

Evaluation of dimensional reduction methods on urban vegetation classification performance using hyperspectral data

C. Brabant, Emilien Alvarez-Vanhard, Gwenaël Morin, Thanh Ngoc Nguyen, Achour Laribi, Thomas Houet

▶ **To cite this version:**

C. Brabant, Emilien Alvarez-Vanhard, Gwenaël Morin, Thanh Ngoc Nguyen, Achour Laribi, et al.. Evaluation of dimensional reduction methods on urban vegetation classification performance using hyperspectral data. IGARSS 2018, Jul 2018, Valencia, Spain. halshs-02191363

HAL Id: halshs-02191363
https://shs.hal.science/halshs-02191363

Submitted on 23 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation of dimensional reduction methods on urban vegetation classification performance using hyperspectral data

Brabant C., Alvarez-Vanhard E., Morin G., Thanh Nguyen K., Laribi A., Houet T.

Université de Rennes II, UFR Science sociale, Place Recteur Henri le Moal, CS 24307 - 35043 Rennes cedex, France

Unité Mixte de Recherche 6554 LETG, INEE-CNRS, Université Rennes II, Place du Recteur Henri le Moal, 35043 Rennes cedex, France

I. Introduction

Urban vegetation : a critical issue

In a context of urbanization of urban spaces, urban vegetation is a major issue particularly to the local and global climatic concerns of the territories. Many services are rendered by urban vegetation, such as the improvement of air quality, the cooling power of the heat island or even habitats for biodiversity (Buyadi, 2015). Differentiating various vegetations types or species is therefore relevant to improve the mitigation of urban heat islands (adapting water supplies) and manage sanitary issues (release of allergic pollens). Mapping and monitoring urban vegetation is a critical issue for decision makers and many studies highlight remote sensing as a powerful approach.

Exploring the potential of hyperspectral imagery

Vegetation monitoring, especially in cities, requires high spectral and spatial resolutions to discriminate vegetation types. Compared to multispectral images, hyperspectral data offers the possibility of studying complex and dynamic environments such as cities and discriminating various natural or anthropogenic elements *(*Adeline, 2014*)*. However, hyperspectral data can exhibit spectral redundancies that may affect the performance of classification algorithms. Moreover, high spectral and spatial dimensionality may be highly time consuming. Many studies claim to reduce the spectral dimension of the dataset in order to optimize the efficiency of classification methods in discriminating urban vegetation type for instance (Zhou et al, 2015; Erudel et al, 2017).

The objective of this study is to evaluate the influence of three different methods of dimension reduction on the performance of urban vegetation classifications.

