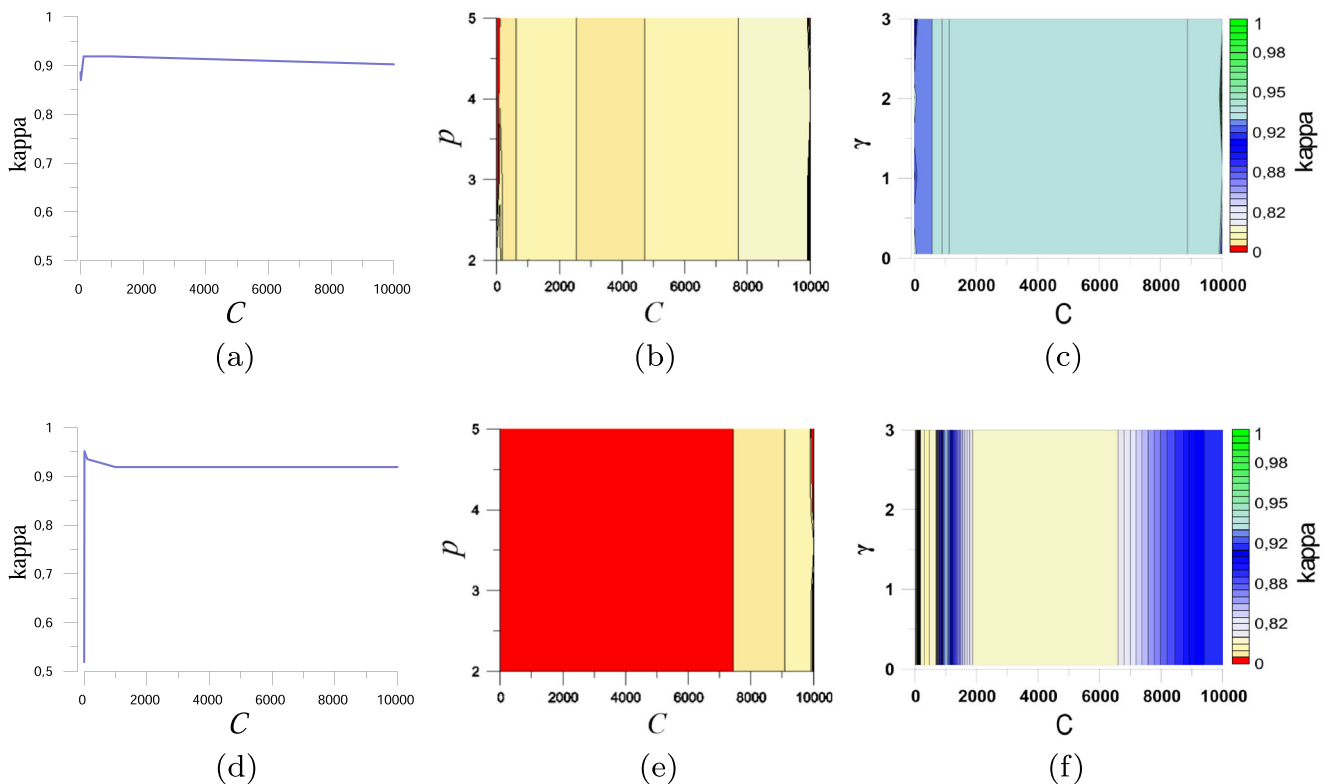

Fig. 6 Accuracy of primary classifications for Area 1 according to distinct kernel functions and multiclass strategies, where (a) Linear/OAA, (b) Polynomial/OAA, (c) RBF/OAA, (d) Linear/OAO, (e) Polynomial/OAO, and (f) RBF/OAO
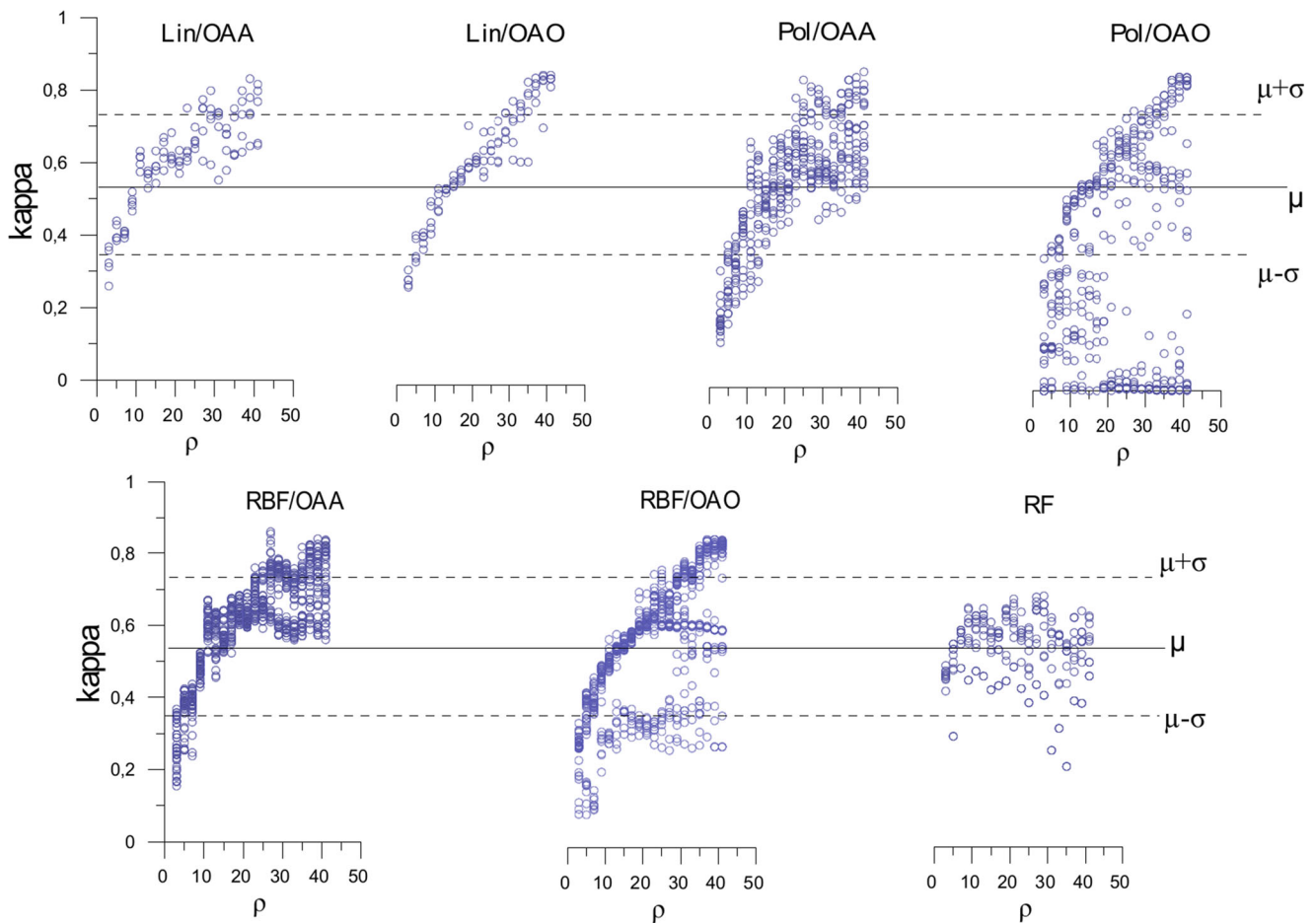
Fig. 7 Accuracy of UST classifications for Area 1 obtained by the proposed and alternative method. The mean kappa value (μ) and standard deviation limits ($\mu \pm \sigma$) are included for reference

UST classification provided by the proposed and alternative methods, respectively. As a supplementary check on the efficiency of the proposed method, a manual mapping of the study area was made in terms of UST, as presented in Fig. 8d.

Although the spatial metrics are calculated considering a context based on the primary classification, the proposed method involves a pixel-based classification. In turn, the alternative method adopts a object-based approach. Therefore, the divergence of kappa values shown by each method is explained by the effectiveness of the spatial metrics in expressing the analyzed USTs. The pixel-based classification approach followed by the proposal also plays a strong influence on the quality of the results.

