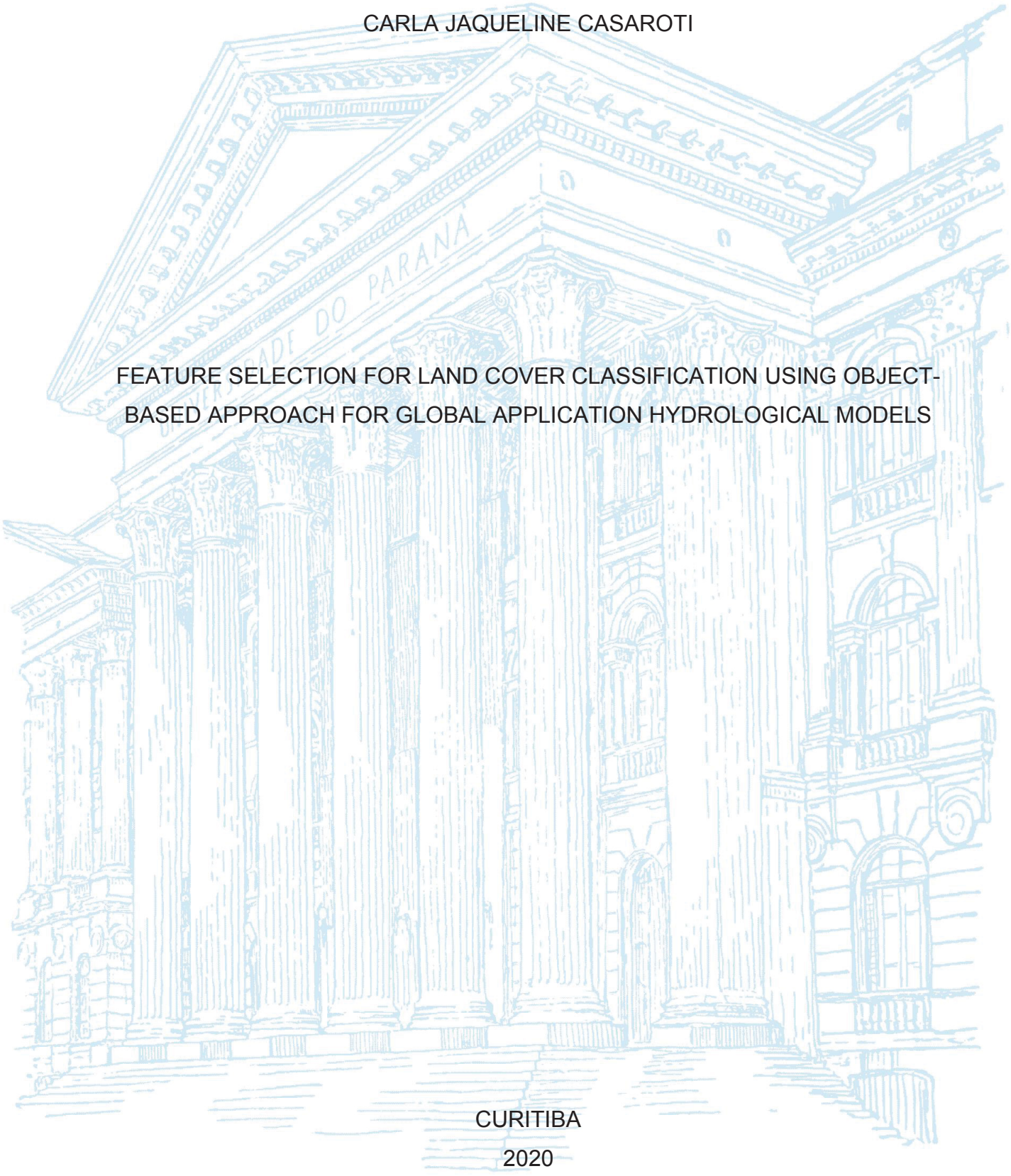
UNIVERSIDADE FEDERAL DO PARANÁ

CARLA JAQUELINE CASAROTI



FEATURE SELECTION FOR LAND COVER CLASSIFICATION USING OBJECT-BASED APPROACH FOR GLOBAL APPLICATION HYDROLOGICAL MODELS

CURITIBA

2020

CARLA JAQUELINE CASAROTI

FEATURE SELECTION FOR LAND COVER CLASSIFICATION USING OBJECT-BASED APPROACH FOR GLOBAL APPLICATION HYDROLOGICAL MODELS

Tese apresentada como requisito parcial à obtenção do título de Doutora em Ciências Geodésicas no Curso de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra da Universidade Federal do Paraná – UFPR.

Orientador: Prof. Dr. Jorge Antonio Silva Centeno.

CURITIBA

2020

CATALOGAÇÃO NA FONTE – SIBI/UFPR

C335f Casaroti, Carla Jaqueline

Feature selection for land cover classification with object-based approach for global application hydrological models [recurso eletrônico]/ Carla Jaqueline Casaroti , 2020.

Tese (Doutorado) - Programa de Pós-Graduação em Ciências Geodésicas da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Ciências Geodésicas.

Orientador: Prof. Dr. Jorge Antonio Silva Centeno.

1. Geodésia. 2. Seleção de variáveis. I. Centeno, Jorge Antonio Silva. II. Universidade Federal do Paraná. III. Título.

CDD 526.1

Bibliotecária: Vilma Machado CRB9/1563

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS DA TERRA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS GEODÉSICAS - 40001016002P6

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **CARLA JAQUELINE CASAROTI** intitulada: **FEATURE SELECTION METHODS FOR LAND COVER CLASSIFICATION USING OBJECT-BASED APPROACH FOR GLOBAL APPLICATION HYDROLOGICAL MODELS**, sob orientação do Prof. Dr. JORGE ANTONIO SILVA CENTENO, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua aprovação no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 21 de Fevereiro de 2020.

JORGE ANTONIO SILVA CENTENO
Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

MARIA DE LOURDES BUENO TRINDADE GALO
Avaliador Externo (UNIVERSIDADE EST.PAULISTA JÚLIO DE MESQUITA FILHO/PR.PRUDENT)

ALVARO MURIEL LIMA MACHADO
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

RODRIGO DE CAMPOS MACEDO
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Dedico essa tese em especial a dona Maria do Socorro, e que pode soar clichê, mas foi a melhor mãe do mundo.

AGRADECIMENTOS

Agradeço a todos que fizeram parte dessa caminhada, desde os amigos, aos mestres e mesmo aos percalços da vida que me trouxeram até aqui.

Em especial, agradeço ao meu adorado pai Carlos, minha irmã Gracieli que sempre foi nossa fortaleza, e minha mãe, que virou anjo antes mesmo de ver a sua segunda filha graduada.

AO FIM DO DIA, PODEMOS AGUENTAR MUITO MAIS DO QUE
PENSAMOS QUE PODEMOS.
(FRIDA KAHLO)

RESUMO

Para apoiar os estudos hidrológicos baseados em física ambiental, há a necessidade de encurtar a distância entre imagens de satélite bruto e modelos de gestão de recursos hídricos no que diz respeito às classes de cobertura da terra. À medida que as melhorias na resolução espacial de imagem permitem uma visão mais detalhada da superfície da Terra, o número de classes possíveis também aumenta. No entanto, os modelos hidrológicos ou de qualidade da água ainda dependem de dados grosseiros, como mapas de solo em pequena escala ou estimativas de chuva interpoladas que não são compatíveis com uma descrição detalhada da superfície. Por outro lado, manter os modelos mais gerais permite usá-los em diferentes regiões ao redor do mundo. Nesse sentido, é proposta uma rede semântica hierárquica de classes com base nos modelos hidrológicos e de qualidade da água mais comuns e, essa rede também está relacionada a resoluções de imagem. Os resultados mostraram a relação quase direta entre parâmetros de diferentes modelos e, comprovou-se que as classes necessárias pelos modelos podem ser organizadas dentro de uma rede hierárquica, da mesma forma que os sistemas de classes propostos por agências internacionais como CORINE e LCCS. Este esquema de classificação desenvolvido foi então aplicado a um caso de estudo no Brasil, a bacia de Vossoroca. As imagens Rapideye da área são classificadas usando a análise de imagem baseada em objetos (OBIA). O uso da abordagem de análise de imagem baseada em objetos em imagens digitais para fins de classificação em imagens de satélite de alta resolução espacial pode ser dividido em dois passos principais: o primeiro seria a etapa de segmentação e o segundo está relacionado à rotulagem esses segmentos ou objetos de acordo com um determinado conjunto de recursos e classificadores. As árvores de decisão são comumente usadas para representar o conhecimento humano em relação às classes de interesse na etapa de classificação. A questão nessa pesquisa foi sobre como selecionar uma quantidade menor ou combinação de características de um espaço de características disponível; que poderiam ser espaciais, espectrais e/ou texturais, a fim de descrever melhor as classes de interesse. Essa pergunta leva à escolha do melhor ou mais conveniente método de seleção de características. Para a etapa de seleção de características, diferentes algoritmos de seleção foram comparados: um modelo wrapper, um baseado em perceptron e outro no algoritmo SBS. O efeito do conjunto escolhido de variáveis derivadas das características disponíveis foi avaliado classificando as regiões de teste da mesma imagem. Acurácias globais da classificação foram satisfatórias, com valores maiores que 90%, revelando que, usar o método desenvolvido de seleção de características (modelo wrapper), junto com OBIA e o classificador pelo método do vizinho mais próximo é um bom caminho.

Palavras-chave: Seleção de Variáveis, Perceptron, Árvore de Decisão

ABSTRACT

To support environmental physics-based hydrological studies there are several studies made throughout the years in order to shorten the distance between raw satellite imagery and water resources management models as regards land cover classes. As improvements in image spatial resolution enable a more detailed view of the Earth's surface the number of possible classes also increases. Nevertheless, hydrological or water quality models still rely on coarse data, such as small-scale soil maps or interpolated rain estimates that are not compatible with a detailed description of the surface. On the other hand, keeping the models more general allows using them in many different regions around the world. In this sense, in this thesis it is proposed a semantic hierarchical network of classes based on the most common hydrological and water quality models based on the water body components' samples and this net is related to image resolutions. Findings showed the almost direct relation between parameters of different models and it is proved that the classes needed by the models can be organized within a hierarchical net, in the same way as the classification systems CORINE and LCCS proposed by international agencies. This classification schema is then applied to a study case in Brazil, the Vossoroca basin. RapidEye images are classified using the object-based image analysis (OBIA). The use of object-based image analysis approach on digital imagery for classification purposes on high spatial resolution satellite imagery can be divided into two main steps: the first one would be the segmentation step and the second one is related to labeling these segments or objects according to a certain set of features and classifier. Decision trees are commonly used to represent human knowledge regarding the classes of interest in the classification step. The issue here is regarding on how to select a smaller amount or combination of features from a feature space; that could contain spatial, spectral and/or textural features, in order to improve the classes of interest description as well as the feature selection in this context. That question leads to choosing the best or more convenient feature selection method. For the feature selection step, different feature selection algorithms were compared: the wrapper model, the perceptron-based model and the SBS algorithm. The effect of the chosen set of variables derived from the available features was evaluated classifying test regions from the same image. Satisfactory overall image classification accuracy larger than 90% were revealed when using the developed feature selection method along with the OBIA approach using the nearest neighbor as classifier.

Keywords: Feature Selection (FS); Perceptron; Decision tree.

FIGURE LIST

TABLE LIST

ABBREVIATION LIST

CEU – Single Stage Classifier
CN – Curve Number
CORINE – Coordination Of Information On The Environment
CWJM – Complex Wishart J-M separability
DT – Transformed Divergence
FNEA – Fractal Net Approach
GLCM – Gray Level Co-occurrence Matrix
GLDV – Gray Level Difference Vectior
GNDVI – Green Normalized Difference Vegetation Index
HSRI – High-spatial-resolution Image
JS – Jensen-Shannon
JSD – Jensen-Shannon Divergence
KLD – Kullback-Leibler divergence
LC – Land Cover
LCCS – Land Cover Classification System
LU – Land Use
MDS – Digital Surface Model
NDGR – Normalized Difference Green and Red-Edge
NDNB – Normalized Difference NIR1 and Blue
NDRR – Normalized Difference Red-Edge and Red
NDVI – Normalized Difference Vegetation Index
NDVIre – Normalized Difference Vegetation Index using Red Edge
NIR – Near Infra Red
NDWI – Normalized Difference Water Index
OBIA – Object-based Image Analysis
RFE – Recursive Feature Selection
SAR – Synthetic Aperture Radar
SBE – Sequential Backward Elimination
SBS – Sequential Backward Search
SCS – Soil Conservation Service Model
SCS – American Soil Conservation Service
SFA – Sequential Forward Algorithm

SFS – Sequential Forward Search

UAVs – Unmanned Aerial Vehicles

USLE – Universal Soil Loss Equation

SUMÁRIO

1 INTRODUCTION

Water is one of the most vital resources for human beings and it is used for various purposes, from drinking water supply to electric power generation. Many surface water bodies worldwide are strongly impacted and threatened in terms of quantity and quality. Water bodies are tightly related to their catchment and associated natural processes, for instance, erosion and sediment transport. Nevertheless, human activities can change the natural equilibrium and introduce severe changes that alter strongly the natural relationships between living species and water.

Today, a more conscious relation between the human race and water resources is being developed all over the world, as the problem gained global and huge implications. Human beings become more and more aware that natural resources need to be preserved to grant life and, that a sustainable approach regarding water management is needed to guarantee human life as well as social development.

Global initiatives are now looking for solutions to monitor and manage water quality and availability all over the world. Relevant national and international institutions are working towards normalizing, sharing and grouping data used for water and catchment management. A common tool to forecast changes as a basis for the proposal of use and preservation strategies is the use of models.

Although the quantity of models is considerable, depending on the aim and regional characteristics, many of them share common characteristics in terms of the type of information that is used as input. When the catchment is part of the study, information about its surface is needed to provide reliable estimates. For example, estimating runoff along the watershed demands knowledge of the amount of precipitation, as well as the physical characteristics of the catchment. Even the simpler model, such as the rational method model, needs such information. Its runoff coefficient (c) is a function of the soil type and drainage basin slope. As the model becomes more complex, more details are included and more input information is needed, such as land cover, land use, conservation practices or topography. Of course, more accurate results are expected with more complex models, but the information increase makes the model more difficult to use or applied in different regions. That is the reason why many hydrologists around the world look for simpler

models that can be easily adapted in order to perform historical or geographical comparative studies (Ragan and Jackson, 1980; Tucci, 1993; Bruno and Stein, 2004, Barbosa et al. 2015). A well-known simple model is the one developed by the American Soil Conservation Service (SCS), which relies on the use of a parameter that describes the catchment based on the area's hydrologic soil group, land use, treatment and hydrologic condition.

Among other parameters, freshwater service models are designed to predict water quality in a catchment and also depend on the reliable description of the surface and soil. Many of them are based on the USLE (universal soil loss equation) and its family (RUSLE – revised USLE, RUSLE2 – revised USLE 2, and MUSLE – modified USLE). They are physics-based and use spatially explicit environmental input data (Hamel et al, 2015). USLE-based models are widely used for water resources management since soil erosion and hence siltation of water bodies are major problems leading to the degradation of the water quality, regarding their physical and chemical parameters (Benavidez el al, 2018; Wohl et al, 2015; Beskow et al, 2009; Borrelli et al, 2017). The main equation for the USLE family has six factors; one of them is directly related to the basin land use and land cover, the cover and management factor (CP). Then, the USLE model also describes the catchment using coefficients, among which the cropping management factors (C and P or CP) is the key to describe surface conditions. The classes of interest and the coefficients used in this factor were both chosen to represent the river basin in a universal and simpler way. Since the C and P factors depend on the classified land use type, the correct identification of land use classes is of very high relevance. Due to high C values, the precise identification of agricultural and urban land is of particular importance (Hamel et al, 2015).

One of the aims of hydrological modeling and water quality models is to predict the consequences of land use changes (Brauman et al., 2007), which implies changes on the coefficients that describe the physical properties of the catchment. The importance of these changes will increase due to strong climate and direct anthropogenic changes (Beck et al, 2018).

In this context, remote sensing is the only viable option to create full-coverage, high-resolution catchment information of increasing quality also in terms of seasonal information (Borrelli et al, 2017). It plays an important role in current water and environmental management, providing information for a large number of

applications related to land cover studies and assessment, mainly for urban and semi-urban catchments.

The improvements in terms of spatial resolution and the advent of (very) high-spatial-resolution imagery either derived from satellites such as GeoEye and Quickbird or coming from the so-called unmanned aerial vehicles (UAVs), opens a wide research field in terms of image analysis. As the spatial resolution of images is improved, featuring pixel sizes of less than 5 meters, the variability between digital values of the same target (object) on the scene also increases (Carleer and Wolff, 2007), reducing the efficiency of spectral-based classifiers. On the other hand, the higher spatial resolution allows analyzing new variables such as the spatial ones.

Dealing with a detailed image of the Earth's surface proved that pixel-based methods have limitations. The use of spectral information, the digital numbers in the spectral bands, to take a decision about the possible class was widely spread in the 70s to the 90s, but the paradigm changed as high resolution images were introduced, because in high resolution images the classes are less uniform and object mixtures are more visible. Such new reality created space for alternative image analysis approaches such as the so-called object-based image analysis (OBIA) that adopts image regions as elements instead of isolated pixels. Uniform regions are delimited in the image and it is assumed that these regions are associated to objects in the scene. Therefore, the region concept in the image space is linked to the object concept in the scene. Using the object as a processing element opens up a wide array of ways in image classification, where it is possible to combine both spectral and spatial features to describe the classes of interest. Meanwhile, choosing the features that describe the classes as reliably as possible can be a challenging task due to a large number of combinations that can be made from all of the spatial and spectral descriptors available. Thus, within this framework of image analysis, it is reasonable to state that the OBIA classification is highly dependent on the proper selection of the most significant variables.

1.1 HYPOTHESIS

Considering the discussions from the above paragraphs some opportunities can be addressed and can be taken into account in this study. These main lines determine the aims of the present study:

- The first question can be formalized as: “is it possible to propose a class schema that fulfills the needs of broad hydrological models used in water quality?”
- The second question is: “once the classification schema is defined as a semantic hierarchical network, is it possible to automatically select the best features for the classification step and bring some interpretation from it? Differing from method that use dimensionality reduction, for instance.”

Formally, this study aims can be divided into two main parts:

a) To propose a semantic hierarchical network of land cover classes that can support hydrological studies at different scales considering several studies to base the network in; and
b) To propose a feature selection method that can be used within the context of the defined class hierarchy and brings more efficiency and interpretability to this matter as well.

1.2 CONTRIBUTION

One of the main desires regarding water management is that the hydrological model could be used worldwide, meaning in different scenarios and landscapes and, using data that should be as free and as easy to handle as possible It is also desired that the input data is similar in different studies to enable comparisons (Ragan and Jackson, 1980). For this purpose, in this study common water resources models’ parameters derived from the land cover are analyzed and compared with the aim of reducing complexity and building clusters at different generalization levels. This conceptual hierarchical network can be translated into a classification schema that can be solved using the OBIA approach. As an alternative to the statistical algorithms commonly used to conduct image classification, the use of a decision trees has been proposed by several studies, it became popular mainly by its use in the object-based image analysis algorithms (Van Coillie et al., 2007; Mahmoudi et al., 2013; Hamediantar and Shafri 2016; Wang et al., 2016). When it comes to object-based image classification, the system first divides the image into segments according to partitioning parameters and then classifies the segments using spatial, spectral or textural features.

The classes are defined by the hydrology models demands in terms of inputs information about the Earth's surface. They are also used to build a semantic hierarchical network. One of the most important aspects of defining the classes of interest and the semantic hierarchical network is to increase a good transferability on a global scale, also based on the study made by Yang et al (2017). There is the need for interoperability between different land cover datasets in order to harmonize all of the available land cover/land use systems and to improve the comparison, validation and understanding of patterns and changes in a universal way (Yange et al, 2017).

As the classes will be organized within a hierarchical network, it is reasonable to use also a hierarchical classification schema, a decision tree, product of the former class hierarchy analysis. The tree I then used to solve the classification problem by applying rules at each node, decision based on features that can be computed for each segment. The question addressed here is on choosing the features to be analyzed at each node of a given decision tree and the expected solution can also be applied to other watersheds using the same decision tree. Once established a hierarchical class network, a feature selection method is necessary to solve the classification problem.

