

Heterogeneous forest classification by creating mixed vegetation classes using EO-1 Hyperion

Journal:	<i>International Journal of Remote Sensing</i>
Manuscript ID	TRES-PAP-2016-0483.R2
Manuscript Type:	IJRS Research Paper
Date Submitted by the Author:	30-Dec-2016
Complete List of Authors:	Deák, Márton; Eötvös Loránd University, Department of Physical Geography Telbisz, Tamás; Eötvös Loránd University, Department of Physical Geography Árvai, Mátyás; Hungarian Academy of Sciences, Department of Soil Mapping and Environmental Informatics Mari, László; Eötvös Loránd University, Department of Physical Geography Horváth, Ferenc; Hungarian Academy of Sciences, Centre for Ecological Research Kohán, Balázs; Eötvös Loránd University, Department of Environmental and Landscape Geography Szabó, Orsolya; Partium Christian University, Department of Economics Kovács, József; Eötvös Loránd University, Department of Physical and Applied Geology
Keywords:	hyperspectral data, classification, forestry
Keywords (user defined):	band selection

SCHOLARONE™
 Manuscripts

Reply to reviewer 1**Question 1:**

The sentence in page 11, lines 14-21 (Studies about ... full spectra) refers to the number of spectral bands that were used for classification. Spectral indices are based on 2-3 bands in general whereas full spectra is more than 100 bands in general.

Answer 1:

Corrected like the referee suggested – please see in Answer 3.

Question 2:

The next sentence (Classification is ... be very effective) in page 11, lines 21-27 refers to the usage of other features in addition to spectral information. Dalponte et al. (Dalponte et al. 2008) utilize LiDAR data whereas Tarabalka et al. (Tarabalka et al. 2009) utilize segmentation map based on spatial distribution. On the other hand, neural network approaches (Gong et al. 1997, Wang et al. 2008, Caiyun 2014) use only spectral information. As I have repeatedly said in the previous comments, a data processing technique (neural network) cannot be compared with types of data (LiDAR data and segmentation map). In that sense, I don't think referring to neural network in this sentence is appropriate. Considering that the neural networks utilize whole spectral bands, I suppose one option is to move the references (Gong et al. 1997, Wang et al. 2008, Caiyun 2014) to the references with respect to full spectra (Williams et al. 2003, Clark et al. 2005, Boitt et al. 2014, Atzberger et al. 2015, Burai et al. 2015) in page 11, lines 14-21.

Answer 2:

Also corrected like the referee suggested – please see in Answer 3.

Question 3:

2. high resolution (Page 11, line 10, 14) and lower resolution (Page 11, line 12)

Please specify the types of the resolutions, i.e., spectral or spatial resolution.

Answer 3:

Also corrected. Most of the paragraph looks like this now:

1
2
3 “Using this technology species are mostly identified using high spatial resolution
4 hyperspectral data but we can find less papers in the topic of the application of lower spatial
5 resolution hyperspectral satellite images. Studies about the application of high spatial
6 resolution data are often concentrating on the biochemical properties of the plants which are
7 connected to spectral indices or extracted from different parts of the full spectra. Studies using
8 high spectral resolution field and laboratory spectrometry are even more chemistry-dependent.
9 These studies are using different techniques and methods utilizing field/laboratory
10 spectrometer, aerial hyperspectral imagery or both (Gong et al. 1997, Williams et al. 2003,
11 Clark et al. 2005, Boitt et al. 2014, Caiyun 2014, Atzberger et al. 2015, Burai et al. 2015,
12 Wang et al. 2015).”

13
14
15
16
17
18
19
20
21 **Question 4:**

22 3. classification accuracy (Page 19, lines 50-53)

23 At this point, the accuracies have not been shown. Please specify where (i.e., section 5) the
24 evidence (85.78 %) is shown.
25
26
27

28
29 **Answer 4:**

30 Corrected like this:

31
32 “As stated in section 5 in the previous step the highest accuracy was achieved by using our
33 influential band reduction method (84.28%), the rule images created during that classification
34 process were further utilized.”
35
36
37
38

39
40 **Question 5:**

41 4. other combinations (Page 20, lines 32-43)

42 The authors describe that there is a priori information that the possible classes are limited to
43 oak, maple, pine, beech, ground vegetation, oak-pine, oak-maple, oak-beech and maple-
44 beech. But the a priori constraints are not used in the classification process (Page 22, lines 19-
45 33). Without the constraints, there is a possibility that some pixels are classified into other
46 combinations, e.g., maple-beech, pine-beech. However Table 3b excludes the impossible
47 combinations. Were the impossible combinations (e.g., maple-beech, pine-beech) avoided
48 without any constraint?
49
50
51
52
53
54