Table 2 presents the *p*-values from a bilateral statistical hypothesis test, with 5% significance, adopted to compare the best results of the proposed method under distinct values for ρ. The alternative method is also analyzed (ref. "Best RF" column), and the proportion $\rho \approx (\sqrt{minimum\ area}/2)$ -1 is assumed for comparisons, once this method was

carried with *minimum area* parameters equivalent to each ρ value assessed by the proposal.

In general, some equivalences (represented in bold values at Table 2) are observed when using images of metrics derived from similar ρ. Also, better classifications come from bigger neighborhood influence radii. As already mentioned, the magnitude of influence radius has an essential role in the proposed method. Regarding comparisons with the alternative approach, the significance (and superiority) of the proposed method is verified in all cases.

When compared to the reference manual classification, the proposed method achieved similar results. Since it follows a pixel-based classification approach, a more detailed mapping is provided, leading to the identification of nuances that are not included in the empirical classification.

Lastly, regarding the final mapping from the proposed method, we may observe the predominance of low- and mid-level residential patterns. The high-level pattern is concentrated in specific areas, usually far from downtown or industrial areas. On the other hand, downtown is located
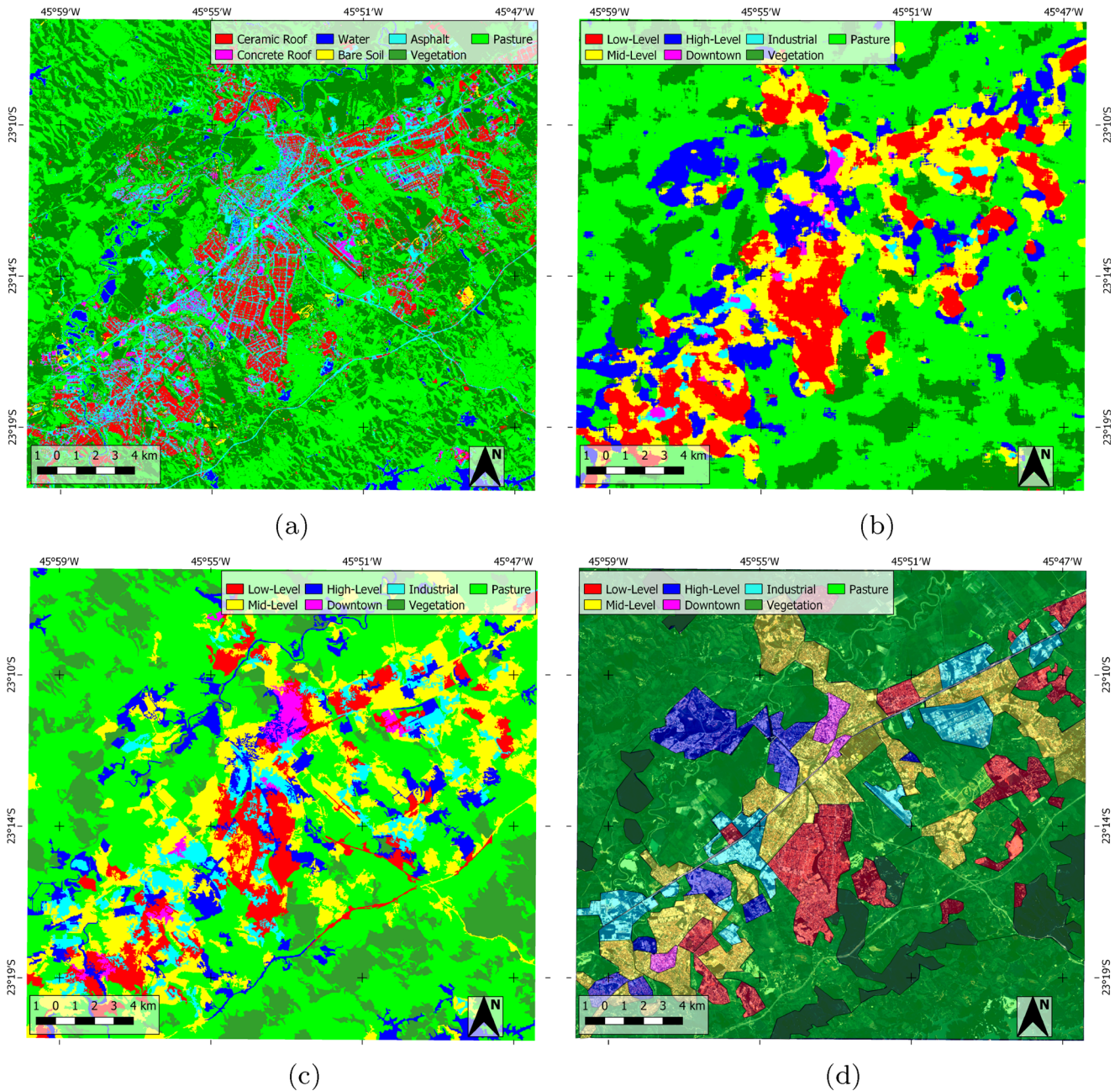
Fig. 8 Best results for the (a) primary, (b) proposed method's, and (c) alternative method's classifications, and (d) an empirical UST mapping for Area 1