The constant improvement of sensors as well as the object-based image analysis techniques for classifying high spatial resolution satellite imagery led to a large number of available variables, for instance, spatial and spectral features. Nevertheless, the large number of features (feature space) causes problems, as it happens when statistical methods are used. According to Haertel and Landgrebe (1999), a high dimensional feature space might cause problems in the estimation of the classes' covariance matrices. As the dimensionality increases, so do the number of samples requires a reliable estimation of the covariance matrix, which is known as the Hughes phenomenon, when dealing with a parametric classifier. In his work, Hughes (1968) concluded that as the number of features increases, so the classifier accuracy also does until a maximum accuracy value is reached. In that way, adding new features to the classification algorithm would reduce the accuracy instead of improving it. The solution for that issue would be to increase the set of training samples, which is not easy mainly because it is a truly time-consuming task. Therefore, methods to reduce the feature space dimensionality have been studied by several authors (XIE et al., 2018). Furthermore, reducing the number of descriptors in the feature space can also reduce computing and storage capacity. In this study, a

different method to solve the feature selection problem was introduced and discussed, playing the major role in explaining the features used also giving them interpretation through the classes used inside the hierarchical network, and using the most from the features themselves. Meaning that, in the wrapper model developed there were no hyper-parameters or parameters, such as happens when using Support Vector Machines or Random Forests, for instance.

2 REVIEW OF LITERATURE

In the following subsections, all the techniques and contextualization that were used as base for this work are presented in order to keep the context and give an overview of what has been made and what would be possible to be done.

2.1 SEGMENTATION

The first step in the image classification using the object-based image analysis (OBIA) approach is to group similar adjacent pixels in homogeneous regions that should correspond to an object in the scene or part of it. This is achieved applying segmentation algorithms. In the present study, the FNEA (Fractal Net Approach) was applied to perform segmentation.

The segments, also called regions, are spatially connected set of pixels that share common attributes. The algorithm used for this study was the well-known multi-scale resolution segmentation (FNEA – fractal net evolution approach) that can be stated as a region merging technique. This method depends on homogeneity definitions in combination with local and global optimization techniques, and comprises the choice of three main parameters being: the scale, shape and compactness parameters more about these and the segmentation approach can be found in Baatz and Schäpe (2000).

One of the prerequisites needed for a successful segmentation is to consider the paradigm of object-oriented analysis in image processing (BAATZ and SCHÄPE, 2001). According to Baatz and Schäpe (2000), this is achieved through a general segmentation algorithm, based on homogeneity definitions in combination with local and global optimization techniques.

In the FNEA (Fractal Net Evolution Approach) method available in the eCognition software, the grouping criterion is defined based on spectral and spatial heterogeneity (Definiens User Guide).

The parameter “f” is responsible for the production of more spatially uniform segments, if desired. This parameter is computed, according to equation 1, by the weighted sum of two

$$f = w \cdot h_{espectral} + (1 - w) \cdot h_{forma} \quad (1)$$

Where:

w : weight of the spectral uniformity;

$h_{espectral}$: spectral uniformity;

$h_{espacial}$: spatial uniformity;

The two criteria used to describe the separation of the image object are: spectral heterogeneity (color-related) and spatial heterogeneity (related to smoothing and compaction), which control the homogeneity of segments and regions (BLASCHKE et al, 2000). Color is directly related to spectral homogeneity and shape to spatial homogeneity (HOFMANN, 2002). A parameter is obtained by the weighted sum of these two criteria. (BATISTA, 2006). The spatial criterion is, again, the weighted sum of two parameters (compactness and softness).

The possibility of grouping the pixels in regions, the so-called objects, and even super objects, by combining adjacent regions, creates a hierarchical network in the image. Therefore, these different levels of generalization enable the generation of different levels of resolution in the same image, so that it is obtained from each available resolution level, which is appropriate for the extraction of information of interest. This hierarchical network is defined topologically, that is, the edge of a super object is consistent with the edge of its sub objects (KERSTING, 2006). The process ends when the smallest possible growth of a pair of objects exceeds a certain threshold, the scale parameter, according to Baatz and Schäpe (2000), the main components of multi-resolution segmentation are:

- **Heuristic decision –** it is the stage in which picture elements (initially pixels, then segments, are grouped within an iterative process. Assuming that there are two adjacent objects A and B; there are several heuristic approaches to group the elements despite other possible combinations in the image. eCognition has four approaches:
    - Adjustment: Groups object A with any neighboring object B when a heterogeneity criterion is achieved.
    - Optimal Adjustment: Object A is fused with a neighbor B if this is the best possible fusion in the image. That is, after analyzing all

possible fusions of adjacent objects, it is chosen the one that introduces less loss of uniformity in the image as a whole.

- Optimal Local Mutual Adjustment: Two objects (A and B) are combined if the fusion is the best possible for element A, at the same time that is the best option for element B.
- Optimal Global Mutual Adjustment: Groups the pair of adjacent objects that achieve the best heterogeneity criterion for the entire scene, considering mutual adjustment.

- **Definition of the fusion criterion** that determines the degree of agreement for a pair of objects**:**
  - Being X an n-dimensional feature space, in which each dimension is represented by a spatial or spectral variable, two objects A and B will look similar when they are as close as possible to each other, relative to all n-dimensions of X.

According to Baatz and Schäpe (2000) and Blaschke et al (2000), a multi-resolution or object-based segmentation should consider beyond the similarity criteria, the scale parameter, which determines the average size of objects. The definition of the segmentation parameters is still considered directly dependent on the characteristics of the image, the purpose of the segmentation – linked to the objects of the classes of interest that is desired to classify, and the knowledge of the analyst. In Benz et al (2001) it is emphasized that the choice of the scale parameter and similarity criteria is related to spatial and spectral resolutions of the image.

2.2 IMAGE FEATURES

The following subsections stand for the concepts related to the descriptors that were used in this research. These are described in three categories: spectral, spatial and textural. Although the texture can also be considered a spatial variable, here it is treated separately. Texture concerns the variation of the values of the digital value within the object and the spatial variables refer to the external characteristics of the segment, such as its shape or size.

2.2.1 Spectral features

Spectral variables are related to the response to reflected and/or emitted radiation by targets in the various regions of the electromagnetic spectrum and they can be (SONKA, HLAVAC, BOYLE, 1998.):

- Spectral mean: the average value of the digital values of a spectral band for each object of the segmented image.
- Spectral standard deviation: Quantifies the variation of digital values for each object in a spectral band.
- Spectral Ratio: is defined by the ratio between the spectral mean of the segment of a band, by the sum of spectral means over all segments.

$$r_L = \frac{\bar{c}_{L,obj}}{\sum_{i=1}^{m} c_{Li}} \quad (2)$$

Where

r_L: spectral ratio of the segment in the band L,

$\bar{c}_{L,obj}$: spectral mean of the segment in the band$L$,

j: segment identification,

m : total number of segments of the L band.

- Brightness: it is the sum of the values in all spectral bands.
- Maximum Difference: indicates the maximum spectral difference of a segment considering each spectral band.

2.2.2 Spatial fetures

Shape, size, and texture variables, which are the spatial attributes of the object, carry important information to describe the objects. They also provide information about the context of the object, that is, introduces the idea of human cognition in the perception of objects in a scene, considering information beyond spectral attributes.

Spatial attributes can be divided by according to their characteristics, commonly with geometric and texture properties. Thus spatial variations of the objects can be described primarily with geometric features.

- Size: It can be seen as the total number of pixels that build the segment.
- Length: It can be defined by linear extension in space from one end to another on an object.
- Width: describes the extension of the segment along the perpendicular direction to the main axis:
- Length/Width ratio: describes the shape of objects. A ratio close to one describes a compact object, while larger values stand for elongated shapes (ANDRADE & CENTENO).
- Compactness: It represents the concentration of pixels around a point, for example the centroid (SONKA et al, 1998).
- *Elliptic Fit*: compares the border of the segment to an inscribed ellipse.
- Rectangular Fit: It has the same purpose as the elliptical fit but using a rectangle.
- Border Length: The perimeter is the linear extension of the segments border.
- Shape Index: this variable describes the smoothness or roughness of the borders. Thus, the lower the shape index, the more smoothed the contour.
- Density: It is another feature to describe the concentration of pixels.
- Main Direction: designates the most significant spatial orientation for each object. This variable can be obtained computing the first eigenvalue of the line/row coordinates of all the pixels of the region.
- Asymmetry: demonstrates the geometric regularity of the segments. The more the shape of the segment approaches a circumference or square, the lower the value of its asymmetry.

2.2.3 Texture

One can understand texture as the variation of the digital values of the pixels in the image, or within a given region. If the values are very similar, then the region is said to be uniform or smooth in terms of texture. According to Gonzalez and Woods (1993), the hue and texture are highly correlated, although they are considered

independent properties, because if there were no hue variations, there would be no texture.

There are several proposals for the extraction of texture attributes in the literature, one of which is described by SALI & WOLFSON (1992), who describe the main statistical methods from the first-order statistics, the mean, variance and moment, considering the pixels neighborhood. They also include second-order statistics that are related to the co-occurrence matrix.

Haralick (1979) proposes the use of the co-occurrence matrix to derive a set of texture variables. For this purpose, the relative frequency of the digital values within the region is analyzed. Statistics are computed considering how many time combinations of different digital values happen at close pixels, preserving the spatial relationship. That means that it is counted how many time a digital value “I” is found in a pixel at the same time that the value “j” is present in the neighbor located in a given distance in a given direction (θ). These values are stored in a matrix which is later used to compute the features.

Haralick (1973) proposes a set of features that can be computed from the covariance matrix in order to describe the texture. Among these variables are: homogeneity, contrast, dissimilarity, entropy, second-order angular moment, mean, standard deviation and correlation, which are described below, based on individual digital counters for each segment (HARALICK, 1973).

- Homogeneity is a measure of local uniformity in terms of gray levels. An extreme homogeneous region would have all pixels of with the same value.
- Contrast refers to the difference between the highest and lowest values within the region.
- Dissimilarity quantifies the internal difference between the elements, the greater the difference between the elements, the greater the value of dissimilarity, which is very similar to contrast.
- Entropy measures gray levels disorder. The lower the values presented in the GLCM, the lesser the texture uniformity will be.
- Second order angular moment: evaluates texture uniformity, according to the repetition of pairs of digital values.
- Mean: This parameter is the mean value of the frequencies stored in the co-occurrence matrix.

- Standard Deviation: Provides a measure of dispersion of the data around the mean, within the matrix.

2.3 CLASSIFICATION

On a simplified way, the classification steps consist of labelling pixels or objects based on their characteristics, depending on the analysis approach. Since the region growing algorithm consists on dividing the image into segments or objects, that can be describe in terms of spectral features, spatial and textural features. These features build up a feature space that can be used to label the segments according to previously defined classes of interest, which leads to the next processing step in OBIA, the classification.

Image classification aims at, given a set of available features, estimating the membership relation between the object and the classes and deciding for the most similar class. There are several approaches to perform classification, however the best-known algorithms can be grouped into parametric and nonparametric. Parametric methods compare the hypothetical values of discriminating functions computed for each class, using parameters such as the mean vector and variance-covariance matrix and generally assume normal distribution. Conversely, nonparametric algorithms don't rely on parameters and evaluate relationships of dependency. Generally nonparametric techniques are called robust because they are applied to a wide variety of class distributions, when class signatures are reasonably distinct (SCHOWENGERDT, 1983).

Some criteria must be respected in order to classify remote sensing images into land use and coverage classes, (ANDERSON et al, 1976 apud ANDERSON, 1971):

- The minimum accuracy of the product should be at least 85 percent;
- The classification should be valid for extensive areas;
- The classification system must adapt to different dates;
- Class clustering into more general classes should be possible;
- A possible comparison with future soil cover classifications should be granted.

- Multiple land cover classes and uses should be mapped when possible.

From these considerations, it is concluded that the classifier methods used must comply with the minimum requirement of global accuracy (85%), as well as, depending on the purpose of the mapping to be carried out, the classes can be defined using decision trees, or directly. In this study, a semantic hierarchical network was developed to define de classes of interest in different levels as well as to be used to classify the high spatial resolution images.

2.4 DECISION TREE CLASSIFIER

Object-based classification methods provide the combination of several spatial, spectral, texture and neighborhood relationships in class description, which is closely close to human cognition in the perception of objects, as well as whether linked to topological relations between objects (MOLENAAR 1998). One option would be to use all computed variables to build up a feature vector for each segment and use it in the classification process. This option would lead to errors because some features are not significant for the description of the classes. Therefore, a preliminary feature selection step would be necessary.

Another option is to build up a decision tree that represents human knowledge using selected features, which the user considers relevant at each step, in a hierarchical classification. Nevertheless, a human is not always able to reproduce his/her knowledge with the available variables (features) and has difficulties to select them. Therefore, automatic variable selection methods for decision trees are necessary, as described in Patel and Rana (2014).

According to NAVULUR (2008), decision tree models are useful for both classification and regression problems. The problem domain consists of a set of classification or forecast variables and a dependent variable. In image analysis, typically, the forecast variables are the spectral and spatial features and other quantitative or qualitative variables, as a land cover class. The root node of the decision tree is located at the so-called level 0 (zero), which contains all the existing classes. In the following nodes decisions are taken to refine the groups according to (NAVULUR, 2008):

a) A decision rule;

b) A specific feature set (variables);

c) A specific set of classes.

Safavian and Landgrebe (1991) relate to potentialities of the Decision Tree Classification, point out the optimized form of decision model algorithms and their computational performance, as well as improvement in the discriminating performance of the classifier. In a Decision Tree Classification complex decisions can be replaced by a set of simple local decisions structured in hierarchical net, through the use of various levels of the tree. The advantage is, according to BATISTA (2006) the fact that, in the decision tree each sampled is evaluated with a restricted set of features, optimizing computational performance.

The decision tree classification is considered optimal or sub-optimal, based on the comparison between a pair of classes, which can be made through separability measures such as: the Euclidean distance, Bhattacharyya distance, the Kolmogorov-Smirnoff distance, among others (BATISTA, 2006).

2.5 FEATURE SELECTION FOR A DECISION TREE

When a decision tree is proposed, the next problem is to find the best feature (or set of features) that can be used in each node to support the decisions. A decision tree uses an organized tree-like model of decisions and relation between the decisions to obtain the classification result of the branches. The principle is to break up a complex decision (the desired classes of interest) into a union of several simpler decisions (nodes), hoping that the final product is the desired and most correct classification (Safavian and Landgrebe, 1991). Despite the fact that the principle is simple, the algorithm requires applying rules at each node, which should be able to perform the binary classification based on a reduced set of variables and thresholds.

Feature selection is part of the dimensionality reduction techniques that can be performed according two approaches: feature extraction and feature selection. In feature extraction, the original variables are combined in order to create a smaller set of new features that preserve the information of the original feature space. A typical feature extraction example is to apply the Principal Components transformation in the original feature space and then select the most significant new features (Weinberger and Saul, 2006), based on their percentage of representation of the entire dataset. Other well-known approaches to extract information such as Locally Linear Embedding and Multidimensional Scaling were also proposed by Zhang and Chau

(2009). In recent years, researches on deep learning techniques are also bringing alternative solutions, for instance, using the restricted Boltzmann machine (Sohn et al. 2013). A drawback of such approaches is based on their intentions to represent the original information from the dataset using a lower number of variables without considering on what the selection is aiming to. In that way, the same set cannot be suitable for different classification schemas.

On the other hand, feature selection methods are used to select a subset from the original set of variables with the most informative and distinctive variables to solve a certain problem according to a search strategy and an evaluation criterion. Firstly, the search strategy is based on looking for the most suitable subset among all the features available. Then, the evaluation criterion measures the success of each selected feature subset, for example, evaluating the classification results. Examples of how the search strategy and evaluation criterion work can be found in Aguilar et al. (2012), Guo et al. (2016) and/or Xiurui et al. (2014).

The problem of feature selection is basically to select the best features to solve a given problem among a set of available features. In this process, variables that are similar to another, or that are highly correlated to it, are discarded to reduce the computational effort. The optimal solution should still be able to solve the classification problem with enough accuracy and include, it is expected, the most significant variables.

The similarity between variables can be measured by statistical methods, such as the transformed divergence or the Jeffries-Matusita distance. Such distances measure the distance in the feature space considering the position of the classes, given by the mean vector, and the distribution of the elements, as described by the variance-covariance matrix. The next subsections present some of the statistical distances used in feature selection.

2.5.1 Transformed divergence

According to Jensen (1996), the transformed divergence measures the distance between the classes based on the provided samples. First, the divergence is computed, which is later normalized for comparison purposes. The scale of the transformed divergence values varies between 0 and 2000, thus, the higher the transformed divergence, the better the classes can be separated using the

considered features. It is also valid that, for a good separation of the samples, the values of the transformed divergence should be greater than 1900. The distance is considered reasonable when the values are between 1700 and 1900, and poor if it lies below 1700. The Transformed Divergence (DT) is given by:

$$DT_{a,b} = 2000\left(1 - e^{\left(\frac{-D_{a,b}}{8}\right)}\right) \quad (3)$$

$$D_{a,b} = D_1 + D_2 \quad (4)$$

$$D_1 = 0{,}5T_R\left((V_b^{-1} + V_a^{-1})(m_a - m_b)(m_a - m_b)^T\right) \quad (5)$$

$$D_2 = 0{,}5T_R\left((V_a - V_b)(V_b^{-1} + V_a^{-1})\right) \quad (6)$$

Where:

$DT_{a,b}$ – Transformed Divergence between classes a and b;

V_a –Variance – covarince matrix;

m_a – Mean vector;

$D_{a,b}$ – Distnce between classes a and b

T_R – trace of the matrix.

2.5.2 Jeffries-Matusita distance

The Jeffries-Matusita (J-M) distance can be used to measure the separability between classes and is frequently applied in pattern recognition and feature selection. The J-M distance is the J-criterion that separates two classes wi and wj that are members of a set of classes C (i,j=1,2,…,C,i≠j), as defined by Swain and Davis (1978). The J-M distance between a pair of classes, given by their probabilistic distributions is defined as (RICHARDS and JIA, 2006):

Here P(x/w_i) e P(x/w_j) are conditional density functions and conditional probabilities of the random variable x, given the classes w_i and w_j. It is assumed as the measure of the average distance between two class density functions, assuming the normality distribution, as follows:

$$JM_{ij} = \sqrt{2(1 - e^{-B})} \quad (7)$$

Where, B is the Bhattacharyya distance (RICHARDS e JIA, 2006), defined as:

$$B = \frac{1}{8}(m_j + m_i)^T \left[\frac{\Sigma_i + \Sigma_j}{2}\right]^{-1} (m_j + m_i) + \frac{1}{2}\ln \frac{\left|\frac{\Sigma_i + \Sigma_j}{2}\right|}{\sqrt{|\Sigma_i||\Sigma_j|}} \quad (8)$$

with:

m_j – mean vector,

Σ_i – variance – covariance matrix.

The exponential term B provides an exponentially decreasing weight in order to enhance the spectral separation of classes (RICHARDS and JIA, 2006). The J-M distance is asymptotic to 2, thus, a distance value equal to 2 between two classes would imply that distance is perfectly separable (RICHARDS and JIA, 2006).

Such classical methods are easily applied when the number of available features is reduced, as it happens in a pixel-based classification of a multispectral image, nevertheless, their computation becomes prohibitive or difficult as the number of features increase. They are also restricted by the normality assumption, which may not be granted in all studies. Therefore, other possibilities were searched to solve the problem, especially when dealing with object-based classifications. Some of them are discussed below.

2.5.3 Kullback-Leibler (KL) Divergence And -Shannon (JL) Distance

Within a decision tree, at each node, a decision is made to split the samples into two clusters that are supposed to be different. In statistics, to verify if two populations can be considered different from each other a suitable test must be applied and a hypothesis verified (MOURIER, 1946). One available option is to measure the similarity (or difference) between the classes using the KL divergence, also called relative entropy. A low, close to zero, KL divergence value indicates that the distributions of the two clusters have similar behavior, otherwise, they can be considered different (KULLBACK e LEIBLER, 1951).

Cantú-Paz (2004) developed a hybrid classification method, based on a genetic algorithm and the KL divergence as separability measure between the classes, using a public domain database. The performed experiments suggested that the proposed hybrid method was able to find more compact feature subsets providing

better results when compared to those coming from the Sequential Forward Algorithm (SFS) and the Sequential Backward Elimination (SBE).