II. Data and Methodology

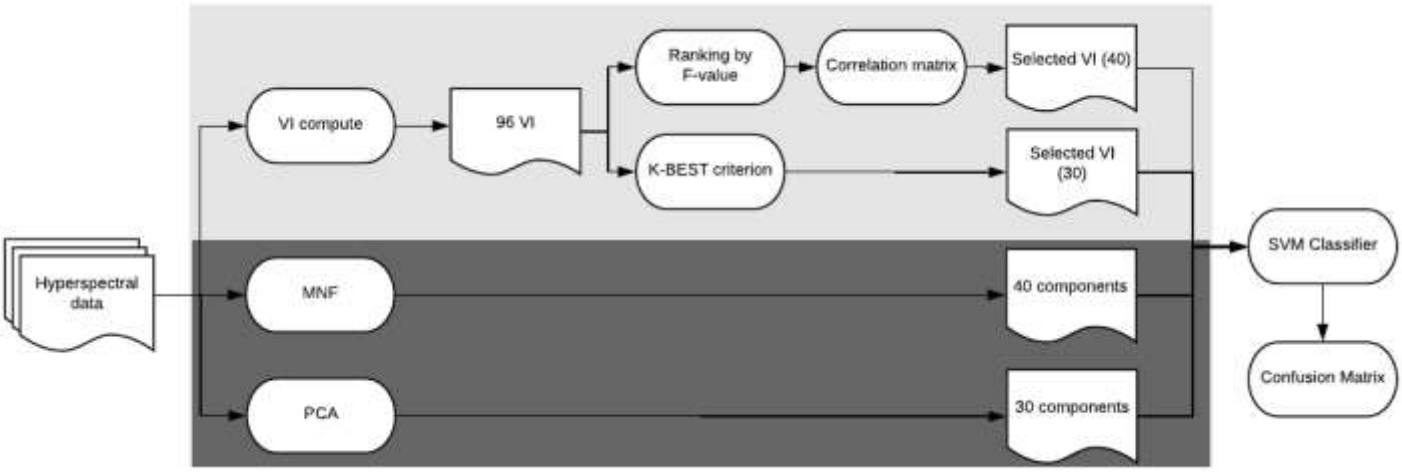


Figure 1 : General methodology of dimension reduction comparison. VI: vegetation indices. In light grey : feature selection approach. In dark grey : feature extraction approach.

Study area

The studied site is the city center of Toulouse (43°36'16''N, 1°26'39''E) - France. The urban area is strongly artificialized and the vegetation is strongly interlocked between buildings and houses, present along the roads or in the yards of building.

Data

Airborne hyperspectral data were acquired using the Hyspex sensor (2m spatial resolution, > 480 spectral bands ranging from 0.4 to 2.5µm) in July 2015. In order to assess the potential of the future Hypxim spaceborne sensor (192 bands), hyspex data have been resampled accordingly to Hypxim specificities at 4m and 8m spatial resolution. Atmospheric corrections were applied using the Cochise method (Poutier et al, 2002).

A field observation made it possible to identify and locate, using a differential GPS (with an absolute geometric accuracy below 10 cm), 510 trees in the swath of the hyperspectral dataset. 19 different types of tree families have been identified. These points of trees are used as learning and validation datasets with no redundancy between both, but training data (class representativeness) exhibit strong heterogeneous sizes (some tree families are well represented while others exhibits only few points).

Methodological approach

The purpose of the hyperspectral images dimensional reduction is to select or extract an optimal subset from spectral bands that are relevant for the issue at stake. Reducing dataset dimensionality helps to reducing the redundancy of the variables and facilitate further processing and interpretation of the data (Khoder, 2013).

Two main types of methods can be distinguished: those based on features extraction and those relying on features selection. Features extraction methods (PCA, MNF for instance) are easy to implement but difficult to interpret. The extracted characteristics doesn't relate to any biophysical or reflexive properties of the land surface but only to a synthesis of information. These algorithms are known to overcome the Hughes phenomenon (Tarabalka, 2010). Features selection method regroups informations that have a biophysical meaning and the lowest redundancy with all others selected informations. Thus, it make it possible to identify the most discriminant characteristics distinguishing the different families of trees. We hypothesize that a selection of that non-redundant indices related to vegetation properties can be appropriate to reduce hyperspectral images. The developed feature selection method is summed up in figure 1. 96 vegetation indices (VI) have been computed, thus integrating various potential discriminant characteristics (Chlorophyll, water stress, stress, soil litter, cellulose, carotenoid ect…) (Erudel et al, 2017). Two reduced VI datasets have been computed using two selection methods:

1) The K-Best method. This algorithm uses ANOVA (ANalysis Of VAriance) as a ranking criterion to identify the best separability between classes. The ANOVA compares intra-class and inter-class dispersion by decomposition of variance.
2) A second method is tested accounting for the correlation between variables. It ranks variables using the ANOVA F-value. Variables are then eliminated according to their degree of correlation with other variables and their ranking position. If two variables overpass the threshold of correlation ($r > 0.85$), the one with the lower F-value is eliminated.

This study seeks to compare two features extraction and two feature selection methods to reduce the high dimensionality of Hypxim imageries on the performance of vegetation classification. A Support Vector Machine (SVM) classifier (Melgani et Bruzzone, 2002) is performed on the 4 reduced datasets and on the non-reduced Hypxim imageries at both 4m and 8m resolution images. SVM finds a hyperplane that best separates the data by maximizing the distance between classes. SVM is prefered here as it is less sensitive to small and non homogeneous training datasets (compared to Random Forest algorithm for instance) (Melgani and Bruzzone, 2002).