55
56 **Answer 5:**
57
58
59
60

1
2
3 In the sample area there were some species which were not mixing due to natural
4 biogeographical reasons – these are maple-pine and beech-pine mixtures (maple-beech was
5 present as stated in the paper). Although we might not have been clear only by listing the
6 possible classes. In this case we used some constraints since there still is a possibility that one
7 pixel contains eg. 50% pine and 50% maple in a non-mixing way (eg. there is a sharp
8 boundary between a homogenous maple and a homogenous pine block which is running in the
9 middle of a pixel), but the probability of that is very low.

10
11 Considering this we think that the reviewer is right at this point. We think by studying a larger
12 sample area there is a small possibility that such “false-positive” mixed pixels will be present
13 in small quantities. To avoid detecting them and to clarify we added some more sentences in
14 section 4.3:
15

16
17 “Pine-maple and pine-beech mixtures were not present, so we excluded them from the further
18 steps.”
19

20
21 and the end of the section:
22

23
24 “However pine-maple and pine-beech mixtures in the study area are not present there is still a
25 possibility of classifying some pixels as such (eg. at the boundaries of two touching
26 homogeneous pine and maple blocks). To avoid this, we introduced the constraint: if
27 $P_{pine} \geq 0.25$, then *maple* and *beech classes* were not investigated.”
28
29

30 31 **Question 6a:**

32
33 5. Fig. 3
34

35 From the view point of band selection, the comparison should be based on the identical
36 number of bands. For example, the classification accuracy based on 51 (proposed method)
37 should be compared with the classification accuracy based on 51 (SDA). When the number of
38 bands is 47, the classification accuracy based on SDA outperforms the classification based on
39 the proposed method. However, when the number of bands is increased to 51, the
40 classification accuracy based on the proposed method slightly outperforms the classification
41 based on SDA.
42
43

44 45 **Answer 6a:**

1
2
3 We think that comparing the best results produced by each method shows the real difference
4 between them, but still any other comparisons can be seen in Fig. 3. But for further
5 clarification we added the following sentences to the “6. Discussion” section:
6
7

8
9
10 “We also compared the two optimal number of bands (47 for SDA and 51 for influential band
11 selection). By 47 bands SDA outperformed our influential band selection method (82.85% for
12 SDA and 62.48% for our method) but at 51 bands and above our influential band selection
13 method yielded somewhat better results (83.61% for SDA and 84.28% for our method). This
14 difference is small, but remained persistent in case of even more bands.”
15
16
17

18
19
20 **Question 6b:**

21 Fig. 3 also indicates that, when the numbers of bands exceed some value (50 (proposed) and
22 46 (SDA)), the classification performances are saturated. If we assume the difference in
23 accuracies between 50 (proposed) and 46 (SDA) is not significant, we could say that SDA
24 could achieve good accuracy with fewer number of bands. Please give more discussion that
25 supports the advantage of the proposed method.
26
27
28

29
30
31 **Answer 6b:**

32 We don't really think that the small (1-2%) gap between the results mean that this difference
33 is not significant. If it was random we think that sometimes SDA would produce slightly
34 better results than our method, but in this case both are varying around the accuracy value
35 produced using the optimal number of bands (remaining SDA less accurate in every case than
36 our influential band selection). In conclusion in this case our method managed to find a
37 somewhat better combination, in conclusion SDA ruled out some key wavelengths – probably
38 in the NIR region (eg. 711.72 nm and 884.7 nm) which could have been important for an even
39 more accurate classification. Since the results are slightly but persistently better, we think that
40 this difference in the band composition is a big enough difference: 5 bands are not used by
41 SDA but used by our method and 1 is used by our method and not by SDA.
42
43
44
45
46
47

48 We would like to point out that this study is only about mixed vegetation classes. In case of
49 other mixtures (eg. urban or mineral) the results may be better or worse.
50

51 To clarify this we also added the following sentences to the “6. Discussion” section:
52
53

54
55
56 “As seen in Table 1. bands corresponding to 711.72 nm, 884.7 nm, 1013.3 nm, 1114.19 nm,
57 1205.07 nm were selected by our influential band selection method and were not selected
58
59
60