Therefore, the KL divergence was proposed as separability criteria between histograms, which represent the distributions and their frequencies, for each pair of two different classes for all the available features (CANTÚ-PAZ, 2004):

$$D_{KL}(p||q) = \sum_1^i p(i).\log\frac{p(i)}{q(i)} \quad (9)$$

Where, D_KL is the KL divergence;

p(i) is the histogram of one class; and

q(i) is the histogram of another class in a specific feature.

The KL divergence is a not symmetric measure between two probability distributions, in other words; it is a measure of information loss when q is used to approach p.

There are some limitations to the use of the KL divergence when dealing with clusters in a segmented image. Therefore, alternatives were proposed, such as the Jensen-Shannon (JS) distance, which is based in the KL divergence, and includes symmetry and a finite value (LIN, 1991).

The Jensen-Shannon (JS) divergence is basically a distance metric that measures the similarity of two groups based on the comparison of the distributions of given features that describes the groups (ENDRES and SCHINDELIN, 2003). One advantage if the JS divergence is that it is always finite when finite random variables are used. Formally, the JS divergence is always positive and values close to zero describe would indicate that the clusters (or populations) are similar or even equal.

The JS divergence is based on the comparison of two entropy values: the entropy of the mixture and mixture of the entropy, according to equation 15 (ENDRES and SCHINDELIN, 2003). Considering two distributions (P and Q):

$$JSD(P||Q) = H\left(\frac{P+Q}{2}\right) - \frac{H(P)+H(Q)}{2} \quad (10)$$

If

$$M=(P+Q)/2 \quad (11)$$

then

$$H\left(\frac{P+Q}{2}\right) = -\sum M.\log M \quad (12)$$

$$H(P) + H(Q) = -1/2(\sum P.\log P \sum Q.\log Q) \quad (13)$$

By developing the equation, we achieve the JSD:

$$JSD(P||Q) = -\sum M.\log M + 1/2\left[\sum P.\log P \sum Q.\log Q\right]$$

$$JSD(P||Q) = -\sum \frac{P}{2}.\log M - \sum \frac{Q}{2}.\log M + \frac{1}{2}.[\sum P.\log P \sum Q.\log Q] \quad (14)$$

2.6 FEATURE SELECTION APPROACHES AND THE OBIA CONTEXT

Feature selection techniques require the establishment of a proposed aim as well as corresponding evaluation criterion and search strategy (Xie et al. 2018). The search strategy can be simple, such as the evaluation of all possible feature combinations, or optimized search algorithms, for instance, based on genetic algorithms (RIBEIRO, 2006). A comparison of feature selection methods in remote sensing is available in Serpico et al (2003). The most popular feature selection methods can be grouped into two groups: sequential backward search (SBS) and sequential forward search (SFS). The SBS method is an iterative process in which the original set of variables is analyzed, in each iteration the less significant variable is discarded until a desired number of variables is reached or when the derived set reaches a satisfactory stopping criterion for the problem. On the other side, the SFS method starts by selecting the most significant variable and subsequently new variables, the ones that are more significant between the remaining variables are added iteratively. It can be found different studies on feature selection in the literature. Some of them are introduced here.

Over the last years, scientific studies have been using object-based image analysis along with feature selection methods to improve high-spatial-resolution image (HSRI) classification. Examples can be found in the literature e.g., Van Coillie et al (2007) who used genetic algorithms as feature selection methods, Mahmoudi et al (2013) who applied multi-agent recognition systems, Hamediantar and Shafri (2016) who relied on the C 4.5 algorithm, or Wang et al (2016) who used support vector machine and random forest classifiers. Meanwhile, even though using spatial information (objects/segments) is highly indexed to classify HSRI, some authors e.g.,

Jung and Ehlers (2016), Persello and Bruzzone (2016), and Pu and Bell (2017) still used the pixel-based approach along with feature selection.

Using the object as a processing element opens up a wide array of ways in image classification, where it is possible to combine both spectral and spatial features to describe the classes of interest. Meanwhile, choosing the features that describe the classes as reliably as possible can be a challenging task due to a large number of combinations that can be made from all of the spatial and spectral descriptors available, thus providing a large dataset.

A set of well accepted methods is based on random forest classifiers. For example, Watmough et al (2017) introduced an operational structure for classifying high resolution three satellite images of a heterogeneous rural landscape – where land cover varied between agriculture and urban areas. The authors reinforce that the methodology is flexible enough to be applied in data from different platforms, with different geometries of target and lighting, from different dates of the year, producing results in the classifications of the different images that are reliable and can be compared between them.

To evaluate and select spatial and spectral attributes, within the object-based image analysis, Batista (2006) proposed the use of the Bhattacharyya distance at each node of the binary decision tree, using a Quickbird image. The results obtained from the proposed methodology were compared to the traditional process that employed a Single Stage Classifier (CEU), which showed a significant gain in the accuracy of the classification of images produced by the proposed method in relation to the traditional method, especially when the form/spatial attributes were used.

Chen et al (2010) developed a semi-supervised feature selection method, which consisted of evaluating the separation of classes into unbalanced sample sets (unequal quantities between the pixels of class samples). As part of the methodology, object-based image analysis was used for classification. This method incorporates asymmetric costs of incorrect classification, using weight matrices, as well as locally explores various types of relationships between sample pairs to more accurately evaluate the capacity of resources in preserving geometric and discriminating structures. The experimental results on satellite images of very high spatial resolution and airborne images proved the effectiveness and viability of the method.

In order to characterize patterns of the differences between Alzheimer's patients and cognitively normal people, as well as differences between people with and without cognitive impairment, Chu et al (2012) used four feature selection methods, and used high dimensional images of two different dates. The methods are: the univariate analysis with the t-student test, a method based on “a priori knowledge”, the recursive feature selection (RFE) and using the t-Student test using samples. The predictive accuracy obtained from different sample sizes, with and without selection of characteristics, were statistically compared. It was shown that more accurate classifications were achieved using “a priori knowledge” in patients with neurodegeneration, and that feature selection improves the accuracy of classifications using samples (CHU et al, 2012).

Mahmoudi et al (2013) studied the possibilities of multi-agent systems in object detection, given the difficulties encountered in the classification of objects in high and very high spatial resolution of urban areas. In the study a Worldview images and a digital surface model (MDS) were used as a data source. The authors concluded that the use of multi-agents easies the use of the object-based image analysis (with contextual relationships and structural descriptors) in object recognition, and improves the accuracy of the classification by about three percent.

Goshi and Joshi (2014) proposed another methodology when dealing with a study to map bamboo fragments using worldview-2 images, based on both pixel-based and object-based classification techniques. The object-based classification generated better results in thematic images, whose accuracy of the producer reached 94%.

Genetic algorithms are also a valid option for feature selection, as proved by Van Coillie et al (2007). The methodology was divided into three stages: segmentation, feature selection with by genetic algorithms (GOLDBERG, 1986) and finally the classification by neural networks (GIACINTO and ROLI, 2000; KITTLER et al., 1998; COILLIE et al., 2004; XU et al., 1992). The proposed method prove to be effective when the set of training samples is small, achieving up to 97% in terms of accuracy.

Dabboor et al (2014) used the J-M distance as criterion for separation between classes, when dealing with polarimetric Synthetic Aperture Radar (SAR) data. Although it is assumed that the data are normally distributed when using the J-M distance; the authors used the Wishart distribution. Two experiments were

performed to evaluate the CWJM method (Complex Wishart J-M separability), using two study areas. Better global accuracy was obtained with the Applying the Wishart algorithm the overall accuracy was around 85.1%, while when using the CWJM criterion it reached 88.5%. Thus, it was concluded that the CWJM separability criterion is useful for the evaluation of the separability of classes from a set of selected samples in the case of SAR images.

Ehlers (2016), in his study of spectroradiometer data to distinguish 14 sediment classes, proposed a new feature selection method based on the Jeffries-Matusita Distance (Jeffries-Matusita Distance Based Feature Selection). The results obtained by the proposed method were compared to the method of selecting attributes known as ReliefF (ROBNIK-SIKONJA and KONOKENKO, 2003); cited by the authors as the state of the art regarding selection methods. Both methods presented the ability to improve the separation of classes with global accuracy obtained above 82%, said, the advantage offered by the proposed method revolves around the shorter computational processing time compared to the method used in the comparison (ReliefF).

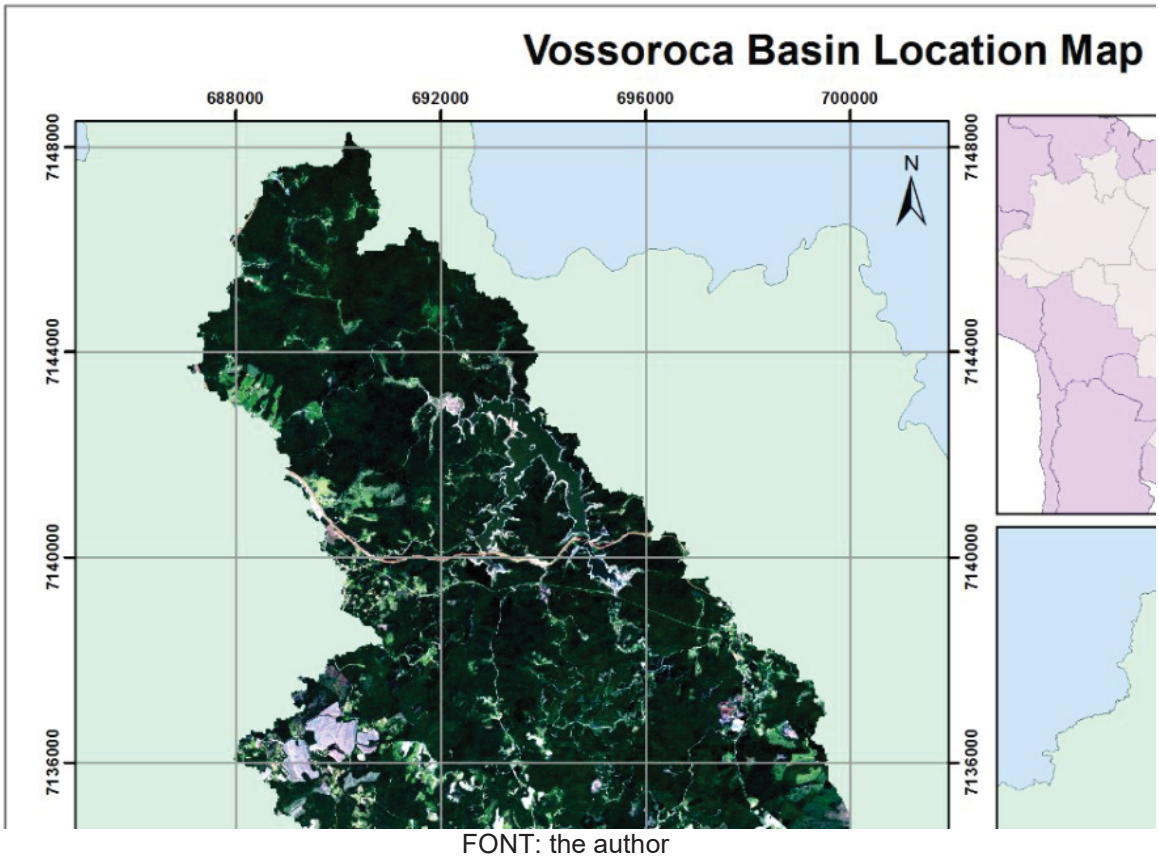
3 MATERIAL AND STUDY AREA

The area of interest is the Vossoroca reservoir catchment. It is located in the state of Paraná, at approximately 50 kilometers from it main city Curitiba, in the Serra do Mar. The reservoir water body covers an area of around five square kilometers and was created in 1940 to regulate the Chaminé dam water flow, located seven kilometers downstream. The basin area of the reservoir is predominantly rural, mostly covered by Atlantic forest, including Paraná pine (Araucaria angustifolia). Most of the reservoir catchment became part of the environmental protection area of the municipality of Tijucas do Sul. However, due to the agricultural land use and high relief energy in the upstream area of the catchment, erosion processes as well as siltation of the reservoir represent relevant environmental threats to water quality and electricity generation. This is especially the case since the metropolitan area of Curitiba consumes high amounts of water and the Vossoroca reservoir may one day become a source of drinking water.

Considering that the changes in river basins are important as regards to possible hydrological impacts (Welde and Gebremariam (2017), Temesgen et al (2018) and Lei and Zhu (2018)) the study was developed using two mosaics of images from the RapidEye sensor. A comparison between the features used to classify the two mosaics was possible since they had the same classes of land use and hierarchical network scheme. The mosaic images were acquired from the years of 2009 and 2014 for the Vossoroca basin. They are georeferenced to the WGS84/UTM22 system coordination. The images have five meters of spatial resolution and five radiometric bands; blue (440 – 510 nm), green (520 – 590 nm), red (630 – 685 nm), red edge (690 – 730 nm) and near-infrared (760 – 850 nm).

The Vossoroca dam is used for power generation and the basin is predominantly rural. The frequent land cover classes are "forest cover", "agriculture", "bare soil" and small "urban areas". Figure 1 displays the basin in RGB composition as well as its geographic location.

FIGURE 1 - AREA OF INTEREST LOCATION MAP – VOSSOROCA BASIN



FONT: the author

4 METHODS

The Object-based analysis of high-resolution image differs from pixel-based analysis because segments are used as image elements instead of pixels. That is, in addition to the spectral information used in pixel-based image analysis, object-based analysis also uses spatial and texture parameters. This concept had a very important milestone in July 2006, with the 1st International OBIA Conference held in Salzburg – Austria. From that, from a rigorous process of reviewing the articles sent to the conference, the book OBIA – Spatial Concepts for Knowledge-Driven Remote Sensing Applications originated, edited by the most well-known and respected researchers in this subject: Thomas Blaschke, Stefan Lang and Geoffrey J. Hay (BLASCHKE and LANG, 2006).

Therefore, the methodology is divided in steps:

1. Survey of the land cover classes used in hydrology studies
2. Definition of a class hierarchy for hydrological studies by clustering
3. Image segmentation
4. Proposal of feature selection methods for the decision tree, considering the study Survey of the land cover classes used in hydrology studies: Study case of the Vossoroca catchment.
    a. Feature selection using a perceptron
    b. Feature selection using the JS distance
5. Image classification

4.1 SURVEY OF THE LAND COVER CLASSES USED IN HIDROLOGY STUDIES

The discussion about the classes of interest is based on the demands of environmental physics-based models, such as USLE (WISCHMEIER and SMITH, 1960), or more complex USLE-based models such as MoRE (modeling of regionalized emissions) in Fuchs et al (2017), in terms of land use and land cover (LU/LC). A literature analysis was performed to find the most recent works regarding surface runoff and erosion prediction and determine how the classes of interest from LULC were defined. Even though they use different hydrological models, Rizeei et al (2017) and Demirel et al (2018) applied different classes of interest and used satellite imagery to obtain the most recent scenario of LU/LC classes. In comparison, Welde

and Gebremariam (2017), Temesgen et al (2018) and Lei and Zhu (2018) worked in the same direction analyzing how the LU/LC changes could cause hydrological impacts on river basins. Welde and Gebremariam (2017) and Lei and Zhu (2018) used already processed data from other institutions, whereas Temesgen et al (2018) used Landsat imagery to classify three different scenes. For all of the studies mentioned, the classes of interest were partly different or had different nomenclatures and different levels of detail regarding the map scale.

Considering previous studies, an important point is revealed regarding the classes of interest: even if classes had different names, they could be directly related to each other. This means that although their names may be partly or completely different, they represent comparable land cover or land use characteristics depending on the level of detail.

A similar problem was found when remote sensing was applied to map large areas, covered by different countries, and a standardized product was expected. Classification schemas were proposed to overcome the problem, such as the CORINE (coordination of information on the environment) and LCCS (land cover classification system) systems. A pioneer approach, in the field of hydrology, was proposed by Ragan and Jackson (1980), based on Landsat images. These three main studies were compared and used as a basis for the definition of classes of interest as well as the semantic hierarchical network for image classification. For instance, the study by Ragan and Jackson (1980) introduces a classification system that can be applied to Landsat images to map land cover classes according to the soil conservation service model (SCS, developed by the US Department of Agriculture) for urban areas. So, a hierarchical net of classes, similar to the one available in CORINE or LCCS, was proposed analyzing the class hydrological properties, as described by the values of the curve number (CN) of the Soil Conservation Model as well as the C factor of the USLE.

The soil conservation service curve number model (SCS-CN) defined a set of land cover and land use classes according to the soil, and covering conditions in watersheds in the USA. Translations and adaptations of this model can be found in the literature, (e.g. in Tucci (1993)). As described by Tucci (1993), the values in Table 1 can be used for Brazilian urban basins. Basically, the model computes runoff (Q), according to equation 20, based on the information about precipitation (P), initial

abstractions (Ia) and the potential maximum soil moisture retention (S) after runoff begins.

$$Q = \frac{(P-I_a)^2}{P-I_a+S} \quad (15)$$

The local soil conditions, given by S, are described in terms of soil type and land cover/land use and computed using a value (curve number CN) that ranges from 30 to 100, where lower numbers are related to low runoff potential and larger numbers describe larger runoff potential.

$$S = \frac{1000}{CN} - 10 \quad (16)$$

TABLE 1 - CURVE NUMBER (CN) VALUES FOR THE SCS MODEL FOR URBAN AND SEMIURBAN BASINS, ACCORDING TO TUCCI (1993).

Cover type and hydrological condition	Code	CN for hydrological soil			
		A	B	C	D
Cultivated agricultural lands:					
With soil conservation	'CultSC'	72	81	88	91
Without conservation	'CultCC'	62	71	78	81
Grass or bald land, bad conservation	'PastMa'	68	79	86	89
Bald land preserved	'Baldio'	39	61	74	80
Open space (lawns, parks, golf courses, cemeteries, etc.)					
Poor condition (grass cover < 50%)	'Aber50'	68	79	86	89
Good condition (grass cover > 75%)	'Aber75'	39	61	74	80
Meadow—continuous grass, protected	'PradoB'	30	58	71	78
Woods—grass combination	'Bosque'	45	66	77	83
Forest	'Floret'	25	55	70	77
Urban districts: Commercial APIA=85	'Comer'	89	92	94	95
Urban districts: Industrial APIA=72	'Indust'	81	88	91	93
Residential, according to average lot size (ALS) and average percent of impervious area					
1/8 acre or less, APIA=65 (town houses)	'Resi65'	77	85	90	92
1/4 acre APIA=38	'Resi38'	61	75	83	87
1/3 acre APIA=30	'Resi30'	57	72	81	86
1/2 acre APIA=25	'Resi25'	54	70	80	85
1 acre v APIA=20	'Resi20'	51	68	79	84
Paved parking lots	'estTel'	98	98	98	98
Streets roads					
drained	'RuasDr'	98	98	98	98

paved	'RuasPl'	76	85	89	91
dirt	‘RuasTr'	72	82	87	89

Another well accepted model is the USLE, equation 22, which estimates soil loss (A) using values representing four major factors: climate erosivity represented by R, soil erodibility represented by K, topography represented by LS, and land use and management represented by CP.

$$A = R \times K \times LS \times CP \quad (17)$$

An example of the USLE in a Brazilian watershed was described by Bueno & Stein (2004). Their study focused on the Brotas region in the state of Sao Paulo and considered the land use classes displayed in Table 2 to apply the CP factor. In a similar study, Barbosa et al (2015) included other land cover classes. More about other works regarding CP or C factor for other regions can be found, for example, in Shi et al (2002) or Lee & Lee (2006). However, to help create the semantic hierarchical network, Table 2 was used, where both CP factors and land use categories applied by Bueno & Stein (2004) and Barbosa et al (2015) were related.

TABLE 2 - CP FACTOR VALUES USED IN BRAZILIAN STUDIES ACCORDING TO (A) BUENO & STEIN (2004) AND (B) BARBOSA ET AL (2015).

Land use categories	CP factor
Dense urban /water (a)	0,0000
Dense vegetation (b)	0,00004
Cerrado (a, b)	0,0007
Urban residential (a)	0,0080
Woods (a)	0,0158
Farmlands (a)	0,0400
Sugarcane (a)	0,0500
Grass (a, b)	0,0050 - 0,01
agriculture	0,2
Dirty agricultural fields (b)	0,25
Bare soil (b)	1

4.2 CLASS HIERARCHY AND CLUSTERING

A set of land cover classes that can be used as input in the mentioned models was proposed. In this attempt, it was considered that the same class may be described with a different name, according to the source. In this effort it was noticed that the most complete description is given by the Soil Conservation Curve Number. Then, just few additions were necessary to turn the list complete. Each class was the described by a vector that includes the curve number values, as well as the CP factor of the USLE. The next step was to group similar classes into more general groups, making it possible to reduce the number of classes and, also, build up a hierarchical network that is controlled by the hydrologic information. This net was obtained by hierarchical clustering.