at the center of the São José dos Campos city, characterized as a commercial area. Industrial areas are also concentrated in regions of industrial activities. This kind of information is useful to understand the arrangement of the city, and our proposed method is shown to be effective in such understanding.

Area 2 – Sentinel-2 MSI

Regarding the second study area, primary classifications were derived from the Sentinel-2 MSI image. High kappa

values were achieved using the RBF kernel function and OAO multiclass strategy. The best performance found stands for a kappa value of 0.941 when $\gamma = 0.25$ and $C = 10^3$. Figure 9a depicts kappa values profiles relative to the mentioned kernel function and multiclass strategy.

In a second moment, the best primary classification was submitted to spatial metrics computing. The range for neighborhood influence radius considered in this process were $\rho \in \{15, 16, \ldots, 24\}$. Whereas four spatial metrics are computed for the six primary classes, the generated images of metrics have 24 features. The best UST classification

Table 2 *p*-values ($\times 10^{-3}$) from a bilateral test to compare kappa values from Landsat-8 UST classification of proposed and alternative methods

ρ		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Best RF
ρ	1	–	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2		–	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3			–	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4				–	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5					–	1	1	**298**	0	0	0	0	0	0	0	0	0	0	0	0	10
	6						–	**478**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7							–	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8								–	1	3	0	0	0	0	0	0	0	0	0	0	61
	9									–	**405**	0	0	0	0	0	0	0	0	0	0	4
	10										–	0	0	0	0	0	0	0	0	0	0	10
	11											–	0	0	4	**309**	**471**	0	0	0	0	0
	12												–	59	1	0	0	**145**	**127**	17	2	0
	13													–	0	0	0	4	**334**	**295**	**102**	0
	14														–	1	5	27	0	0	0	0
	15															–	**284**	0	0	0	0	0
	16																–	0	0	0	0	0
	17																	–	14	1	0	0
	18																		–	**165**	44	0
	19																			–	**232**	0
	20																				–	0

Bold values represent statistic equivalence between methods

result showed a kappa value of 0.848 and was obtained using the polynomial kernel function with $p = 3$, OAA multiclass strategy, $C = 10^4$, and $\rho = 21$. Figure 9b represents the kappa behavior for different ρ values according to the best kernel function and multiclass strategy (i.e., polynomial kernel and OAA strategy) for the UST mapping by the SVM classification.