The accuracy assessment of the SVM classification is based on the overall accuracy (OA = number of well-ordered pixels/ total number of pixels) and the Kappa Index of Agreement (KIA) as it compares the correct proportion observed with the correct proportion linked to luck (KIA = 1 : classification not linked to luck, KIA = 0 : Classification linked to luck).

III. Results

Vegetation indices selection

The K-Best method contributes to select 30 vegetation indices of which 23 are related to chlorophyll contents. The method using the correlation matrix (Figure 2) allow selecting 40 vegetation indices among which almost 50% are Chlorophyll related. Although, the latter selection method introduces others biophysical indicators, the global separability of classes is similar of K-Best selection. Furthermore, the 13 indices obtaining the ANOVA highest F-score in both methods represent Chlorophyll (38.25 < score-F > 14.46). May be inferred that Chlorophyll is the principal factor for vegetation families' discrimination in this study area.

Influence of reduction dimension methods on vegetation classifications

SVM classifier applied on the 4m resolution Hypwim dataset reduced using the MNF gives an overall accuracy score of 51.92% (KIA : 0.41) (Table 1). At 8m, the OA is 46.92% (KIA : 0.35). Features selection methods provide lower OA of 45.77% (KIA : 0.33) and 45% (KIA : 0.32) for the confusion matrix and the K-Best methods respectively. Results from the PCA are the worst ones with OA below 30% (Table 1) and similar to those results obtained with non-reduced hyperspectral images.

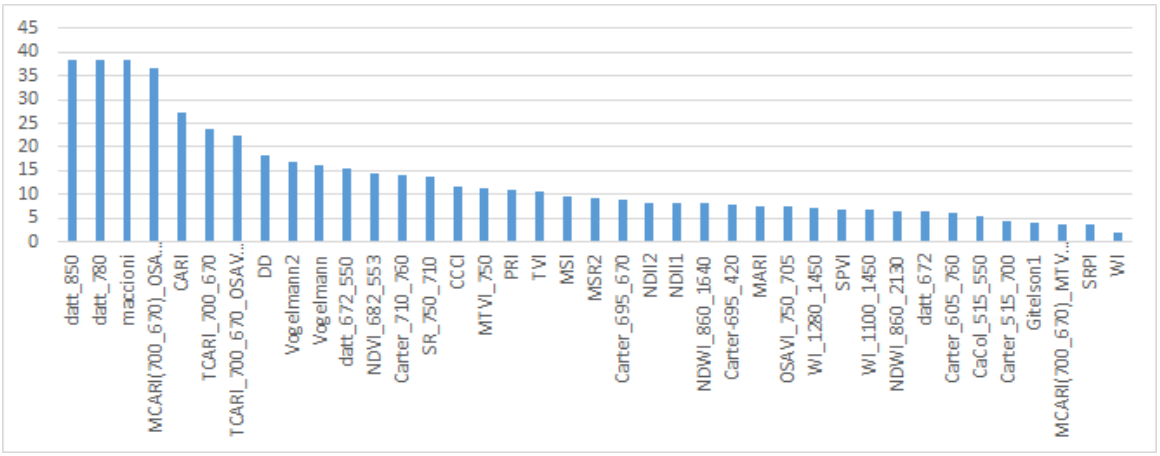


Figure 2: F-Score results of selected vegetation indices by correlation matrix feature selection method

Table 1: Performances of SVM on vegetation classifications using different dimension reduction methods applied on Hypxim hyperspectral imageries (at both 4m and 8m)

Datasets	Spatial resolution	Overall Accuracy (OA)	KIA
Hypxim (192 spectral bands)	4m	31.62 %	0.19
	8m	28.21 %	0.16
Reduced with PCA	4m	28.08 %	0.01
	8m	27.70 %	0.04
Reduced with MNF	4m	51.92%	0.41
	8m	46.92 %	0.35
Reduced with VI (correlation matrix)	4m	45.77 %	0.33
Reduced with VI (K-best)	4m	45 %	0.32