Clustering is the technique that groups similar elements (land cover classes in the present case) such that the elements in one group are more similar to each other than to other elements in the set. A group of similar classes is called a cluster and therefore the grouping technique is known as clustering.

Hierarchical clustering was applied because it has two advantages: first, it reduces the number of classes; it also combines similar classes building a new generalization level that can be linked to the image resolution, it is straight forward and is easily computed. Clustering can be agglomerative or divisive. In the present study, the agglomerative approach was used.

In the first case, agglomerative, each class is considered as an initial cluster. Within an iterative process, similar clusters are merged. The process ends when a predefined number of clusters is achieved.

It is necessary to evaluate the similarity between classes at each iteration. This is achieved computing a similarity matrix, a matrix that stores the similarity, or distance, between each pair of classes. This matrix allows identifying the most similar pair of classes that can be merged at this stage.

The result can be visualized as a dendrogram. A dendrogram is a tree-like diagram that represents the sequence of decisions taken at each iteration, fusions.

The divisive clustering is exactly the opposite of the agglomerative clustering. It starts with a unique class, that is composed by all the elements, and in each iteration the cluster is progressively divided if a uniformity criterion is not satisfied. It

is expected that new clusters are built and the classes better specified. This will not be detailed here because it was not applied in the study.

The agglomerative approach needs a similarity measurement to take the decisions. There are many options to measure similarity, as the already mentioned divergence, but simpler options are more common, because of the reduced number of available elements. The most common is the Euclidian Distance Between Centroids. Although other methods, like the single linkage can be used. In this case, the similarity is given by the shorter distance between elements of two groups. In a similar manner, the complete linkage computes similarity as the largest distance between elements of two different groups.

In the effort to combine land cover classes according to their hydrological properties, the Euclidian Distance Between Centroids was used here.

4.3 IMAGE PREPROCESSING

The images used in this work were acquired from the German satellite constellation RapidEye. They have the 3A level of preprocessing, which means that they are radiometrically, sensor- and geometrically corrected and are aligned to a cartographic map projection (UTM) and a horizontal datum (WGS84).

Even though the images were radiometrically corrected, their coming from different dates and times of the day required a normalization between the data. To solve this problem, new normalizations between the bands of the images were performed in order to make uniform mosaics for the area of interest. Ten images were used to produce the two mosaics for the Vossoroca basin – five images for time one and five for time two. The mosaics were built using the ENVI georeferenced mosaicking tool.

4.4 SEGMENTATION

The next step is image classification with object-based image analysis. When it comes to performing OBIA, there are two major steps: image segmentation and image classification. Due to the high level of information as regards class descriptors in high spatial resolution imagery, it is important to consider the feature selection step before using the classifier.

In the segmentation step, the interest area is divided into (geographical) objects using a segmentation algorithm. The one used for this study is the well-known multi-scale resolution segmentation (FNEA – fractal net evolution approach) that can be stated as a region merging technique. This method depends on homogeneity definitions in combination with local and global optimization techniques (BAATZ and SCHÄPE, 2000), which means that it works with spectral and/or spatial information as homogeneity criteria and a scale parameter to control the average image object size. This method was developed with a view to six design goals: high-quality image objects, multiresolution - objects of interest appear at different scales in an image, from a finer to a coarser level of detail, similar resolution, reproducibility, universality, and speed. For details on this technique and details on how the algorithm works, see (BAATZ and SCHÄPE, 2000).

As a consequence of image segmentation, dividing the image into uniform regions based on spectral and spatial characteristics, the spatial and spectral features were computed for all image objects. Therefore, to continue with the processing steps, the samples of the classes were collected and used regarding each level of detail. Different samples were collected to be used as information to perform the feature selection step, and then samples to perform the nearest neighborhood classification step, which are more detailed in the next sections.

Regarding the focus of this work, in one of the classifications' approaches a search strategy was used in order to choose the best feature or a combination of two or more features that could better describe and distinguish two classes at each tree node. This proposed method was based on the principles of a single perceptron; one advantage regarding the method used here is that, it does not require a specific statistical distribution, as occurs in the Transformed Divergence, for instance. Another one remains on proposing a linear combination of two or more features when necessary. Along with the used method advantages, one of the most important goals is to improve high spatial resolution satellite imagery classification using a simple and yet effective Machine Learning algorithm.

4.5 CLASS FEATURES

Before explaining how the feature selection methods worked, the part of how and from what kind of data the features come is an important step. The segmented image allows the computation of a huge number of features, beings spatial, spectral and textural for each object and each available class. Considering the objects or segments, the features that were taken into account were:

1) Spectral features:
   - Mean of the available bands;
   - Standard deviation of the available bands;
   - Maximum difference;
   - Brightness;
   - Water and vegetation normalized indexes.
2) Spatial features:
   - Based on the object area and dimensions: border length, number of pixels, length, length / width, length / thickness, area, width, volume, border-to-border ratio of the image;
   - Based on geometry - shape: shape index, circularity, rectangular fit, radius of the largest enclosed ellipse, radius of the smallest enclosed ellipse, main direction, elliptical adjustment, density, compactness, border index, asymmetry;
   - Based on geometry – polygons: standard deviation of edges, polygon auto-intersection, perimeter, number of edges, number of internal objects, length of the longest edge, volume, average length of edges, area including internal polygons, area excluding external polygons;
   - Based on geometry – skeletons: main line width, standard deviation of the area represented by segments, standard deviation of curvature, maximum branch length, circle line length, main line length, main line length / width, skeleton branching degree, curvature / length of the main line, average area represented by segments, average length of branches, number of segments;
3) Textural:

- The texture features were computed considering four directions (0°, 45°, 90° and 135°), according to Haralick (1979), using the co-occurrence matrices of gray levels, they are:
- GLCM second angular moment; GLCM contrast; GLCM correlation; GLCM dissimilarity; GLCM entropy; GLCM homogeneity; GLCM Average; GLCM standard deviation; GLDV second angular moment; GLDV Contrast; GLDV Entropy and GLDV Average.

The next step, after obtaining that large number of features for all the classes, the JS distance was used in order to choose the most significant ones, for each pair of classes.

4.6 CLASSES AND FEATURE SELECTION METHODS

To perform the feature selection step, a set of samples that would best describe the classes' distributions is needed, Congalton and Green (2009) stated that the minimum sample size acceptable would be 50 pixels per class of interest. From the samples collected for every class, one processing step of feature selection is the algorithm or measure to be used. In this work, the JS distance was used to select the best feature both for threshold and nearest neighborhood classification.

Regarding the sampling step, the samples were arranged into the hierarchical network previously defined in section 4.1. The first level was divided into water and non-water, from the non-water class all the other subclasses two by two were also arranged. In that way, the measure (JS distance) was used to compute the values between each two-classes node and define if and what features are more suitable on differing those classes.

Details on how the JS distance was used to select the suitable features to be used in the two different classification approaches are in the following sections.

4.6.1 Nearest neighborhood classification

The nearest neighborhood classification was used as one of the classifications approaches, once the features were defined by the JS distance-based method. This classification was performed into the eCognition ® software, in which the collected set of samples to perform the classification is used to assign values to

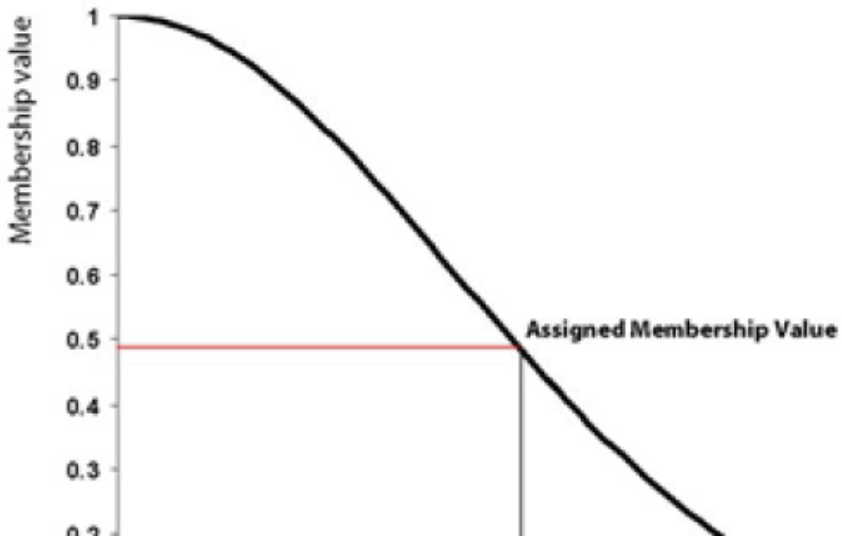
membership functions coming from Fuzzy logics. Two main steps compose the processing:

1) To estimate the classes' averages through the samples collected;
2) To classify the image objects (or segments) in the image domain based on the average classes' distances, selecting then the nearest one.

The pertinence value assigned by the classifier is given for an image object ranging between 0 and 1. This value is computed based on the distance that sets apart the object, in the feature space, from the classes in consideration. The pertinence degree is equal one when the image object is identical to one of the classes. Otherwise, as the pertinence degree decreases, the bigger is the distance between the image object and the considered class in the feature space. Thus, the classifier chooses the class with the biggest pertinence degree in order to attribute to this object a class label.

In the figure 2 bellow, it is possible to remark in the x-axis, 0.5 as pertinence value, and the distance between the object and the sample is of 0.8, leading to the conclusion that the object would have 50% of probability to belong to the analyzed class based on the samples.

FIGURE 2 - MEMBERSHIP FUNCTION EXAMPLE



FONT: *eCognition ®Developer User Guide*

4.6.2 Feature selection in the image analysis context

Although the two parts in OBIA might seem trivial, classification process can be difficult due to the huge amount of class descriptors/features (e.g. spatial and spectral features) that can be used to classify the image (Tang et al, 2014). Considering it is necessary to define the features that best describe the classes of interest for classification purposes, spectral, spatial and/or textural descriptors need to be selected among the large set of available features, being in such case, the objects from the segmentation process allowed to compute. Some examples of all the features available are listed below (Gonzalez & Woods, 2000; Haralick et al, 1973):

- spectral features – mean and standard deviation of the spectral bands,
- spatial features – length/width, area, border length, asymmetry, compactness,
- textural features – according to Haralick (1979), gray level co-occurrence matrix, such as entropy, homogeneity and so on, and
- other possible class descriptors – normalized indexes, such as NDVI (normalized difference vegetation index), NDRR (normalized difference red and red edge index), NDNB (normalized difference near-infrared and blue index), NDGR (normalized difference green and red index).

In that way, to select the best features and then perform the classification final step, a wrapper approach was developed for this work. Firstly, it may be interesting to clarify or remember the differences between filter and wrapper approaches. Theodoridis and Koutroumbas (2009) said that, in order to choose the best features in an optimal way, it is necessary to check all the possible feature combinations available. There are two ways to find and check those combinations:

- by using filter models: in which the feature selection works independently from the classifier method chosen. For each combination of features, a separability criterion should be used to compute and select the best feature or the best vector of feature combinations. This method has no control regarding how selecting one feature or a combination of them can affect the classification step, other than that, a huge number of combinations can provide good separability measures, bringing the question on which combination should be used towards generating good results;

- by using wrapper models: this approach is used considering that, as several or the most part of the times, a decision is needed to help choosing the feature or combination of features is worth using, other than only using some kind of separability measure. It consists of verifying the classification error that can be derived from a feature or combination of features used in a classifier method previously chosen, and then choosing the best one(s).

4.6.3 THE PERCEPTRON-BASED FEATURE SELECTION

Throughout the recent decades, studies introduced advances and new methods in feature selection. Gasca et al (2006) and Ruck et al (1990) proposed the use of multilayer perceptron to solve the problem of high dimensionality in the feature space. Such algorithms select a subset of relevant variables by estimating the relative contribution of the input variables from the corresponding classes' problem to the output neurons (Gasca et al 2006). Another recent example using the perceptron in feature selection was described in Habermann et al. (2018).

The perceptron is the basic learning algorithm in neural networks and machine learning. Although its formulation is relatively simple, it proved to be very efficient. A Perceptron is also described as a "binary classifier" because it proposes a function, which can decide whether an input represented by a feature vector belongs or not to a specific class (1 in the positive case, 0 in the negative case). Therefore, its input is a set of values (feature vector) and its output is either one or zero. This algorithm is considered as supervised because it develops its classification rule based on inputs with known output given by the user. Formally, the classification rule combines the inputs (feature values) to produce an output according to equation 18.

$$f(x) = \begin{cases} 1 \rightarrow if \ \vec{w} * \\ \end{cases} \tag{18}$$

When the input vector contains two values, equation 23 can be written as displayed in equation 24 Note that when the decision border is a line in a bidimensional feature space.

$$f(x) = w_1 * x_1 + w \tag{19}$$

When more input variables are used, the function becomes the equation of a plane or a hyperplane in a polynomial behavior with n-weights and a-bias for n-input

variables. The weights and b-bias computation depends on the given training dataset and it is achieved with the Backpropagation algorithm. This algorithm perform the weights correction according to the difference between the computed output and the desired one.

In this study, the idea was to use a single perceptron to find the best fitting of variables (or variable combination) for each classification network node using only one variable, a pair of variables and the SBS. It is relevant to highlight that the decision tree has nodes with only binary solutions. As described in equation 25, the decision is taken based on a polynomial whose size depends on the number of input variables. The problem that a perceptron can solve is, given a set of input features and a set of training samples known as the “expected classification output”, it computes the a priori unknown parameters of w. The Backpropagation algorithm is used to estimate the optimal set of parameters (wi) for the linear function used on the classification node, using the selected features as (xi) well.

$$f(x) = w_1 * x_1 + w_2 * x_2 + .. \quad (20)$$

The evaluation criterion is the degree of partial accuracy achieved with the probable solution. As there are just two possible classes at each node, the accuracy is measured as the percentage of correct classified elements.

Regarding this work methodology, in a first attempt, the perceptron was used to select the best feature for each node. For this purpose, all the features (Tables 1 and 2) were evaluated. This was a relatively simple task because just two parameters need to be computed: the weight w1 and the bias b in equation 26. The feature that best separates the two proposed classes was the solution and the value of the bias was the desired threshold in separating the classes.

$$0 = w_1 * x \quad (21)$$

In the second experiment, an exhaustive search was performed using pairs of features. As the search can achieve big proportions when the number of original variables increases, the process was divided into two steps. In the first step was conducted using only the spectral variables displayed in Table 1, under the assumption that the probability to solve the nodes containing the superclasses with spectral variables is high. From that point and variables, if no solution was achieved, then the spatial variables (Table 2) were included, but only for the nodes that remained without a solution. Pairs of variables were also evaluated because in many

cases one variable was not enough to describe the difference between the classes and superclasses. As displayed in equation 27, the decision boundary assumes the form of the equation of a line in the bidimensional space, which allowed finding solutions that were not necessarily parallel to the axes.

$$0 = w_1 * x_1 + w_2 \tag{22}$$

Finally, the perceptron was used to select the less significant feature within a SBS. The search started with all possible variables and computed the set of weights that enabled an optimal solution. The feature associated with the lower weight was then discarded and the process repeated. This iterative process was repeated until the accuracy of the binary classification at each node was above a given value. The results were finally evaluated using test samples. For the evaluation, the final classification was considered, not the results at each node. A confusion matrix was also computed for comparison purposes.

4.6.4 THE JS-DISTANCE-BASED FEATURE SELECTION

In this work, the wrapper model developed for the feature selection step was based on the Jensen-Shannon Divergence (JSD) and global accuracy using minimum distance. The JSD is a method of measuring the similarity between two probability distributions, also known as total divergence to the average (Dagan et al, 1997), based on the Kullback-Leibler divergence (KLD). It differs from KLD in terms of symmetry and finite value, on other words, JSD is a symmetrized and smoothed version of the KLD, and it is defined by (Lin, 1991):

$$JDS(p||q) = \frac{1}{2}D(p||m) + \frac{1}{2}D(q||m) \tag{23}$$

Where m is the mixture distribution defined by:

$$m = \frac{1}{2}(p + q) \tag{24}$$

And the KLD for the two distributions $D(p||m)$ and $D(q||m$are defined by;

$$D(p||m) = \sum_1^i p(i).\log\frac{p(i)}{m(i)}$$

$$D(q||m) = \sum_1^i q(i).\log\frac{q(i)}{m(i)} \tag{25}$$

(26)

To perform the calculations, the first set of class samples were separated to be used as training and test sets. The training samples were used to compute the JSDs for all the network nodes and features. Then, also for each node, the minimum distances were used to classify the training samples and the test samples, so the overall accuracy could be computed.

The KLD was computed based on the features' distribution for all the classes and nodes, that is to say, when combining "water" and "non-water" classes, for instance, the KLD values were performed for all the spectral, spatial and textural features from the training samples, using their probabilities that came from their histograms, and then used in the JDS formula. With this large amount of scores, the next step was to choose the best ones, thus, the minimum distance was used to test the overall accuracy for all the JSD values on the test samples. Only the five features with the best values of JSD and overall accuracies bigger than 70% were considered.

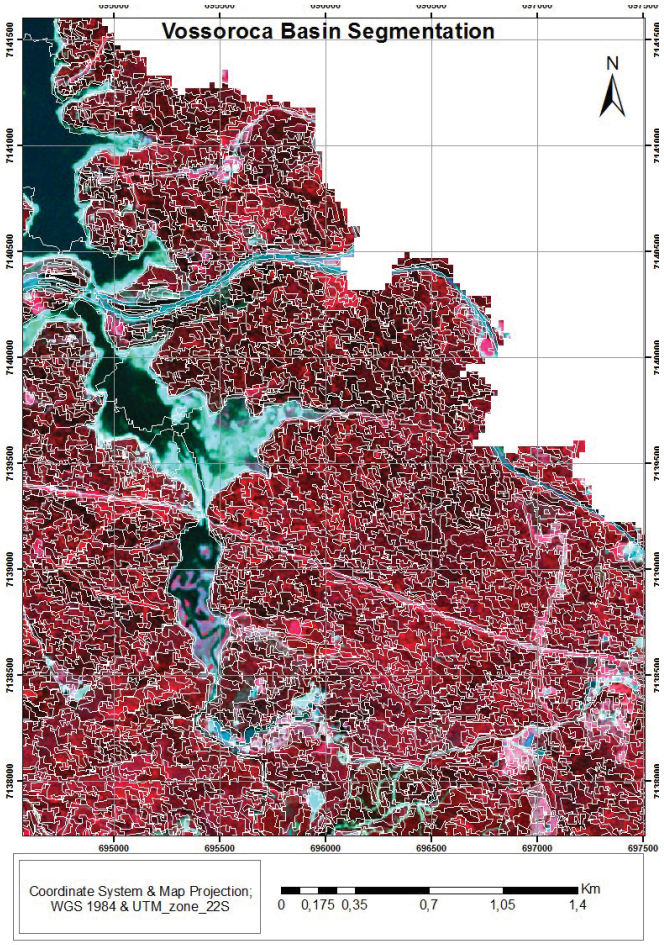
With the classes and their descriptors chosen from the features available at each level of the semantic hierarchical network, the classification step itself was then performed using the thresholds and nearest neighbor classifier. The thresholds were computed using one feature for each node with the best values of JSD and overall accuracy, then, based on the features' distribution, using the medium value between the end of the first distribution and the beginning of the second one.

On the other hand, the features to be used in the nearest neighbor classification in eCognition (based on a fuzzy classification algorithm) were chosen respecting the following rules: the ones with high JSD and overall accuracy bigger than 0.9, then, if it was not possible, only features with global accuracy bigger than 0.7. Different training and test samples were collected to be used in that step too.

5 EXPERIMENTS

Regarding the image segmentation, only one level was used. This decision was made in order to test the feature selection algorithm without a bigger influence of the user that could optimize the segmentation using his/her knowledge. The parameters for the segmentation were 100 for the scale parameter, 0.5 both for shape and compactness parameters. All the spectral bands were used to segment the image, only differing that for the NIR band weight 2 was used and weight 1 for the other ones (figure 3). The NIR band was used with more importance, one might say, due to scene land cover that was mainly covered by vegetated areas and water bodies, in which, the NIR plays an important role at describing them and their spectral curves.