Regarding the UST classification provided by the baseline method, the most accurate result shows a kappa value of 0.498, achieved when using as parameter configuration
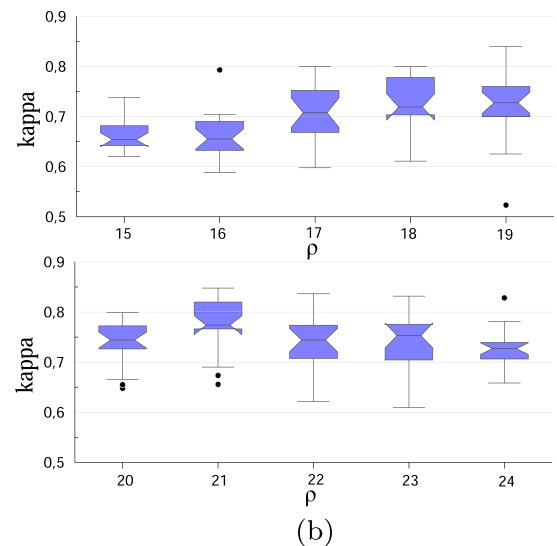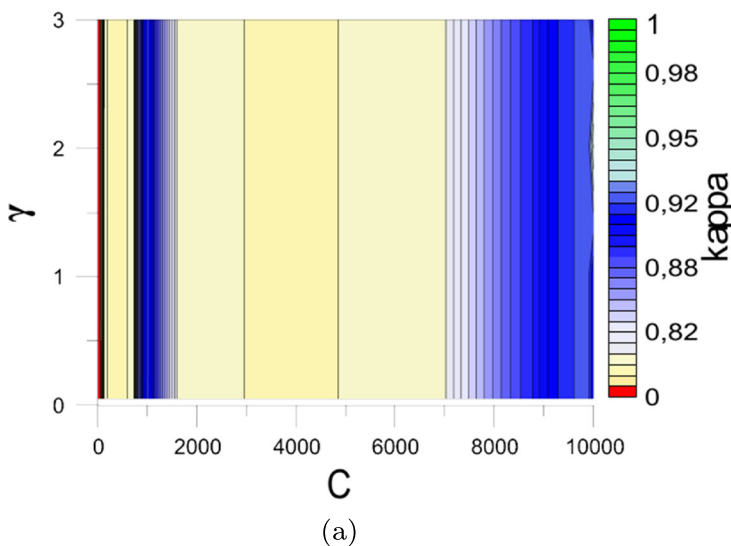

Fig. 9 (a) RBF/OAO SVM configuration for primary classification and (b) Pol/OAA SVM configuration for UST classification for Area 2, that is, the best kernel and multiclass strategy for each classification kind. In (b), the whiskers representing the minimum-maximum accuracy range, notches express the 95% confidence range around the median, and black dots as extreme values

a *maximum depth of trees* of 7, *minimum number of samples in each node* equal to 2, and a segmentation generated by *minimum area* around 961 pixels (equivalent to $\rho = 15$). In analogy with Area 1, Fig. 10 shows the better results for primary and UST classifications for Area 2, including the baseline method output and a manual classification for additional comparison. Moreover, Table 3 presents the *p*-value from a bilateral statistical hypothesis test, also with a significance level of 5%.

It is observed a statistical superiority of the proposed method is comparison to the baseline method. Such results

allow concluding that the use of spatial metrics favors a better UST mapping. However, it is worth highlighting the statistical equivalences among the proposed method's results when considering high values of ρ. This behavior can be assigned to the existence of an optimum value for the neighborhood influence radius. By gradually increasing, it is observed a maximum point of accuracy at a particular value ($\rho = 21$) and a loss of performance for radius values above it (Fig. 9b).