1
2
3 using SDA while the band corresponding to 742.25 nm was ruled out by influential band
4 selection but it was selected using SDA. In conclusion in this case we managed to find an
5 even better combination of bands for mixed vegetation classes by using more bands in the
6 NIR region and in the starch, protein, oil and H₂O absorption bands (Kumar et al. 2001).
7
8
9

10
11 **Other corrections:**

12 We also corrected some previously unseen typos and added one more band in the 47 bands
13 list (it was accidentally left out after the first revision).
14

15 Also one of our author changed his affiliation in the meantime (Mátyás Árvai).
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7 **Heterogeneous forest classification by creating mixed vegetation**
8 **classes using EO-1 Hyperion**
9

10
11 Márton Deák

12
13 *Department of Physical Geography, Eötvös Loránd University, 1117 Pázmány Péter*
14 *promenade 1/c., Budapest, Hungary, dmarton@elte.hu; corresponding author*
15
16

17
18 Tamás Telbisz

19
20 *Department of Physical Geography, Eötvös Loránd University, Budapest, Hungary,*
21 *telbiszom@ludens.elte.hu*
22
23

24 Mátyás Árvai

25
26 *Institute for Soil Sciences and Agricultural Chemistry, Department of Soil Mapping and*
27 *Environmental Informatics, Hungarian Academy of Sciences, Budapest, Hungary.*
28 *Department of Physical Geography, Eötvös Loránd University, Budapest, Hungary,*
29 *arvai.matyas@agrar.mta.hu, matyas.arvai@gmail.com*
30
31
32

33 László Mari

34
35 *Department of Physical Geography, Eötvös Loránd University, Budapest, Hungary,*
36 *mari.laci@gmail.com*
37
38

39 Ferenc Horváth

40
41 *Centre for Ecological Research, Hungarian Academy of Sciences, Tihany, Hungary,*
42 *horvath.ferenc@okologia.mta.hu*
43
44
45

46 Balázs Kohán

47
48 *Department of Environmental and Landscape Geography, Eötvös Loránd University,*
49 *Budapest, Hungary, kohan.balazs@gmail.com*
50
51

52 Orsolya Szabó
53
54
55
56
57
58
59
60

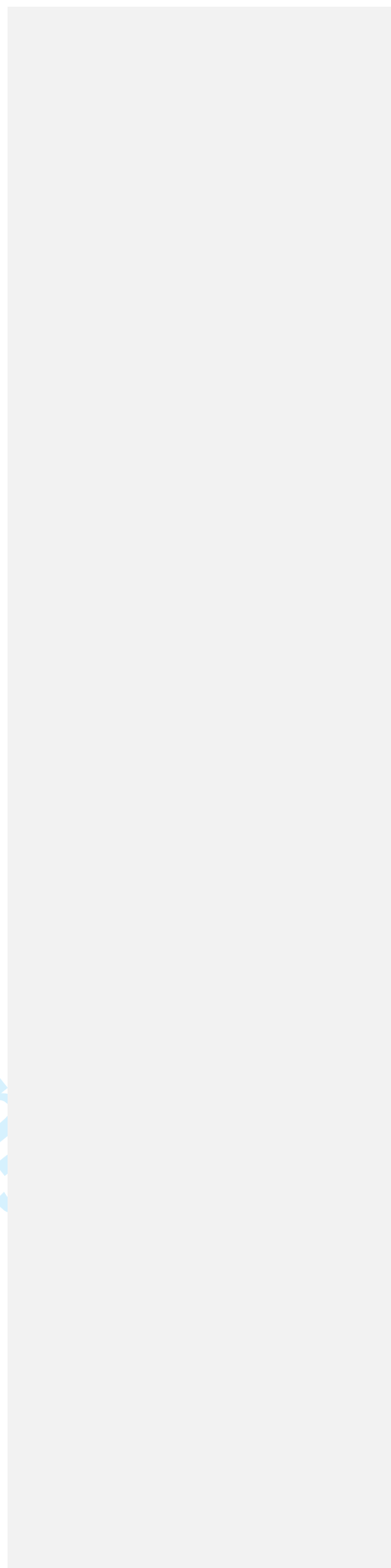
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Department of Economics, Partium Christian University, Oradea, Romania,
orsolya_szabo@yahoo.es

József Kovács

Department of Physical and Applied Geology, Budapest, Hungary,
kevesolt@geology.elte.hu

For Peer Review Only



Heterogeneous forest classification by creating mixed vegetation classes using EO-1 Hyperion