FIGURE 3 - MAP WITH THE IMAGE SEGMENTS FROM 2009



FONT: the author

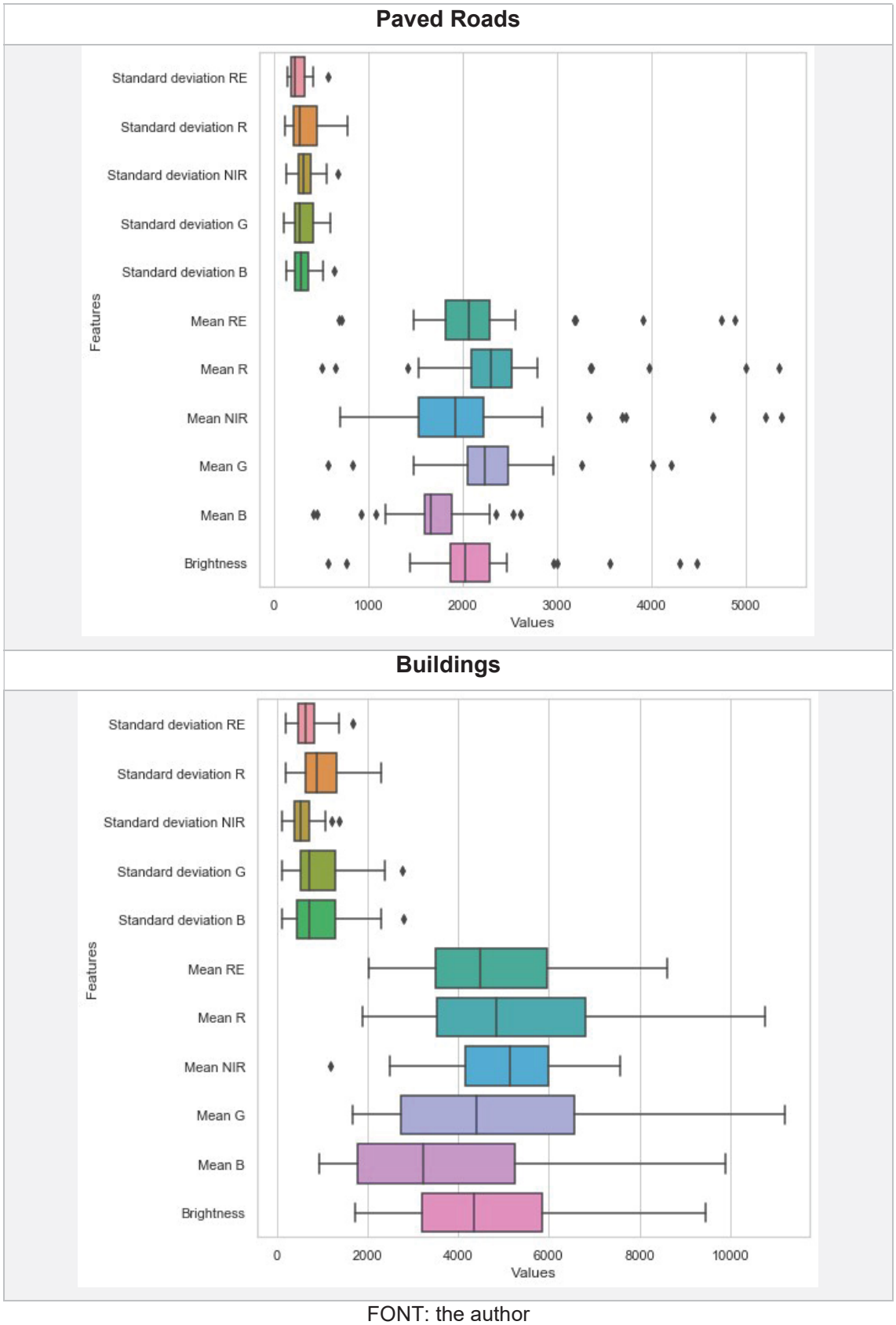
The NIR band was used with more importance, one might say, due to scene land cover that was mainly covered by vegetated areas and water bodies, in which, the NIR plays an important role at describing them and their spectral curves. Table 3 presents the list of features used in the feature selection step.

TABLE 3 - LIST OF USED FEATURES

Feature	Description
NDWI	Normalized difference water index – definition and equation in Gao (1996)
NDVIre	Normalized difference vegetation index – definition and equation in Tucker (1979), computed with the red edge and NIR bands
NDVIr	Normalized difference vegetation index, computed with the red and NIR bands –
Max. diff.	Maximum difference between bands
Max.Std.dev.	maximum standard deviation of the region in each band
Mean(RE)	Mean of the red edge
Mean(R)	Mean of the red band
Mean(NIR)	Mean of the near infrared
Mean(G)	Mean of the green band
Mean(B)	Mean of the blue band
Brightness	Sum of all bands

In figure 4 it is presented the variation of the spectral features of the final classes, using the 2009 training samples, regarding the last level of detail in the hierarchical network. This node was chosen in order to show how only the spectral features are not enough to distinguish these two classes (paved roads and buildings), once it is visible that they have almost the same intervals of values for the different spectral features.

FIGURE 4 - COMPARISON BETWEEN SPECTRAL FEATURES OF THE CLASSES PAVED ROADS AND BUILDINGS



FONT: the author

Table 4 displays the additional spatial features used in the feature selection step, they were computed in the eCognition software, more about how they are calculated can be found in Gonzalez and Woods (2002).

TABLE 4 - LIST OF ADDITIONAL SPATIAL FEATURES

Features	Features
Border length	Compactness (polygon)
Number of edges (polygon)	Average branch length
Width	Std. dev. of length of edges (polygon)
Asymmetry	Volume
Length/Width (only main line)	Radius of the largest enclosed ellipse
Rel. Border to Image Border	Perimeter (polygon)
Avrg. area represented by segments	Length/Thickness
Elliptic Fit	Std. dev. of the area represented by the segments
Std. dev. Curvature (only main line)	Main direction
Length of longest edge (polygon)	Length of main line (no cycles)
Density	Shape index
Average length of edges (polygon)	Degree of skeleton branching
Number of pixels	Thickness
Polygon self-intersection (polygon)	Number of segments
Radius of smallest enclosing ellipse	Maximum branch length
Area (excluding inner polygons)	Compactness
Rectangular Fit	Curvature/length (only main line)
Length of main line (regarding cycles)	Roundness
Length	Area
Area (including inner polygons)	Border index
Number of inner objects (polygon)	Width (only mainline)
Length/Width	

Training samples were selected for each class using the image objects (segments), taking care that use more than 30 segments, as widely used in probability and statistics. The samples collected by the analyst were based on the objects that could best describe the classes of interest. The quantities of each class are presented below, from the 2009 and 2014 mosaic images, respectively:

- Water: 31 and 34 objects;
- Non-water: 357 and 156 objects;
- Vegetation: 342 and 259 objects;

- Non-vegetation: 133 and 319 objects;
- Bare soil: 248 and 252 objects;
- Impervious area: 76 and 72 objects;
- Low vegetation: 99 and 167 objects;
- High vegetation: 267 and 181 objects;
- Reforestation: 89 ad 45 objects;
- Natural forest: 183 and 316 objects;
- Bare land: 139 and 337 objects;
- Bare roads:169 and 136 object;
- Buildings 62 and 105 objects;
- Paved roads: 34 and 21 objects;
- Grass: 93 and 112 objects;
- Agriculture: 172 and 223 objects.

For each training segment, the spectral and spatial features were computed and stored in a database. The samples were separated into training and test sets. It must be remarked that the frequency of the classes is not regular so that more samples are available for some of the classes, as well as the image segment sizes were different.

5.1 FEATURE SELECTION APPROACH AND CLASSIFICATION

Although the two parts in OBIA might seem trivial, the classification process can be difficult due to the huge amount of class descriptors/features (e.g. spatial and spectral features) that can be used to classify the image (Tang et al, 2014). Considering that it is necessary to define the features that best describe the classes of interest for classification purposes, spectral, spatial and/or textural descriptors need to be selected among the large set of available features. Some examples of all the features available are listed below (Gonzalez & Woods, 2000; Haralick et al, 1973):

- spectral features – mean and standard deviation of the spectral bands,
- spatial features – length/width, area, border length, asymmetry, compactness,
- textural features – according to Haralick (1979), gray level co-occurrence matrix, such as entropy, homogeneity and so on, and
- other possible class descriptors – normalized indexes, such as NDVI (normalized difference vegetation index), NDRR (normalized difference red and red edge index), NDNB (normalized difference near-infrared and blue index), NDGR (normalized difference green and red index).

Theodoridis and Koutroumbas (2009) said that, in order to choose the best features in an optimal way, it is necessary to check all the possible feature combinations available. There are two ways to find and check those combinations:

Filter models: Feature selection works independently from the classifier method. For each combination of features, a separability criterion is used to compute and select the best feature or the best sub vector of features. This method has no control regarding how the choice of one feature, or a sub set, can affect the classification step and a huge number of combinations can provide good separability measures, bringing the question on which combination should be used towards generating good results.

Wrapper models: In this approach not only the separability of the clusters is analyzed when selecting the best features, but it is also verified the classification error that can be derived the possible decision.

In this work, the wrapper model developed for the feature selection step was used, based on three concepts: the Jensen-Shannon Divergence (JSD); the minimum distance and the Overall Accuracy to check how well the test samples were classified using the minimum distance. The class samples were separated into training and test sets. The training samples were used to compute the JSDs for all the network nodes and features. In each node, the minimum distance method was used to classify the training samples and the test samples, and finally the overall accuracy was computed.

The KLD was computed based on the features' distribution for all the classes and nodes. For example, when analyzing the "water" and "non-water" clusters, the KLD values were computed for all the spectral, spatial and textural features from the training samples. To compute the similarity between the clusters with the JDS, it was used probabilities derived from the histograms of these features within each class.

This procedure enables a large number of scores, because all the features are used. Then, the next step is to choose the best features, which was performed analyzing the overall accuracy of the minimum distance classification of the test samples. Only the five features with the best JSD values and overall accuracies bigger than 70% were considered.

All the step described regarding the wrapper model were implemented into a MATLAB language script.

5.2 REFERENCE SBS

A comparison of feature selection methods in remote sensing is available in Serpico et al (2003). The most popular feature selection methods can be grouped into two groups: sequential backward search (SBS) and sequential forward search (SFS). The SBS method is an iterative process in which the original set of variables is analyzed, in each iteration the less significant variable is discarded until a desired number of variables is reached or when the derived set reaches a satisfactory stopping criterion for the problem. On the other side, the SFS method starts by selecting the most significant variable and subsequently new variables, the ones that are more significant between the remaining variables are added iteratively.

As the two approaches are new, the result of a established feature selection method was necessary to validate the success or fail of the proposed methods. For this purpose, the traditional SBS method was used here as reference. The search strategy is simple. First, all variables, after normalization, are used to compute the mean vector of each cluster. Then, the distance, or difference between the two clusters in each variable is computed. Finally, the variables are ranked according to their distance.

In a second step, within an iterative process, all available training samples are classified using these variables. The classification is performed using the minimum distance algorithm, using the means as representative value for the clusters. This allows computing the total error rate. If the error rate lies above a given threshold, then it is considered that the classification is a success. Then, it is verified if a variable can be discarded without increasing significantly the error rate. For this purpose, the variable with the lowest distance is discarded and a new set of variables obtained.

This new set of variables is used to compute the new error rate and the process is repeated. The process stops when then error rate achieves a value bellow the given tolerance or when only one variable remains.

This algorithm was applied to all nodes of the decision tree and the obtained accuracies were used as reference.

5.3 CLASSIFICATION

With the selected features for each level of the hierarchical semantic network, the classification step was then performed using thresholds and nearest neighborhood classifier. The thresholds were computed using one feature for each node with the best values of JSD and overall accuracy, then, they were used to classify the segments, based on their features' distribution.

As reference, a traditional classification was performed with the same dataset within the eCognition software. It was used the nearest neighbor classification (based on a fuzzy classification algorithm) and the features were selected using the following criteria: Consider the features with higher JSD and overall accuracy above 0.9. In some cases, this minimal accuracy was not achieved. In these cases, the value was decreased to 0.7. Different training and test samples were collected to be used in that step too, as previously stated.

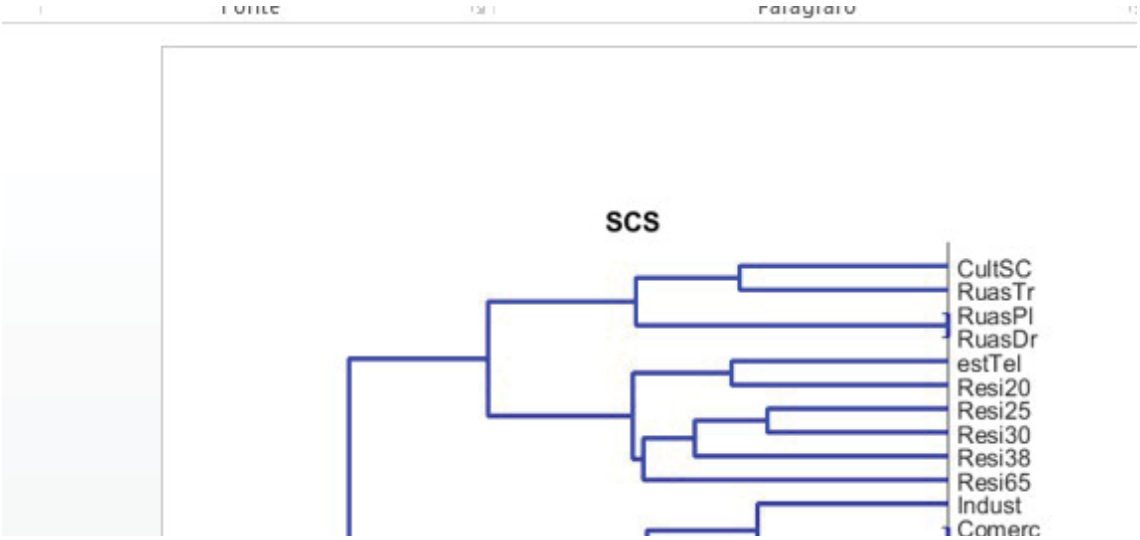
6 RESULTS AND DISCUSSIONS

As the project is composed of different steps, the results were also organized according to the steps described in the methodology. First, the results of the class selection are introduced. Later, the feature selection results are described.

6.1 HYDROLOGICAL LAND COVER CLASSES

As the number of land cover classes is larger for the SCS-CN model and it includes other classifications, such as the most part of the classes in Table 2, two dendrograms of land cover classes of the SCS model were created. The CN values can be stored as a vector for each land cover class that was used to perform a hierarchical clustering, aiming at grouping similar land covers in terms of hydrological properties. The first result is displayed on Figure 2(a). In a second attempt, the dendrogram can be computed including the USLE CP factor for each class. The result is displayed on Figure 2(b). The column “code” in Table 1 is used to help on identifying the real label of the land cover classes in the dendrogram (Figure 2), since they were of considerable size to be represented in the image.

FIGURE 5 - DENDROGRAM OF LAND COVER CLASSES OF THE SCS MODEL, ACCORDING TO THE CURVE NUMBER VALUES



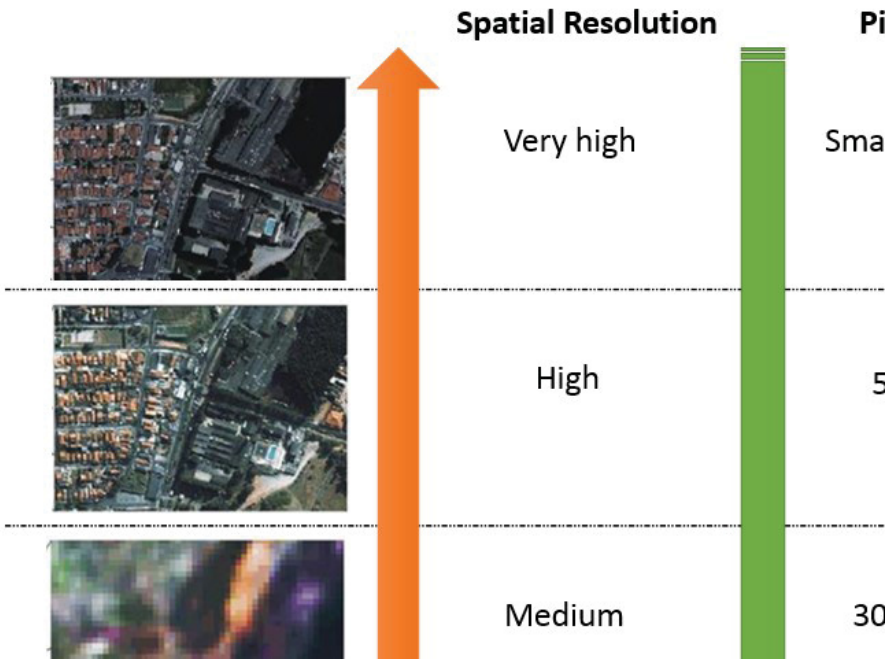
Comparing the dendrograms table 5 was created based on the relations between their classes as well as a semantic hierarchical scheme to classify images from a coarser to a finer level of detail.

TABLE 5 - COMPARISON BETWEEN SCS-CN AND BRAZILIAN STUDIES ACCORDING TO BUENO & STEIN (2004) AND (B) BARBOSA ET AL (2015)

Level 1	Level 2	Level 3		Level 4
SCS and CP factor	SCS and CP factor	SCS	CP factor	-
Vegetation	Low vegetation	Cultivated	Agriculture	-
		Grass		-
	Medium/high vegetation	Forest and woods	Dense vegetation	-
		Meadow	Medium vegetation	-
Level 1	Level 2	Level 3		Level 4
SCS and CP factor	SCS and CP factor	SCS	CP factor	SCS
Non-vegetation	Paved areas	Buildings	Roads	Paved
				Dirt
		Urban residential	Dense urban	-
	Bare land	Bare land	Bare soil	-
			Dirt agriculture	-

Another important aspect to be taken into account regarding the thematic image classification process is the spatial resolution that is directly related to the scale. The spatial resolution of images can range from low to medium through to high and very high spatial resolutions. Figure 6 shows a simple scheme of different spatial resolutions related to each other concerning the level of detail and the minimum area represented in the respective images.

FIGURE 6 - IMAGE SPATIAL RESOLUTION AND SCALE COMPARISON (ADAPTED FROM MELO (2002))

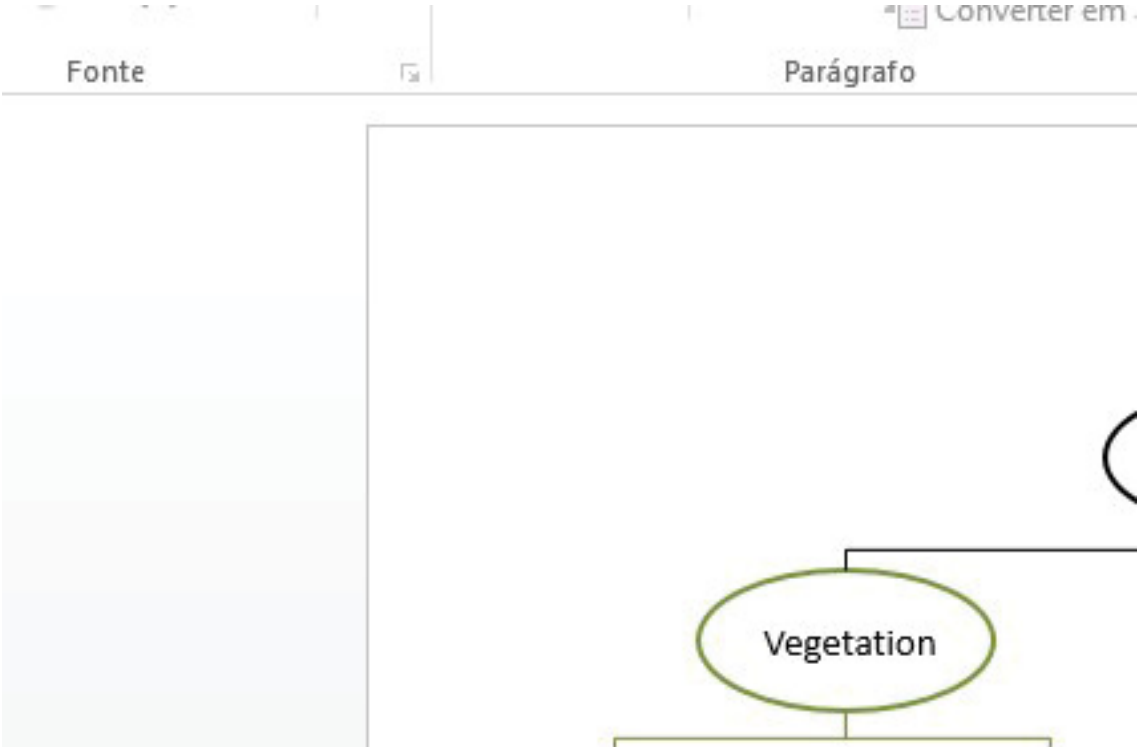


FONT: the author

Based on ongoing discussions in river basin management and the need for harmonization of existing land cover classification systems (Yang et al, 2017) and a global interoperability to use them in environmental physics-based models, nine classes of interest were defined in a first step. The final classes to be represented in that context were water bodies, grass, agriculture, natural forest, reforestation, bare roads, bare land, buildings and paved roads. Bare roads and bare land are the most important areas as regards erosion processes leading to strongly increased loads. Natural forest and reforestation are the classes of interest that represent an elevated infiltration capacity and thus a very low delivery area. Agriculture comprises the agricultural activities that are also important to CP factor choosing. The buildings and paved roads classes are where the urban and industrial areas are located – houses, industrial plants and paved roads, leading to peak surface runoff.