As previously mentioned, Area 2 comprehends a portion of the São Paulo city. Most of this study area is covered by
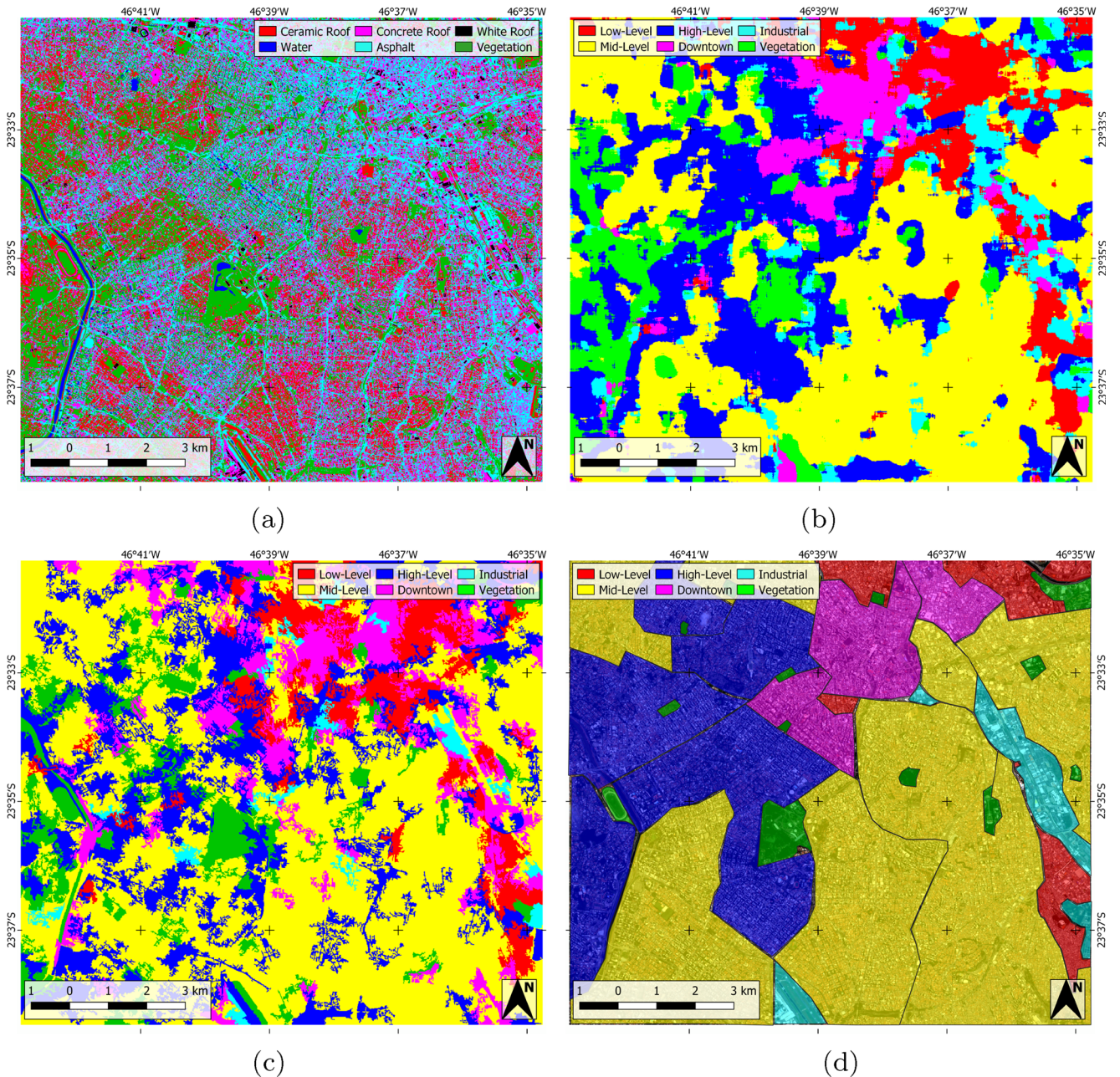

Fig. 10 Best results for the (a) primary, (b) proposed method's, and (c) alternative method's classifications, and (d) an empirical UST mapping for Area 2

Table 3 p-values ($\times 10^{-3}$) from a bilateral test to compare kappa values from Sentinel-2 UST classification of proposed and alternative methods

ρ		15	16	17	18	19	20	21	22	23	24	BestRF
ρ	15	–	73	0	**115**	0	**125**	0	0	0	0	0
	16		–	0	**399**	0	**385**	0	0	2	5	0
	17			–	0	**378**	0	**124**	**487**	**299**	**188**	0
	18				–	0	**484**	0	0	1	2	0
	19					–	0	**199**	**365**	**201**	**115**	0
	20						–	0	0	1	2	0
	21							–	**117**	49	20	0
	22								–	**310**	**196**	0
	23									–	**359**	0
	24										–	0

Bold values represent statistic equivalence between methods

mid- and high-level residential patterns. This city has several urban peculiarities, as different kinds of commercial and residential patterns. São Paulo's downtown, for example, is composed of high-rise buildings (at its business centers), high-density small shops (at its commercial centers), and the historical center, with unique morphology. The residential patterns, particularly the high-level, also may have different configurations over this study area. A common element over the residential areas is the presence of vegetation, where, depending on the ρ value, it can be misclassified as the UST vegetation class. Despite the high complexity of São Paulo, the proposed method showed a satisfactory in recognizing the urban patterns, proving then its effectiveness.