IV. Discussion & Conclusion

This study aimed at evaluating the influence of reduction dimension methods on urban vegetation (tree) classifications. As a first challenge, 19 classes of tree families have been considered. Preliminary results shown that the SVM classifier perform better when using the MNF method followed by defined features selection methods. This paper demonstrates the ability of dimension reduction method to improve detailed urban vegetation classification using the hyperspectral imagery. Even if resolution improvement increases OA and KIA scores, confusions between vegetation types remain. Over and under estimations of some classes can be explained by the internal spectral variability of a single tree family. Here, we hypothesize that trees can have exhibit different spectral characteristics (due to the species considered or their sun exposure, the water availability, etc.) while belonging to the same family (e.g. Magnoliaceae family : Tulips tree from Virginia and Magnolia may have different spectral signatures). The finer the spatial resolution and the timing of observation, more these variabilities tend to dissipate (Xie et al. 2008).

Omission and commission errors can also inherit from the unbalanced training samples that influence the classifier performance. These preliminary results will be completed with some forthcoming works. Various different classifiers (SVM, Random Forest, and Deep Learning) will be assessed and compared. Results from Hypxim datasets will also be compared with those computed from Hyspec 2m data (492 spectral bands) and simulated Sentinel-2 imageries. Furthermore, vegetation types will be simplified and detailed into 5 and 38 classes respectively.

The overall goal of this contribution would be to determine define and criticize the most performant methodological design (classifier, reduction dimension method) according to the training dataset available, on the one hand, and the selected vegetation nomenclature, on the other hand. Potentialities of future spaceborne Hypxim imageries (to be launch in 2021) would thus be evaluated for urban vegetation classification.

References

K. Adeline, “Classification des matériaux urbains en présence de végétation éparse par télédétection hyperspectrale à haute résolution spatiale”. Thèse de doctorat. Toulouse, ISAE. 2014

S. N. A. Buyadi, W. M. N. W. Mohd, et A. Misni., “Vegetation's Role on Modifying Microclimate of Urban Resident”. *Procedia-Social and Behavioral Sciences*, vol. 202, p. 400-407, 2015

J. Degerickx *et al*. “Mapping Functional Urban Green Types Using Hyperspectral Remote Sensing.” *2017 Joint Urban Remote Sensing Event (JURSE)*, 2017

T. Erudel *et al*. “Criteria Comparison for Classifying Peatland Vegetation Types Using In Situ Hyperspectral Measurements.” *Remote Sensing*, vol. 9, no. 8, p. 748, 2017

J. Khoder. “Nouvel algorithme pour la réduction de la dimensionnalité en imagerie hyperspectrale”. Thèse de doctorat. Université de Versailles-Saint Quentin en Yvelines. 2013

F. Melgani et L. Bruzzone. Support vector machines for classification of hyperspectral remote-sensing images. In : *Geoscience and Remote Sensing Symposium. IGARSS'02. 2002 IEEE International*. IEEE, 2002. p. 506-508. 2002

L. Poutier, C. Miech, X. Lenot, *et al.* “COMANCHE and COCHISE: two reciprocal atmospheric codes for hyperspectral remote sensing”. In : *2002 AVIRIS Earth Science and Applications Workshop Proceedings*. 2002.

Y. Tarabalka, J.A. Benediktsson, J. Chanussot *et al.* “Multiple spectral–spatial classification approach for hyperspectral data”. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no 11, p. 4122-4132. 2010

Y. Xie, Z. Sha, & M. Yu. Remote sensing imagery in vegetation mapping: a review. *Journal of plant ecology*, *1*(1), 9-23. 2008

Y. Zhou *et al*. “Dimension Reduction Using Spatial and Spectral Regularized Local Discriminant Embedding for Hyperspectral Image Classification.” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, p. 1082–1095. 2015