These nine classes were arranged into a hierarchical semantic network (Figure 7) going from a coarser (water and non-water classes) to a finer level (most detailed) of classification (water bodies, grass, agriculture, natural forest, reforestation, bare roads, bare land, buildings and paved roads).

FIGURE 7 - SEMANTIC HIERARCHICAL NETWORK



FONT: the author

As shown in Figure 7, the semantic network can be extended when the spatial resolution is increased and allows distinguishing more classes when level 3 from Table 5 is considered. Nevertheless, based on the works employed to build Tables 1 and 2, it can be assumed that a simpler network would be more useful for environmental modeling to reproduce effects such as:

- diffuse pollution emissions,
- dependence of different land covers on soil infiltration,
- increase in the suspended sediment load in runoff,
- plant cover protection against erosion,
- water interception by vegetation, and
- Increased response of impervious areas in terms of surface runoff.

The final classes used to represent the land cover area could either have a different name or could be in a different hierarchical network disposition. Anyhow, the decision to go ahead with the one developed and showed in figure 6 makes sense due to all the work involved in finding in the literature the most used and suitable classes of interest in this context.

6.2 VOSSOROCA BASIN SEGMENTATION PARAMETERS

The first part of the practical experiment refers to the segmentation within the OBIA approach. The values for the two times mosaics were:

- time 1 (2009) – 100 of scale parameter, 0.5 both for shape and compactness parameters,
- time 2 (2014) – 80 of scale parameter, 0.5 both for shape and compactness parameters.

It is important to stress that only one level of segmentation was used for the benefit of simplifying all other next steps. Anyhow, the one level of segmentation for all mosaics was enough to separate and define the most important classes' objects.

Even though the level of detail in the land cover classes was generalized, the OBIA approach was used for image classification because of the high spatial resolution of the available images. A discussion of using or not this approach and dataset is open, mainly because it is already possible to use multispectral images with lower spatial resolution, free of charge from satellites such as Sentinel 2, together with a pixel-based processing approach if the four classes of level 2 and the water class were to be used. However, other than providing high spatial resolution images, RapidEye satellite constellation has a bigger archive in terms of temporal resolution along with the red edge band having the same spatial resolution as the other spectral bands, which, for instance, do not occur in Sentinel 2.

6.3 RESULTS FROM THE FEATURE SELECTION METHODS

In this section, the results of the feature selection approaches are described and discussed. The experiments were organized as follows: first, a RapidEye mosaic of 2009 was selected to compare the algorithms. Then, after a discussion, it was decided to use the wrapper algorithm to study the evolution of the land cover in the Vossoroca basin between 2009 and 2014.

6.3.1 Feature selection with the traditional SBS[

As explained before, the simple SBS algorithm was applied in order to obtain reference values in terms of feature selection results. The algorithm discards, within

an iterative process, less relevant variables and keeps the best ones, considering a maximal tolerated of error rate.

Table 7 displays the obtained result using the spectral and spatial variables of the 2009 image when only one variable is chosen.

TABLE 6 - FEATURES SELECTED WITH THE SBS METHOD FOR 2009.

Node Classes	Variable	Threshold	Training samples accuracy
Water x Non-water	NDVIr	-0,20	0.850
Veg. x Non-vegetation	NDVIr	0,3	0.533
High veg. x Low veg.	Mean(NIR)	789	0.500
Forest x Reforestation	Brightness	480	0.500
Bare roads x Bare soil	NDVIre	0.5	0.950
Bare land x Imperv. area	NDWI	0.1	0.250
Grass x Agriculture	Density	1.2	0.500
Paved roads x Buildings	Compactness	0.3	0.933

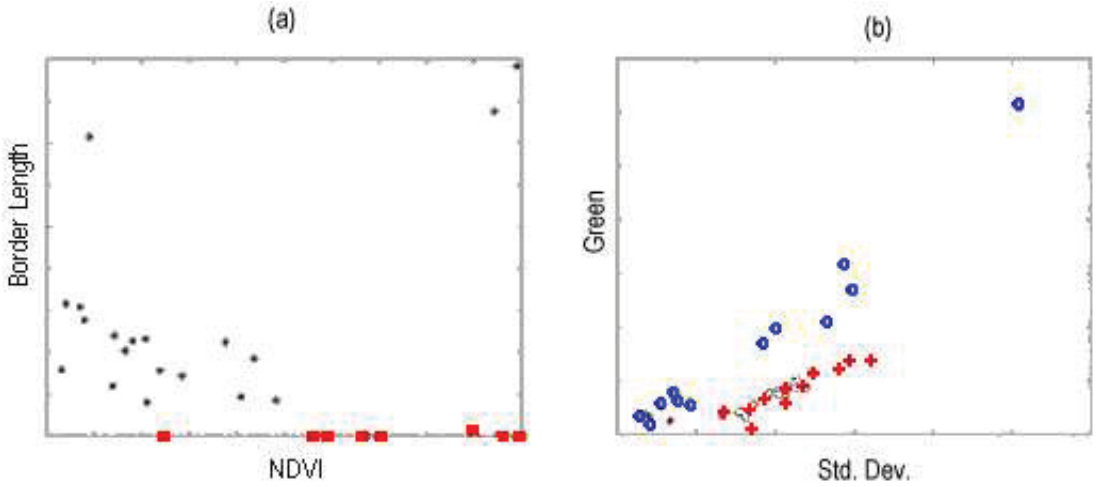
The result regarding the features selected was reasonable and can be partially explained in terms of the classes and variables. For example, the use of a vegetation index is an acceptable choice to identify areas covered by vegetation or to separate water bodies. The use of the near infrared seems also reasonable to discriminate high and low vegetation, because high vegetation appears brighter in the near infrared. To separate paved roads from buildings the SBS algorithm proposes the se of a geometrical property, compactness, which can be used to separate elongated areas that can be roads. It must be pointed out that there are other variables combinations that can produce the same results in terms or error rate.

In the same table, it is presented the accuracy of the classification of the training samples. The values were not satisfactory, for instance, the accuracy between bare soil and impervious area classes was the smallest one, 25%. It must be pointed out that only 30 samples were randomly selected to represent each cluster, so that might have a share on the accuracy results. In the first nodes of the tree, more samples were available, while the number of available samples was reduced as the decision tree become more specific.

6.3.2 The perceptron-based method

In the second series of experiments, the perceptron concept was applied at each node to select the best features. The idea was to use the possibility to select a linear combination of two variables instead of just one, as proposed by the SBS method, in the hope that more complex decisions can be taken at each node. Figure 9a displays an example where only one variable is necessary; in this case, the spatial feature “border length” can be used, where one can readily understand that the decision border was a horizontal line. On the other hand, at some nodes, the decision border was only achieved when using a linear combination of two variables, as displayed in figure 9b, where the decision boundary is not parallel to the axes.

FIGURE 8 - EXAMPLES OF THE COMPUTED SOLUTIONS. (A) THE CLASSES (CLASS ONE REPRESENTED BY THE EMPTY CIRCLE AND CLASS 2 BY THE FILLED CIRCLE) CAN BE SEPARATED USING ONLY ONE FEATURE; (B) A LINEAR COMBINATION OF TWO FEATURES IS NECESSARY TO SEPARATE THE CLASSES



FONT: the author

In some cases, the perceptron converged very quickly, showing that most of the times only one feature would be enough to separate two classes. This happened when one of the features' weights were too small compared to the other one. In other situations, the solution was a fair combination of two variables. The perceptron enables obtaining the decision line as a linear combination of two variables. The obtained results are listed in table 7. The last column in table 7 displays the accuracy obtained with the training samples. In some cases, more than one pair of variables has the same accuracy. Then, the author selected one.

TABLE 7 - PAIRS OF FEATURES AND WEIGHT CHOSEN FOR EACH NODE, ACCORDING TO THE PERCEPTRON METHOD.

Node Classes	Variable	Training samples accuracy
Water x Non-water	R= 11.607 Blue -8098 NDWI -9271	1.00
Veg. x Non-vegetation	R=-12.31 NDVIr + 13.310 NDVIre + 2.00	0.98
High veg. x Low veg.	R= -14.410 Mean(RE) + 4751.6 NDVIr + 5265	0.96
Forest x Reforestation	R= -27.710 NDVIre -15.470 NDWI + 1.00	0.97
Bare roads x Bare soil	R=-17.380 NDVIr + 34.09 NDVIre	0.88
Bare land x Imperv. area	R= -76726 Mean(G) +3078 Mean(RE) + 417	0.97
Grass x Agriculture	R= -19.83 Max. diff. -60.84 NDWI + 12.00	0.86
Paved roads x Buildings	R=6912 length/width -27,66 Brightn +669	1.00

The result was suitable in some cases, when one considers the involved variables. For instance, to separate water, the inclusion of the water index is a good choice, but the use of the mean Blue band as a feature is complex to explain. In the second node, when it comes to separate areas covered by vegetation, two vegetation indexes are selected. As the coefficients are very similar, it can be assumed that the use of just one index would also be enough.

Low and high vegetation were separated using a vegetation index and the red edge data, which is acceptable too. To distinguish natural forest from reforestation, the water index is included. This does not sound reasonable at first, but the combination of a vegetation index and a water index would represent the contrast between the near infrared and two regions of the visible, which can help to detect some vegetated areas. In the last nodes, more spatial information was included, because the classes become more similar in spectral terms and new information is needed.

A weak side of the perceptron method is that it demands a high computer effort to analyze each node. This effort grows very fast with the amount of variables (features and their weights), because the algorithm evaluates all possible pair wise combinations. Therefore, it might turn not very interesting when the number of variables grows.

For information purposes, the elected variables were used to classify the image segments and a global accuracy of 77.5% was obtained.

6.3.3 Wrapper method

In the third experiment, the wrapper method was used to select the best variables for the 2009 and 2014 mosaic images. For this purpose, the JD divergence was computed for the available feature set, considering the two clusters of each node of the semantic hierarchical network developed. Table 8 displays the number selected features that were obtained at each node of the decision tree, considering a JSD distance above 0.5. In table 8, the results obtained using the 2014 image are also included.

TABLE 8 - NUMBER OF SELECTED FEATURES FOR EACH NODE WITH THE WRAPPER METHOD.

Node	2009	2014
Water x Non-water	10	10
Vegetation x Non-vegetation	6	10
High vegetation x Low vegetation	11	6
Bare land x bare roads	1	1
Natural forest x Reforestation	7	7
Grass x Agriculture	5	2
Bare soil x Impervious area	8	7
Buildings x Paved roads	11	15

As expected, the number of features varies for each node. In some cases, the number of features is low, for example to separate bare land and bare roads only one feature was obtained in both dates, respecting the selection method stated in the methodology section. On the other hand, at some nodes, a larger number of possible features was available. For instance, 11 features were selected to distinguish low and high vegetation in 2009.

Table 9 displays the best five features for each node. The features were ranked according to its JSD value. When the JSD is equal, then the overall accuracy (OA) was used as second criteria. Here, again, the results of the 2014 image were included for information and later comparison purposes.

TABLE 9 - FIVE BEST SELECTED VARIABLES FOR EACH NODE ACCORDING TO THE JSD AND OVERALL ACCURACY.

Node	2009			2014		
	Features	JSD	OA	Features	JSD	OA
Water x Non-water	NDWI	0.693	0.971	NDVIre	0.693	0.968
	GLCM Mean (45°)	0.693	0.932	Mean RE	0.693	0.905
	GLCM Mean (90°)	0.693	0.932	GLCM Mean (45°)	0.693	0.948
	GLCM Homog. (0°)	0.693	0.750	GLCM Mean (90°)	0.693	0.948
	NDVIre	0.688	0.966	NDWI	0.680	0.979
	Mean NIR	0.683	0.966	NDVIr	0.680	0.989
	GLCM Mean (0°)	0.675	0.932	Mean NIR	0.680	1.000
	GLCM Mean(all dir.)	0.675	0.932	GLCM Mean (0°)	0.679	0.948
	NDVIr	0.670	0.961	GLCM Mean (all dir.)	0.679	0.948
	Std. deviation NIR	0.670	0.692	GLCM Mean (135°)	0.679	0.948
Vegetation x Non-vegetation	NDVIr	0.643	0.975	GLCM Mean (90°)	0.634	0.705
	NDWI	0.626	0.967	GLCM Mean) (all dir.)	0.634	0.705
	NDVIre	0.620	0.959	GLCM Entropy (all dir.)	0.634	0.705
	Max. diff.	0.607	0.975	GLCM Correlation (135°)	0.634	0.705
	Mean R	0.493	0.891	GLCM Homogeneity (all dir.)	0.623	0.705
				NDVIr	0.597	0.958
	Average branch length	0.364	0.727	Max. diff.	0.577	0.951
				NDVIre	0.571	0.944
				NDWI	0.547	0.937
				Standard deviation R	0.518	0.753
High x Low vegetation	Mean G	0.693	0.994	Mean G	0.656	0.925
	Mean R	0.693	0.994	Mean RE	0.655	0.942
	GLCM StdDev (0°)	0.693	0.813	Brightness	0.644	0.954
	GLDV Ang. 2nd moment (135°)	0.693	0.813	Mean R	0.634	0.890
	GLDV Ang. 2nd moment (all dir.)	0.693	0.813	Mean B	0.581	0.896
	GLCM StdDev (45°)	0.693	0.813	Compactness	0.348	0.705
	GLCM Ang. 2nd moment (all dir.)	0.693	0.813			
	Mean B	0.680	1.000			
	Mean RE	0.676	0.978			
	NDVIr	0.665	0.918			
	Compactness	0.344	0.702			
Bare land x bare roads	NDWI	0.555	0.870	NDWI	0.450	0.753
Natural forest x Reforestation	Max. diff.	0.535	0.830	GLDV Mean (quick 8/11) (45°)	0.6743	0.772
	GLDV Mean (quick 8/11) (all dir.)	0.514	0.815	GLCM Dissimilarity (quick 8/11) (45°)	0.674	0.772
	GLCM Dissimilarity (quick 8/11) (all dir.)	0.514	0.815	GLCM StdDev (quick 8/11) (0°)	0.664	0.777
	NDVIre	0.507	0.778	NDVIre	0.540	0.883
	NDWI	0.502	0.822	Mean NIR	0.514	0.878
	Mean NIR	0.447	0.874	Brightness	0.478	0.789
	Standard deviation B	0.440	0.726	Compactness (polygon)	0.433	0.706
Grass x Agriculture	GLDV Entropy (all dir.)	0.543	0.765	Mean NIR	0.436	0.695
	GLCM Contrast (all dir.)	0.532	0.788	NDVIr	0.382	0.695
	GLDV Contrast (all dir.)	0.532	0.788			

	GLCM StdDev (135°)	0.528	0.742			
	GLCM StdDev (90°)	0.527	0.712			
Bare soil x Bare land	Std. deviation G	0.602	0.934	NDVIre	0.447	0.906
	Std. deviation RE	0.567	0.928	Mean NIR	0.438	0.762
	Std. deviation R	0.539	0.902	NDVIr	0.418	0.902
	Std. deviation B	0.507	0.830	NDWI	0.410	0.855
	Std. deviation NIR	0.494	0.836	Std. deviation RE	0.374	0.923
	Density	0.456	0.784	Density	0.315	0.834
	main line width	0.444	0.823	Length/Width	0.283	0.813
	Width	0.422	0.712	NDVIre	0.447	0.906
				Mean NIR	0.438	0.762
				NDVIr	0.418	0.902
Buildings x Paved roads	GLDV Ang. 2nd moment (45°)	0.693	0.812	Average length of edges (polygon)	0.693	0.9677
	Length/Width	0.693	0.770	GLCM Mean (90°)	0.693	0.9677
	Elliptic Fit	0.693	0.770	Length/Width	0.693	0.9516
	GLCM Entropy (135°)	0.659	0.875	GLCM Mean (0°)	0.693	0.9355
	GLCM Entropy (0°)	0.651	0.854	NDVIr	0.693	0.9194
	GLCM Homogeneity (all dir.)	0.626	0.854	Mean NIR	0.693	0.919
	GLCM Mean (135°)	0.626	0.770	NDWI	0.693	0.919
	Mean NIR	0.623	0.937	Compactness (polygon)	0.693	0.919
	Brightness	0.607	0.937	GLCM Entropy (45°)	0.693	0.919
	Standard deviation RE	0.593	0.895	Standard deviation R	0.693	0.887
	Standard deviation NIR	0.518	0.750	GLCM Entropy (135°)	0.693	0.887
				Max. diff.	0.693	0.871
				Average branch length	0.693	0.758
				Stddev of length of edges (polygon)	0.693	1.000
				GLCM Entropy (90°)	0.658	0.903

From a results comparison between 2009 and 2014, it was possible to see that there are coincidences between the features selected for both dates and same nodes. This is a positive aspect of the wrapper model here proposed. It makes sense that similar variables were chosen for the same area, because the classes that build the clusters were the same. Differences were also expected, mainly due to different illumination or atmospheric conditions when the different mosaic images were acquired.

Taking as example the first node (water x non-water). The spectral features were almost the same (NDVIre, NDWI, NDVIr, and Mean NIR) for both scenes (2009 and 2014). The only difference was the use of “Standard deviation NIR” instead of “NDVIre”. In terms of texture, the selected features are also very similar. For this

node, no spatial features were selected. This can be attributed to the low number of samples, which did not provide enough information about shape.

For the node containing bare soil and impervious area classes, it was selected just one feature in both scenes (NDWI). One first remark is towards these two classes' objects; they are very similar both in spectral and spatial terms. This example shows a classical matter that occurs when dealing with remote sensing image classification. The classes are different in terms of their hydrological properties, but the desired information difference is not visible in the image.

To continue with the classification steps, two approaches were considered to classify the image mosaics. In the first one, only the best feature presented by the feature selection method proposed was used in a thresholding classification. In the second, using the Nearest Neighborhood classification, the best five features, when available, were included into the classifier. After some experiments, it was observed that the computation of the GLCM matrix demands too much computation and time. Therefore, textural variables were not considered even if some of them provided good results.

Table 10 displays the best features for each node and the threshold that was used to perform the first classification in the decision trees for 2009 and 2014.

TABLE 10 - FEATURES AND THRESHOLDS OF 2009 AND 2014 USED IN THE CLASSIFICATION.

	2009		2014	
Node Classes	Variable	Threshold	Variable	Threshold
Water x Non-water	Mean NIR	1314	NDVI – Red band	-0.442
High veg. x Low veg.	Mean– Red band	835	NDVI – Red band	0.330
Forest x Reforestation	Max. Difference	2.83	Mean – Red Edge	1818
Bare roads x Bare soil	Std. dev. Green	261	NDVI – Red Edge	-0.0178
Veg. x Non-vegetation	NDVI (Red)	0.411	NDVI – Red Edge	0.529
Bare land x Imperv. area	NDWI	-0.137	NDWI	-0.0034
Grass x Agriculture	GLCM Contrast (all dir.)	1002	Mean – NIR	5584
Paved roads x Buildings	Mean – NIR	3533	Average length of edges	5.274

Table 11 displays the best five spectral and spatial features used in the second series of classifications. Only the five features with the best JSD values and overall accuracies above 70% were considered.

TABLE 11 - BEST FIVE FEATURES FOR EACH NODE OF THE DECISION TREE FOR 2009 AND 2014.