Conclusions

Understanding urban spatial dynamics is essential for decision making and sustainable planning. Remote sensing data and digital image processing techniques have been highlighted as potential tools for such a process. This study proposed a unique image-based method for urban area classification based on USTs. Two study cases, using Landsat-8 OLI and Sentinel-2 MSI imagery was carried out. Comparisons with an alternative method were also presented.

When considering appropriate parameter configuration, which includes those for the classifier (SVM), and for computing the spatial metrics (neighborhood radius), the proposed method can provide classification results with high accuracy levels. Moreover, it can afford consistent results according to the expected spatial behavior observed over the study area. Furthermore, the significance of the results was analyzed to prove the proposal's superiority when compared with an alternative method based on object-based image classification concepts. Additionally, the increase of the neighborhood influence radius also promotes statistically different results since the amount of information adopted for spatial metrics calculation is crucial for the results' quality. Also, it was noticed a trend of an optimum ρ value; that is, a spatial neighborhood size that sufficiently captures the spatial information and promotes correct UST classification.

Regarding the output maps, the proposed method showed efficiency in classifying the urban space into UST elements. Considering different urban complexities, the method effectively recognized the USTs in both cases. However, the higher complexity of São Paulo city makes it more difficult to separate some of the proposed classes. For example, high-level residential areas were misclassified in some regions as the dense vegetation presence observed in such areas is also associated with other urban standards.

Based on the study cases carried out, the possibility of classifying residential areas into low, medium, and high levels, as well as downtown and industrial regions is worth observing, highlighting the proposed method as a support tool for social actions and urban planning.

As future work, we plan to do the following: (i) consider other spatial metrics; (ii) investigate strategies to produce the image of metrics using a flexible neighborhood influence radius for each primary class; (iii) apply the proposed method to analyze multitemporal urban landscape changes; and (iv) suggest other UST classes according to the urban complexity of the analyzed area.

Declarations

References

Aljoufie M, Zuidgeest M, Brussel M, van Maarseveen M (2013) Spatial–temporal analysis of urban growth and transportation in jeddah city, saudi arabia. Cities 31:57–68. https://doi.org/10.1016/j.cities.2012.04.008

Ananias PHM, Negri RG (2021) Anomalous behaviour detection using one-class support vector machine and remote sensing images: a case study of algal bloom occurrence in inland waters. Int J Digit Earth 0(0):1–22. https://doi.org/10.1080/17538947.2021.1907462

Banzhaf E, Hofer R (2008) Monitoring urban structure types as spatial indicators with CIR aerial photographs for a more effective urban environmental management. IEEE J Select Topics Appl Earth Observ Remote Sens 1:129–138

Banzhaf E, Höfer R, Romero H (2009) Analysing dynamic parameters for urban heat stress incorporating the spatial distribution of urban structure types. IEEE Urban Remote Sens Joint Event 1–4

Belgiu M, Drăguţ L (2016) Random forest in remote sensing: A review of applications and future directions. ISPRS J Photogramm Remote Sens 114:24–31

Berger C, Voltersen M, Schmullius C, Hese S (2018) Robust mapping of urban structure types using high resolution geospatial data. gisScience 2:47–59

Böhm P (1998) Urban structural units as a key indicator for monitoring and optimizing the urban environment. Urban Ecology

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Bruzzone L, Persello C (2009) A novel context-sensitive semisupervised svm classifier robust to mislabeled training samples. IEEE Trans Geosci Remote Sens 47(7):2142–2154

Chavez PS, Kwarteng AY (1989) Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis. Photogram Eng Remote Sensing 55(3):339–348

Congalton RG, Green K (2009) Assessing the accuracy of remotely sensed data: principles and practices, 2nd edn. CRC Press/Taylor & Francis, Boca Raton

Deng JS, Wand K, Hong Y, Qi JG (2009) Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. Landscape Urban Plan 92:187–198

Fukunaga K, Hostetler L (1975) The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inf Theory 21(1):32–40. https://doi.org/10.1109/TIT.1975.1055330

Grizonnet M, Michel J, Poughon V, Inglada J, Savinaud M, Cresson R (2017) Orfeo toolbox: open source processing of remote sensing images. Open Geospatial Data Softw Stand 2(15)

Hecht R, Herold H, Meinel G, Buchroithner M (2013) Automatic derivation of urban structure types from topographic maps by means of image analysis and machine learning. In: 26th international cartographic conference