Classes	Features – 2009	Features – 2014
Water x Non-water	NDWI, NDVI - Red edge band, Mean – NIR band, and NDVI – Red band	NDVI - Red edge band, Mean – Red edge band, NDWI, NDVI – Red band, and Mean Red band
High veg. x Low veg.	Mean – Red band, Mean – Green band, Mean -Red Edge band, and NDVI – Red band	Mean – Green band, Mean -Red Edge band, and Brightness.
Forest x Reforestation	Max. difference	NDVI - Red edge band
Bare roads x Bare soil	Std. dev. – Green band, Std. dev. Red edge band, and Std. dev. – Red band	NDVI - Red edge band, Std. dev. Red edge band, and NDVI – Red band
Veg. x Non-vegetation	NDVI – Red band, Max. difference, NDWI, and NDVI - Red edge band	NDVI – Red band, Max. difference, NDVI - Red edge band, and NDWI
Bare land x Imperv. area	NDWI	NDWI
Grass x Agriculture	GLCM Contrast (all dir.)	Average length of edges (polygon) and Std. dev. of length of edges (polygon)
Paved roads x Buildings	Mean – NIR, Brightness, and Std. dev. – Red edge band	Mean – NIR band, NDVI - Red band, and NDWI

6.3.4 Comparison 2009 – between methods

Table 12 shows a summary of the selected features using the four options, SBS, perceptron, wrapper using one variable and wrapper using the best features. The selected features were not the same, but some similarities can be seen in terms of the classes' spectral properties. For instance, to separate vegetated areas, the tendency is to use a vegetation index, while to separate water the methods proposed the use of a water index of a vegetation index. The use of vegetation indexes was frequent when it was necessary to classify two types of vegetation, which indicates that the methods respected spectral curve's behavior of these classes.

On one hand, the simple SBS method is very fast in terms of time processing but, as their criteria are very simple, the result was not suitable when applied to the

other samples. The perceptron method was adequate in terms of performance, but it is very time consuming. Therefore, the wrapper method was selected to classify the Vossoroca basin image and performs a temporal analysis. As the computational effort is the same to obtain one or five features, test were performed using a single feature or a set of best features.

TABLE 12 - COMPARISON OF SELECTED VARIABLES.

Node Classes	Perceptron	SBS	Wrapper 1	Wrapper 5
Water x Non-water	Blue; NDWI	NDVIr	NIR	NDWI, NDVI - Red edge, NIR, and NDVI–Red
Veg. x Non-vegetation	NDVIr NDVIre	NDVIr	Red	NDVI – Red, Max. difference, NDWI, and NDVI - Red edge
High veg. x Low veg.	Mean(NIR) Mean(R)	NIR	Max. Difference	Red, Green, Red Edge, and NDVI – Red
Forest x Reforestation	NDVIr ; NDVIre	Brightness	Std.dev.(Green)	Max. difference
Bare roads x Bare soil	NDVIr ; NDVIre	NDVIre	NDVI (Red)	NDWI
Bare land x Imperv. area	Mean(G) Mean(RE)	NDWI	NDWI	NDWI
Grass x Agriculture	Max. diff; NDWI	Density	GLCM Contrast (all dir.)	GLCM Contrast (all dir.)
Paved roads x Buildings	length/width; Brightness	Compactness	Mean – NIR	NIR, Brightness, and Std.dev.(Red edge)

6.3.5 Classification with JSD features

In the next step, the selected variables were used to classify the images and evaluate the quality of the produced thematic maps. Four classifications were performed. For each date (2009 and 2014) two classifications were necessary: with one feature and with best features available (max. of 5). Table 13 and 14 contain the confusion matrices for the image mosaic of 2009 using the best feature (Table 6) and the five selected features (table 7). To ease the comparison between the results, no text was places between tables 6 and 7, comments were placed afterwards.

TABLE 13 - CONFUSION MATRIX OF 2009 USING ONE FEATURE.

Classes	1	2	3	4	5	6	7	8	9
Water (1)	15	0	0	0	0	2	0	0	0
Bare land (2)	0	27	4	0	0	0	0	0	0
Bare roads (3)	0	2	24	1	0	0	0	0	0
Buildings (4)	0	0	0	14	0	0	0	0	0
Paved roads (5)	0	1	0	0	15	0	0	0	0
Natural forest (6)	0	0	0	0	0	27	0	0	4
Reforestation (7)	0	0	0	0	0	1	30	0	0
Grass (8)	0	0	2	0	0	0	0	3	4
Agriculture (9)	0	0	0	0	0	0	0	12	7

TABLE 14 - CONFUSION MATRIX OF 2009 USING THE BEST FIVE FEATURES.

Classes	1	2	3	4	5	6	7	8	9
Water (1)	15	0	0	0	0	0	0	0	0
Bare land (2)	0	28	0	0	0	0	0	0	1
Bare roads (3)	0	2	27	1	2	0	0	0	0
Paved roads (4)	0	0	0	12	0	0	0	0	0
Buildings (5)	0	0	1	2	13	0	0	0	0
Natural forest (6)	0	0	0	0	0	26	1	0	0
Reforestation (7)	0	0	0	0	0	4	29	0	0
Agriculture (8)	0	0	0	0	0	0	0	12	2
Grass (9)	0	0	2	0	0	0	0	3	12

There are two interesting points to be highlighted regarding some classes and the classification methods: there are no misclassified samples for water in neither of the classifications. Water bodies are easy to classify because of theie high electromagnetic radiation absorption, and therefore, the decision based on the NIR band or the NDVI is very stable. The second noticeable fact is that, the agriculture classification performed properly in 2009 but in 2014 this class presented misclassification with the grass class.

The same classifications and comparisons were performed using the 2014 mosaic image. The confusion matrices for 2014 are presented in Tables 15 and 16. The number of test samples varied between 15 and 30 and it is not uniform. The number of samples depends on the classes' frequency in the images.

Bare land and bare roads were the classes with higher amounts of misclassification. In hydrological terms, these classes are different because bare roads are compacted by the traffic, which reduces the permeability. In spectral terms, bare land and bare road are very similar because they are covered by dirt. This confusion could be improved if images with higher spatial resolution are used. In this case, using RapidEye images, the lack of information was reflected by the fact that just one feature was selected for the associated node. The differences in the

misclassified samples in both experiments can also be caused by wrong decisions in the previous nodes.

TABLE 15 - CONFUSION MATRIX OF 2014 USING ONE FEATURE.

Classes	1	2	3	4	5	6	7	8	9
Water (1)	15	0	0	0	0	0	0	0	0
Bare land (2)	0	22	1	0	0	0	0	0	0
Bare roads (3)	0	4	27	0	2	0	0	0	0
Paved roads (4)	0	2	0	14	1	0	0	0	0
Buildings (5)	0	2	1	1	12	0	0	0	0
Natural forest (6)	0	0	0	0	0	29	2	0	0
Reforestation (7)	0	0	0	0	0	1	28	0	0
Agriculture (8)	0	0	0	0	0	0	0	14	3
Grass (9)	0	0	1	0	0	0	0	1	12

TABLE 16 - CONFUSION MATRIX OF 2014 USING THE BEST FIVE FEATURES.

Classes	1	2	3	4	5	6	7	8	9
Water (1)	15	0	0	1	0	0	0	0	0
Bare land (2)	0	29	0	1	3	0	0	0	1
Bare roads (3)	0	0	26	0	0	0	0	0	0
Paved roads (4)	0	0	0	13	0	0	0	0	0
Buildings (5)	0	1	4	0	12	0	0	0	0
Natural forest (6)	0	0	0	0	0	29	3	0	0
Reforestation (7)	0	0	0	0	0	1	27	0	0
Agriculture (8)	0	0	0	0	0	0	0	13	2
Grass (9)	0	0	0	0	0	0	0	2	12

The overall accuracies of the classifications are relative good, as displayed in table 17. They vary from 83% to 90%, which might be considered a remarkable amount for this type of measure. Higher overall accuracy index values were achieved with five features, but the values using one variable were not very different when compared result-wise. For example, in 2014 it can be considered that the accuracies were equal using one of the best five features. Furthermore, the Kappa indexes varied from 0.80 to 0.89. The lowest overall accuracy was observed in 2009 when only one feature was used to perform the thresholding classification. Again, in 2014, the use of one or five variables made no difference. This comparison points out that the use of more variables could increase the accuracy, but it also showed that increased overall accuracy value might not be significant. Using a large number of features demands more processing effort and this fact can support the preference for one variable instead of five or the best one, or even classification method.

TABLE 17 - OVERALL ACCURACY AND KAPPA INDEX OF THE CLASSIFICATIONS.

	Overall accuracy (%)		Kappa	
Date	One feature	Five features	One feature	Five features
2009	83	89	0.80	0.88
2014	89	90	0.89	0.89

The analysis was extended to the comparison of the producer's and user's accuracies, as shown in tables 18 and 19 (for 2009 and 2014 respectively).

TABLE 18 - PRODUCER'S ACCURACY FOR 2009 AND 2014 USING ONE OR A SET OF FEATURES.

	2009		2014	
Classes	One feature	Best features	One feature	Best features
Water	1.00	1.00	1.00	1.00
Bare land	0.90	0.93	0.73	0.97
Bare roads	0.80	0.90	0.90	0.87
Paved roads	0.93	0.80	0.93	0.87
Buildings	1.00	0.87	0.80	0.80
Natural forest	0.90	0.87	0.97	0.97
Reforestation	1.00	0.97	0.93	0.90
Agriculture	0.20	0.80	0.93	0.87
Grass	0.47	0.80	0.80	0.80

TABLE 19 - USER'S ACCURACY FOR 2009 AND 2014 USING ONE OR A SET OF FEATURES.

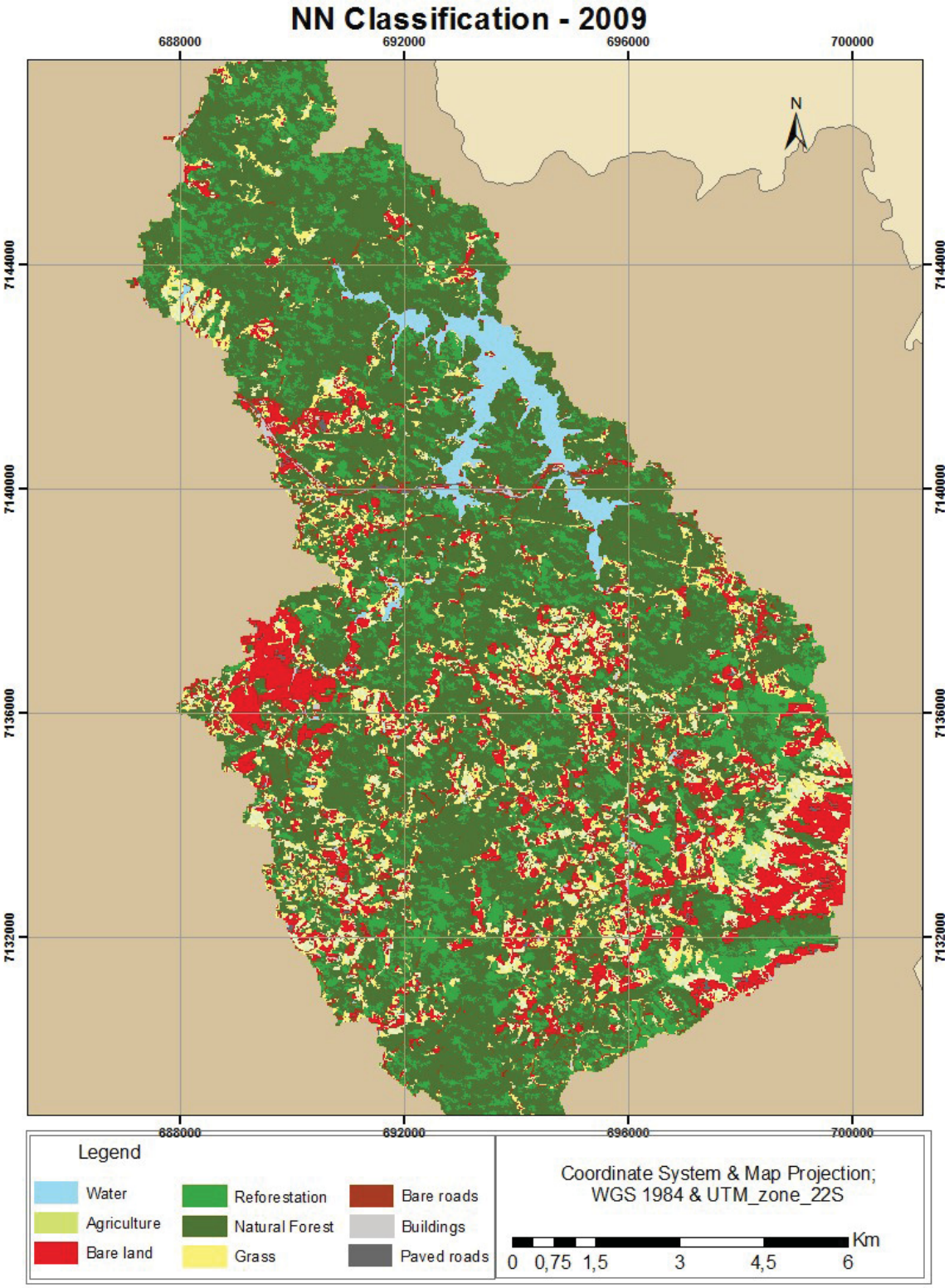
	2009		2014	
Classes	One feature	Best features	One feature	Best features
Water	0.88	1.00	1.00	0.94
Bare land	0.87	0.96	0.96	0.85
Bare roads	0.89	0.87	0.82	1.00
Paved roads	1.00	1.00	0.82	1.00
Buildings	0.94	0.81	0.75	0.70
Natural forest	0.87	0.96	0.93	0.91
Reforestation	0.97	0.88	0.96	0.96
Agriculture	0.33	0.86	0.82	0.87
Grass	0.37	0.70	0.86	0.86
Minimum	0,33	0,7	0,75	0,7
Maximum	1.00	1.00	1.00	1.00
Mean	0,79	0,89	0,88	0,90

The user's accuracy values were satisfactory, reaching more than 80% in most cases. There was a similar trend along time, which means that lower values are visible for the same classes in both dates, which is coherent. The worst results were obtained when the 2009 mosaic image was classified using only one feature per node. Especially agriculture and grass classes were not adequately classified. Fact that did not happen in 2014, when the values lied close to 0,85. Considering agriculture and grass in 2019 in table 19, the remaining values were above 0.70, also

a good indicator. Regarding the number of features, the use of more variables increased the user's accuracy. The lowest performance observed in 2009 for grass and agriculture was corrected when five features were used when the Nearest Neighborhood classification was performed.

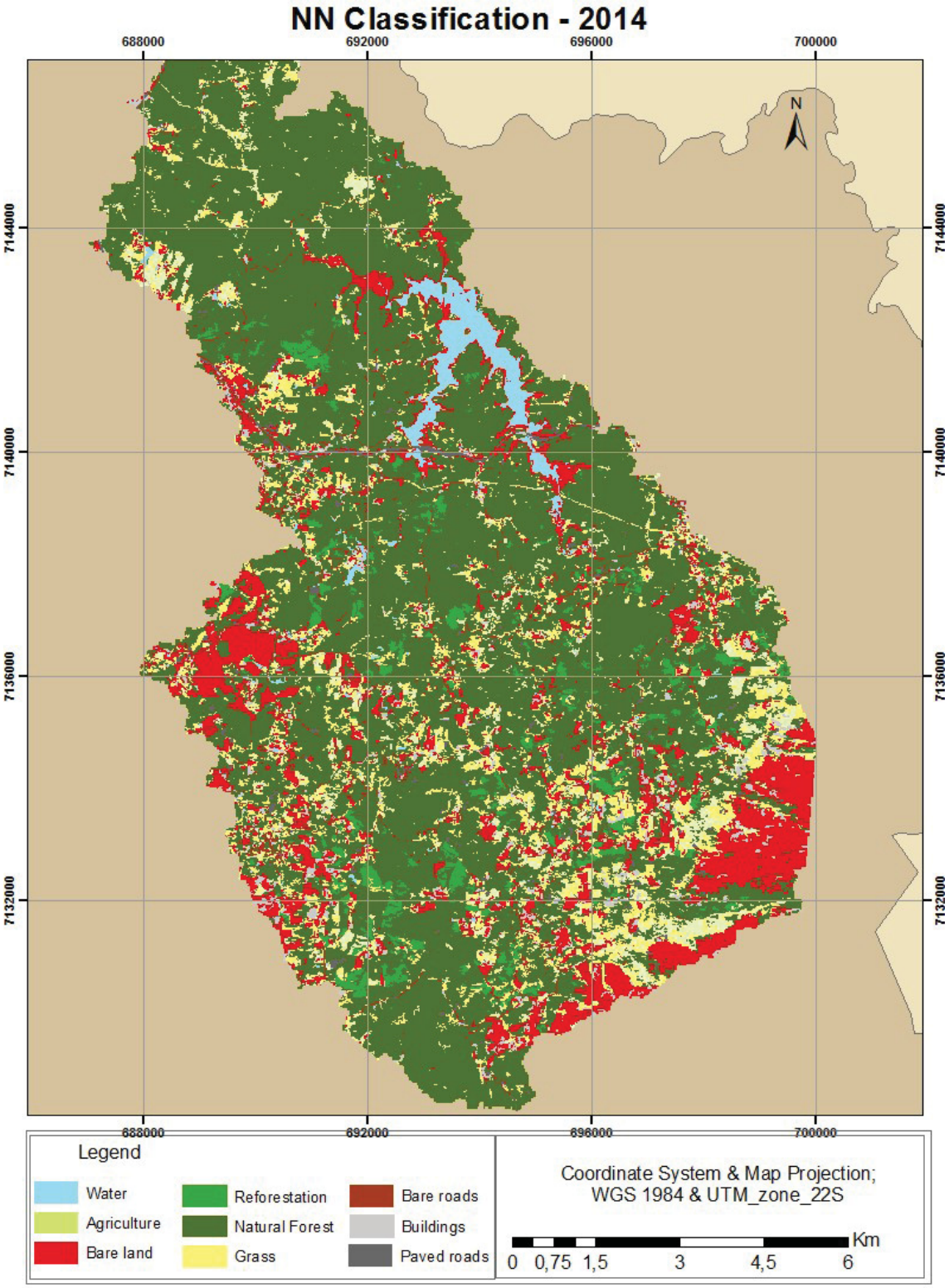
As the best overall accuracies and Kappa values were performed when using the NN classification and the best features, the two thematic maps yielded from 2009 and 2014 and this method are presented below.

FIGURE 9 - THEMATIC MAP FOR 2014 MOSAIC IMAGE



FONT: the author

FIGURE 10 - THEMATIC MAP FOR 2009 MOSAIC IMAGE



FONT: the author

Visually, the comparison between the two maps might turn a bit complicated for various reasons, to help with some conclusions; some facts can be stated. Firstly, table 20 with the area results is presented:

TABLE 20 - COMPARISON BETWEEN LAND COVER QUANTITIES.

Classes	2009		2014	
	One feature (km^2)	Best features (km^2)	One feature (km^2)	Best features (km^2)
Agriculture	6.85	8.21	3.14	6.65
Bare land	14.05	14.18	4.39	15.73
Bare road	4.43	5.16	9.02	4.63
Buildings	0.19	0.61	1.97	2.07
Reforestation	11.25	20.86	9.92	4.21
Natural forest	84.21	72.12	91.81	86.68
Grass	8.24	10.31	10.32	12.17
Paved roads	0.82	1.00	3.36	1.18
Water	6.55	4.08	2.66	2.87

All the areas are in square kilometers. There are naturally multiple changes between the four column values, but they can be explained for some reasons. Regarding the main water body – the reservoir, its shape and the bare land area around it in the two maps, it is possible to spatially see a decrease of the reservoir area between 2009 and 2004. Because of this decrease in the reservoir mirror area, many cities in Paraná state near the reservoir had trouble with water supply. This was also a reflection of the reservoir volume decrease. The same problem occurred in São Paulo state at the same period.

Concerning the natural forest classification, it was possible to remark that, there are more areas labelled as such class in 2014 than in 2009 – 72.12 square kilometers in 2009 and in 86.68 in 2014. One of the biggest characteristics regarding this class was based on the illumination, the more illuminated the “high vegetation” segments were, the bigger the chance it was going to be put in the reforestation class, mainly because the texture in such segment surface was less roughened; otherwise, the segment would be more likely to be labelled as natural forest. Along those lines, and the inconsistence of having a bigger amount of natural forest land cover years after, a misclassification might be considered due to differences in illumination and atmospheric conditions between the two scene dates.

7 CONCLUSION

In this work, three topics were developed. The first one is the proposal of a thematic class schema to support image classification within the hydrology context. The second theme refers to feature selection within the OBIA classification process and the last one is the multitemporal study of the Vossoroca basin between 2009 and 2014.

Concerning classes' definition with a hydrological meaning, the study showed two contributions. The first one concerning the definition of a semantic hierarchical network of land cover classes that could be applied to different test areas and produce thematic maps that could also be compared in hydrological studies. The second one is the fact that the semantic network can be adapted to different image resolutions. In the common case, a lower resolution is enough, because the hydrologist faces the problem of poor data sources. For instance, not always, a detailed soil map is available and therefore a detailed land cover map becomes useless. In such cases, the user would rather to restrict the proposed semantic hierarchical network to lower nodes, avoiding wasting time with a detailed map. Nevertheless, when the study needs more detailed land cover information, the whole network could be used.

In a second part, three feature selection approaches were compared: the simple SBS method, a method based in the perceptron principle that was proposed by the author and the use of the JS divergence in the wrapper model method, also proposed by the author. All that, to test he pros and cons of these different feature selection approaches, as well as comparing results from thresholding and Nearest Neighborhood classifications.

The first advantage regarding the use of perceptron for the methods developed in this work is that, this given technique does not require any prior statistical assumption on data distribution, the assumption of normality within the classes, for instance. This is an important point, mainly because when using the OBIA approach for image classification purposes, pixels are grouped together in the segmentation step, reducing the number of available samples. Nevertheless, the pair wise selection based in the perceptron principle demands higher computational effort comparing to the other ones, because it consists of an exhaustive search within a bidimensional space. As the size of the feature space increases substantially

including the spatial features, the number of possible combinations also increases significantly. The advantage of these methods is that they can find solutions to the classes' separation problem combining more than one feature in a linear discriminant function. Besides, the use of the perceptron with a single input proved to achieve results that can be compared to those using a pair of variables as input, with lower effort.

Concerning the feature selection using wrapper models and the JS divergence, the results in both overall accuracy and Kappa values were very similar to the perceptron one. Which leads to the conclusion that, the developed feature selection had a good performance. Another point to be addressed is towards the classification approaches used from the selected features. The best values for user's and producer's accuracies were given when the best features were used inside the NN classification. In relation to the features selected using the wrapper model, for each node the features selected made sense with the land cover they were representing, which means that, they were chosen based on the behavior of the classes spectral curve.

Now, changing the topic to the third part of this work, the land cover evolution in the Vossoroca basin, other than care information of the different scenes, it was important to test if the methodology using the wrapper model developed and the semantic hierarchical network was going to make sense. From the discussions regarding the problem with water supply in 2014, the maps from 2009 and 2014 showed a difference between the water bodies area, being 4.08 square kilometers in 2009 and 2.86 in 2014, representing properly the phenomenon.

It is important to highlight that OBIA is widely used on high spatial resolution imagery not only with the feature selection methods applied in this work but also with other methods carrying the buzzwords inside Machine Learning, such as support vector machine and random forest. The point is that the classifications were made for the area of interest to test an up-to-date and universal framework that would be useful and provide accurate results on estimating soil loss in river basins. This goal was achieved, since the accuracy assessment results showed overall accuracies of above 90% for both mosaic images and of above 80% for producer and user's accuracy for the most part of the classes of interest. One interesting aspect of the level of detail concerning the classes and the semantic hierarchical network is that,

from a coarser layer, the land cover classes can be the same even though it is a different country.

To wrapper the conclusions, it would be recommended to use the semantic hierarchical network used in this study for other types of images, from others satellite sensor, as well as different spatial resolution.

Multiple aspects could be improved regarding the methodology and the data source, but the one of the most important regarding the wrapper model, would be to improve it towards better results when the set of features is smaller and the way the set of samples should be selected.

REFERÊNCIAS

AGUILAR, M. A., VICENTE, R., AGUILAR, F. J., FERNÁNDEZ, A. AND., SALDAÑA, M. M. Optimizing Object-based Classification In Urban Environments Using Very High Resolution Geoeye-1 Imagery. **ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences**. I-7. p.99–104. 2012.

ANDERSON, J. R., HARDY, E.E., ROACH, J. T. AND., WITMER, R.E. A Land Use Andand Land Cover Classification System For Use With Remote Sensor Data. 1976. Available at: <http://www.pbcgis.com/data_basics/anderson.pdf>. Accessed: 9 May. 2019.

BAATZ, M. and SCHÄPE, A. Multiresolution Segmentation: an optimization approach for high quality multiscale image segmentation. **Proceedings of Angewandte Geographische Informationsverarbeitung XII: Beiträgezum AGIT-Symposium**. Salzburg, Austria, 5-7 July, 2000. Heidelberg: Hebert Wichmann Verlag, p.12-23.

Barbosa, A.F.; Oliveira, E.F. De; Mioto, A.L.; Paranhos F., A.C. The application of the Universal Soil Loss Equation b using free and available softwares. **An. do Instit. De Geoc**, UFRJ, v. 38(1), p. 170-179, 2015. Available at: <http://dx.doi.org/10.11137/2015_1_170_179> .

BARTENHAGEN, C., KLEIN, H.-. U., RUCKERT, C., JIANG, X. AND., DUGAS, M. Comparative study of unsupervised dimension reduction techniques for the visualization of microarray gene expression data. **BMC Bioinformatics**, 11(1), 2010.

BECK, HYLKE E.; ZIMMERMANN, NIKLAUS E.; MCVICAR, TIM R.; VERGOPOLAN, NOEMI; BERG, ALEXIS; WOOD, ERIC F. Present and future Köppen-Geiger climate classification maps at 1-km resolution. **Scientific data**, *v. 5*, p. 180-214, 2018. DOI: 10.1038/sdata.2018.214.

BENAVIDEZ, R..; JACKSON, B.; MAXWELL, D.; NORTON, K. A review of the (Revised) Universal Soil Loss Equation (R/USLE): with a view to increasing its global applicability and improving soil loss estimates. ***Hydrol. Earth Syst. Sci. Discuss.*** in review, 2018. Available at: https://doi.org/10.5194/hess-2018-68.

BESKOW, S.; MELLO, C. R.; NORTON, L. D.; CURI, N.; VIOLA, M. R.; AVANZI, J. C. Soil erosion prediction in the Grande River Basin, Brazil using distributed modeling. ***CATENA***, *v. 79 (1),* p. 49–59, 2009. DOI: 10.1016/j.catena.2009.05.010.

BORRELLI, P.; ROBINSON, D. A.; FLEISCHER, L. R.; LUGATO, E.; BALLABIO, C.; ALEWELL, C. ET AL. An assessment of the global impact of 21st century land use change on soil erosion. ***Nature communications****, v. 8 (1)*, pp. 2013, 2017. DOI: 10.1038/s41467-017-02142-7.

BRAUMAN, KATE A.; DAILY, GRETCHEN C.; DUARTE, T. KA'EO; MOONEY, HAROLD A. The Nature and Value of Ecosystem Services: An Overview Highlighting Hydrologic Services. ***Annu. Rev. Environ. Resour***, *v. 32 (1)*, p. 67–98, 2007. DOI: 10.1146/annurev.energy.32.031306.102758.

BUENO, C.R.P.; STEIN, D.P. Potencial natural e antrópico de erosão na região de Brotas, Estado de São Paulo. ***Acta Scient. Agron.*** *v. 26*, pp. 1-5, 2004. DOI: 10.4025/actasciagron.v26i1.1946

CARLEER, A.; WOLFF, E. Change detection for updates of vector database through region-based classification of VHR satellite data. **SPIE Remote Sensing Europe, Proceedings of SPIE 6749**, Florence, Italy, September 17-21, p. 674911-1-10. DOI: 10.1117/12.737910.

CEC [Commission of the European Communities] 1993. **CORINE Land Cover - Guide technique.** Brussels.

DEMIREL, M.C.; MAI, J.; MENDIGUREN, G.; KOCH, J.; SAMANIEGO, L.; STISEN, S. Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. ***Hydrol. Earth Sst. Sci.*** *v. 22*, p. 1299-1315, 2018. DOI: 10.5194/hess-22-1299-2018.

DI GREGRIO, A.; JANSEN, J.M.L. Land Cover Classification System (LCCS): Classification Concepts and User Manual for Software Version 1; **FAO Land and Water Development Division:** Rome, Italy, 2000; ISBN:9788578110796

ENDRES, D., M.; SCHINDELIN, J., E. A new metric for probability. **IEEE Transactions on Information Theory**, Vol. 49, Number 7, July 2003.

FUCHS, S.; KAISER, M.; KIEMLE, L.; KITTLAUS, S.; ROTHVOSS, S.; TOSHOVSKI, S.; WAGNER, A.; WANDER, R.; WEBER, T.; ZIEGLER, S. Modeling of Regionalized Emissions (MoRE) into Water Bodies: An Open-Source River Basin Management System. ***Water***, *v. 9*, p. 239, 2017. DOI: 10.3390/w9040239.

GAO, B. C. NDWI: A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote Sensing**, 58, p. 257-266, 1996.

GASCA, E., SÁNCHEZ, J. S. AND ALONSO, R.. Eliminating redundancy and irrelevance using a new MLP-based feature selection method. **Pattern Recognition,** 39(2), p.313–315, 2006.

GENG, X.; SUN, K.; JI, L. Band selection for target detection in hyperspectral imagery using sparse CEM. **Remote Sensing Letters**, 5:12, p. 1022-1031, 2014.

GITELSON, A. A.; KAUFMAN, Y. J.; MERZLYAK, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sens Environ**, 58, p. 289-298, 1996.

GONZALEZ, R. and WOODS, R. **Digital image processing.** 2nd Edition, Prentice Hall, Upper Saddle River, 2002.

GONZALEZ, R. C.; WOODS, R. **Processamento de Imagens Digitais.** São Paulo, Editora Edgard Blucher.

GUO, B.; GUNN, S. R.; DAMPER, R. I.; NELSON, J. D. B. Band Selection for Hyperspectral Image Classification Using Mutual Information. **IEEE Geosci. Remote Sens.** Lett. 2006(3): 522–526, (2006).

HABERMANN, M., FRÉMONT, V., SHIGUEMORI, E. Unsupervised Hyperspectral Band Selection Using Clustering and Single-layer Neural Network. **Revue Française de Photogrammétrie et de Télédétection, Société Française de Photogrammétrie et de Télédétection**, In press, p.33-42, 2018.

HAERTEL, V., LANGREBE, D. A. On the classification of classes with nearly equal spectral response in remote sensing hyperspectral image data. **IEEE Transactions on Geoscience and Remote Sensing,** 37(5), p.2374–2386, 1999.

HAMEDIANTAR, A. and SHAFRI, H. Z. M. Integrated Approach Using Data Mining-Based Decision Tree and Object-Based Image Analysis for High Resolution Urban Mapping of WorldView-2 Satellite Sensor Data. **Jour. of Applied Rem. Sens.**, v. 10(2), pp. 025001-1 - 025001-21, 2016. DOI: 10.1117/1.JRS.10.025001.

HARALICK, R. M.; SHANMUGAM, K. AND DINSTEIN, I. 'Textural features for Image Classification', ***IEEE Trans. On Systems, Man and Cybernetics***, 3(6), pp. 610-621, 1973.

HAMEL, P.; CHAPLIN-KRAMER, R.; SIM, S.; MUELLER, C. A new approach to modeling the sediment retention service (InVEST 3.0). Case study of the Cape Fear catchment, North Carolina, USA. ***The Science of the total environment***, *v. 524-525*, pp. 166–177, 2015. DOI: 10.1016/j.scitotenv.2015.04.027.

HUGHES, G. On the mean accuracy of statistical pattern recognizers. **IEEE Transactions on Information Theory**, 14(1), pp.55–63, 1968.

JUNG, R.; EHLERS, M. Comparison of Two Feature Selection Methods for the Separability Analysis of Intertidal Sediments with Spectrometric Datasets in the German Wadden Sea. ***Intern. Journal of Applied Earth Observ. and Geoinf***, *v. 52*, pp. 175-19, 2016. DOI: 10.1016/j.jag.2016.06.009.

LEE, G.S.; LEE, K.H. Scaling effect for estimating soil loss on the RUSLE model using remotely sensed geospatial data in Korea. ***Hydrol. Earth Syst. Sci. Discuss***, v. 3, p. 135-157, 2006. DOI: 10.5194/hessd-3-135-2006.

LEI, C.; ZHU, L. Spatio-temporal variability of land-use/land cover change (LULCC) within the Huron river: Effects on stream flows. **Clim. Ris Manag**, *v. 19*, pp. 35-47, 2018. DOI: 10.1016/j.crm.2017.09.002.

MAHMOUDI, F. T., SAMADZADEGAN, F., REINARTZ, P. Object Oriented Image Analysis Based on Multi-Agent Recognition. **System. Comput. & Geosc**., v. 54, pp. 219-230, 2013. DOI: 10.1016/j.cageo.2012.12.007.

MELO, D. H. C. T. B. Uso de Dados Ikonos II na Análise Urbana: testes operacionais na zona leste de São Paulo. São José dos Campos, Brazil. 2002. Master thesis, INPE (National Institute of Spatial Research).

NOVO, E. M. L. de M. Sensoriamento Remoto: Princípios e Aplicações, 1st ed.; Edgar Blücher Ltda: São José dos Campos, Brazil, 1989. ISBN: 8521208359.

PATEL, B. R.; RANA, K. K. A Survey on Decision Tree Algorithm for Classification. IJEDR 2(1):1-5, 2014.

PERSELLO, C. and BRUZZONE, L. Kernel-Based Domain-Invariant Feature Selection in Hyperspectral Images for Transfer Learning. **IEEE Trans. on Geosc. and Rem. Sens.**, v. 54(5), pp. 2615-2626, 2016. DOI: 10.1109/TGRS.2015.2503885.

PU, R.; BELL, S. Mapping Seagrass Coverage and Spatial Patterns with High Spatial Resolution IKONOS Imagery. ***Intern. Journal of Applied Earth Observ. and Geoinf.***, *v. 54*, p. 145-158, 2017. DOI: 10.1109/IGARSS.2016.7730998.

RAGAN, R.M.; JACKSON, T.J. Runoff Synthesis Using Landsat and SCS Model. ***Journal of Hydraul. Div.*** *v. 106(HY5)*, pp. 667-678, 1980.

RIBEIRO SILVA, C. R. Uso de Algoritmos Genéticos como Redutor de Dimensionalidade na Classificação de Imagens Hiperespectrias. 2006. Dissertação (mestrado) - Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas.

RIZEEI, H.M.; PRADHAN, B.; SAHARKHIZ, M.A. Surface runoff prediction regarding LULC and climate dynamics using coupled LTM, optimized ARIMA and GIS-based SCS-CN models in tropical region. ***Arabian Journ. of Geosc.***, *v. 11:53*, pp. 1-16, 2018. DOI: 10.1007/s12517-018-3397-6.

RUCK DW,, D. W., ROGERS SK,, S. K., KABRISKY, M. Feature selection using a multilayer perceptron. **Neural Network Computing** 2: pp. 40–48, 1990.

SAFAVIAN, S. R.; LANDGREBE, D. A survey of decision tree classifier methodology. IEEE Transactions on Systems. Systems, Man and Cybernetics, **IEEE Transactions** on. 21. pp. 660 - 674, 1991. DOI: 10.1109/21.97458.

SERPICO, S. B.;., D'INCA, M., MELGANI, F.;., MOSER, G. Comparison of feature reduction techniques for classification of hyperspectral remote-sensing data, Proc. SPIE 4885, **Image and Signal Processing for Remote Sensing VIII**, 13 March 2003.

SHI, Z. H., CAI, C. F., DING, S. W., LI, Z. X., WANG, T. W. AND SUN, Z. C. Assessment of Erosion Risk with the Rusle and Gis in the Middle and Lower Reaches of Hanjiang River. **Proceedings of the 12th ISCO Conference**, Beijing, China, May 26-31, 2002.

SOHN, K.; ZHOU, G.; LEE, C.; LEE, H. Learning and Selecting Features Jointly with Point-wise Gated Boltzmann Machines. **Proceedings of the 30th International Conference on Machine Learning**, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28.

TANG, J.; ALELYANI, S. AND LIU, H. **Feature Selection for Classification: a review. In *Data Classification:*** *Algorithms and Aplications*. 1st ed.; Aggarwal, C. (editor); Chapman and Hall/CRC Press, United Kingdom, 2014, pp. 37. ISBN: 9781466586741.

TEMESGEN, G.; TULU, T.; ARGAW, M.; WORQLUL, A.W. Modelling the hydrological impacts of land use/ land cover changes in the Andassa watershed, Blue Nile Basin, Ethiopia. ***Sci. of the Total Environ.***, *v. 619-620*, pp. 1394-1408, 2018. DOI: 10.1016/j.scitotenv.2017.11.191.

TUCCI, C.E.M. ***Hidrologia:*** *ciência e aplicação*, 2nd ed.; Editora da Universidade/UFRGS: Porto Alegre, Brazil, 1993. ISBN: 85-7025-298-6.

VAN COILLIE, F. M. B.; VERBEKE, L. P. C AND DE WULF, R. R. Feature Selection by Genetic Algorithms in Object-Based Classification of IKONOS Imagery for Forest Mapping in Flanders, Belgium. ***Rem. Sens. of Envir***, *v. 110*, pp. 476-487, 2007. DOI: 10.1016/j.rse.2007.03.020.

XIE F, LI, F., LEI C., KE, L. Representative Band Selection for Hyperspectral Image Classification. **International Journal of Geo-Information**. ISPRS Int. J. Geo-Inf, 7(338):1-19, 2007.

XIURUI, G., VITASI, S., LUYAN J., YONGCHAO, Z. A Fast Volume-Gradient-Based Band Selection Method for Hyperspectral Image. Geoscience and Remote Sensing, **IEEE Transactions on**. 52, pp. 7111-7119, 2014. DOI: 10.1109/TGRS.2014.2307880.

WANG, X. Y., GUO, Y. G., HE, J., DU, L. T. Fusion of HJ1B and ALOS PALSAR Data for Land Cover Classification Using Machine Learning Methods. **Intern. Journal of Applied Earth Observ. and Geoinf.**, v. 52, pp. 193-203, 2016. DOI: 10.1016/j.jag.2016.06.014.

WEINBERGER, K. Q.;., SAUL, L. K. An Introduction to Nonlinear Dimensionality Reduction by Maximum Variance Unfolding. **National Conference on Artificial Intelligence (AAAI)**, Nectar paper, Boston MA. 2006.

WELDE, K.; GEBREMARIAM, B. Effect of land cover dynamics on hydrological response of watershed Case study of Tekeze Dam watershed, northern Ethiopia. Intern. ***Soil and Water Cons. Res.***, *v. 5(1)*, pp. 1-16, 2017. DOI: 10.1016/j.iswcr.2017.03.002.

WISCHMEIER, W.H.; SMITH, D.D. Predicting Rainfall Erosion Losses – A Guide to Conservation Planning; United States Dept. of Agriculture. Handbook; Science and Education Administration, U.S. Department of Agriculture: Washington, DC, USA, 1978.

WOHL, ELLEN; BLEDSOE, BRIAN P.; JACOBSON, ROBERT B.; POFF, N. LEROY; RATHBURN, SARA L.; WALTERS, DAVID M.; WILCOX, ANDREW C. The Natural Sediment Regime in Rivers: Broadening the Foundation for Ecosystem Management. ***BioScience***, *v. 65 (4)*, pp. 358–371, 2015. DOI: 10.1093/biosci/biv002.

YANG, H.; LI, S.; CHEN, J.; ZHANG, X.; XU, S. The Standardization and Harmonization of Land Cover Classification Systems towards Harmonized Datasets: A Review. **Intern. Journal of Geo-Info.**, v. 6(5), p. 154-1-16, 2017. DOI: 10.3390/ijgi6050154.

ZHANG S.;. AND CHAU K. W. Dimension reduction using semi-supervised locally linear embedding for plant leaf classification. In: Huang DS., Jo KH., Lee HH., Kang HJ., Bevilacqua V. (eds) **Emerging Intelligent Computing Technology and Applications.** ICIC 2009. Lecture Notes in Computer Science, vol 5754. Springer, Berlin, Heidelberg.