Herold M, Scepan J, Clarke KC (2002) The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. Environ Plann A Econ Space 34(8):1443–1458. https://doi.org/10.1068/a3496

Herold M, Goldstein NC, Clarke KC (2003) The spatiotemporal form of urban growth: measurement, analysis and modeling. Remote Sens Environ 86:286–302

Herold M, Hemphill J, Dietzel C, Clarke KC (2005) Remote sensing derived mapping to support urban growth theory. Joint Symposia URBAN - URS 2005 Remote Sensing and Urban Growth Theory

Ho TK (1998) The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 20(8)

Huang X, Liu H, Zhang L (2015) Spatiotemporal detection and analysis of urban villages in mega city regions of China using high-resolution remotely sensed imagery. IEEE Trans Geosci Remote Sens 53(7):3639–3657

Lehner A, Blaschke T (2019) A generic classification scheme for urban structure types. Remote Sensing 2:1–11. https://doi.org/10.3390/ rs11020173

Lu D, Weng Q (2007) A survey of image classification methods and techniques for improving classification performance. Int J Remote Sens 28(5):823–870. https://doi.org/10.1080/01431160600746456

Mather PM (2004) Computer Processing of Remotely-Sensed Images: An Introduction. Wiley, Hoboken

Montanges AP, Moser G, Taubenböck H, Wurm M, Tuia D (2015) Classification of urban structural types with multisource data and structured models. In: 2015 joint urban remote sensing event (JURSE), pp 1–4. https://doi.org/10.1109/JURSE.2015.7120489

Moon K, Downes N, Rujner H, Storch H (2009) Adaptation of the urban structure type approach for the assessment of climate change risks in ho chi minh city. 45 ISOCARP pp 1–7

Mountrakis G, Im J, Ogole C (2011) Support Vector Machines in Remote Sensing: A review. ISPRS J Photogram Remote Sensing Soc 66(3):247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Novack T, Stilla U (2017) Context-based classification of urban blocks according to their built-up structure. PFG J Photogram Remote Sens Geoinform Sci 85(6):365–376. https://doi.org/10.1007/s41064-017-0039-7

Pauleit S, Duhme F (2000) Assessing the environmental performance of land cover types for urban planning. Landsc Urban Plan 52:1–20. https://doi.org/10.1016/S0169-2046(00)00109-2

Pham HM, Yamaguchi Y, Bui TQ (2011) A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. Landsc Urban Plan 223–230

Pushparaj J, Hegde AV (2017) Comparison of various pan-sharpening methods using quickbird-2 and landsat-8 imagery. Arab J Geosci 10(119). https://doi.org/10.1007/s12517-017-2878-3

Simanjuntak RM, Reckien KMD (2019) Object-based image analysis to map local climate zones: The case of bandung, indonesia. Appl Geogr 106:108–121. https://doi.org/10.1016/j.apgeog.2019.04.001

Stewart ID, Oke TR (2012) Local climate zones for urban temperature studies. Bull Am Meteorol Soc 93(12):1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1

Tam TH, Abd Rahman MZ, Harun S, Kaoje IU (2018) Mapping of highly heterogeneous urban structure type for flood vulnerability assessment. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4/W9 229–235. https://doi.org/10.5194/isprs-archives-XLII-4-W9-229-2018

Theodoridis S, Koutroumbas K (2008) Pattern recognition fourth edition, 4th edn. Academic Press, Inc, Orlando

Tomás L, Fonseca L, Almeida C, Leonardi F, Pereira M (2016) Urban population estimation based on residential buildings volume using ikonos-2 images and lidar data. Int J Remote Sens 37(sup1):1–28. https://doi.org/10.1080/01431161.2015.1121301

Webb AR (2002) Statistical pattern recognition, 2nd edn. Wiley, Chichester

Webb AR, Copsey KD (2011) Statistical Pattern Recognition, 3rd edn. Wiley, Hoboken

Wieland M, Torres Y, Pittore M, Benito B (2016) Object-based urban structure type pattern recognition from landsat tm with a support vector machine. Int J Remote Sens 37(17):4059–4083. https://doi.org/10.1080/01431161.2016.1207261