Hyperspectral satellite data is an efficient tool in vegetation mapping, however previous studies indicate that classifying heterogeneous forests might be difficult. In this study, we propose a mapping method for a heterogeneous forest using the data of NASA's EO-1 Hyperion supplemented by field survey. We introduced a band reduction method to raise classification accuracy of the Support Vector Machine classification algorithm and compared the results to the PCA-reduced, SDA-reduced and the original dataset. We also used a modified version of the V-I-S model to create mixed vegetation classes consisting of the commonly mixing species in the area and classified them using Decision Tree Classification method. We managed to achieve 84.2 % approximately using our band reduction method which is 2.3% increase compared to PCA (81.9 %), 1.4% compared to the SDA (82.8%) and 5.5% compared to the original dataset (76.7%). Introducing the mixed vegetation classes raised the overall accuracy even higher (85.8%).

Keywords: hyperspectral data; classification; forestry; band selection

1. Introduction

In case of vegetation classification it is very important to have high spatial resolution and high accuracy while including as many classes as possible. Fire hazard modeling (Thonicke et al. 2010), ecological (Hufkens et al. 2012) or landscape research (Varga et al. 2015), silvi- and agricultural surveys (Martin-Queller et al. 2012, Nelson et al. 2009) and even agricultural forecasts (Ines et al. 2013) all demand the precise knowledge of the vegetation composition in an area. The traditional ways of forest mapping can be time-consuming and often costly and relying only on field survey, data collection may be difficult in some areas (eg. private properties). Official databases are often outdated and might not contain high resolution data.

1
2
3
4
5
6 Remote sensing might be an ideal choice for forest estimation, since it is fast,
7
8 cost-efficient and data can be collected about large areas at the same time. The
9
10 application of hyperspectral remote sensing data is a good option. Using this technology
11
12 species are mostly identified using high spatial resolution hyperspectral data but we can
13
14 find less papers in the topic of the application of lower spatial resolution hyperspectral
15
16 satellite images. Studies about the application of high spatial resolution data are often
17
18 concentrating on the biochemical properties of the plants which are depending
19
20 connected to or spectral indices or extracted from different parts of the or the full
21
22 spectra. Studies using high spectral resolution field and laboratory spectrometry are
23
24 even more chemistry-dependent. These studies are using different techniques and
25
26 methods utilizing field/laboratory spectrometer, aerial hyperspectral imagery or both
27
28 (Gong et al. 1997, Williams et al. 2003, Clark et al. 2005, Boitt et al. 2014, Caiyun
29
30 2014, Atzberger et al. 2015, Burai et al. 2015, Gong et al. 1997, Wang et al. 2015).
31
32 Classification is also possible through multi-data (Dalponte et al. 2008, ~~Caiyun 2014~~)
33
34 approaches, ~~but neural network (Gong et al. 1997, Wang et al. 2015)~~ and
35
36 ~~spatial~~spectral-spectral-spatial (Tarabalka et al. 2009) methods are also tend to be very
37
38 effective. On the other hand, hyperspectral aerial images are costly and cover only a
39
40 small area, while spaceborne data is much more cost-efficient and can provide
41
42 information about a large area at the same time.

43
44 NASA's Earth Observing-1 was one of the first satellites equipped with a
45
46 hyperspectral sensor, the Hyperion. It was originally designed as an experiment for
47
48 testing spaceborne hyperspectral sensors but is still active and new methods come up in
49
50 various topics. Its applications range from volcanology (Abrams et al. 2013) to
51
52 vegetation analysis and classification (Pu et al. 2005, Mitri and Gitas 2010) and forest
53
54 monitoring (Goodenough et al. 2002). The application of hyperspectral data usually
55
56
57
58
59
60

1
2
3
4
5
6 produced better results in forest classification than multispectral data (Goodenough et
7 al. 2003, Pu et al. 2012). In earlier studies, confusion was high regarding the deciduous
8 species (George et al. 2014, Dalponte et al. 2012) and most studies aimed at classifying
9 forests consisting of (spectrally) well separable, homogeneous patches dominated by a
10 single species in the upper canopy sector (eg. Thenkabail et al. 2003, Galidaki and Gitas
11 2015).

12
13
14
15
16
17
18 The Hyperion's spatial resolution is 30 meters/pixel making spectral mixture
19 one of the most serious problems. This challenge can tackled by using only a subset of
20 bands which can be used effectively for spectral mixture and sub-pixel analysis. Further
21 classification accuracy also tends to be higher, since the spectral separability of the
22 surface cover types becomes also usually higher. Principal component analysis (PCA),
23 minimum noise fraction (MNF) and stepwise discriminant analysis (SDA) turned out to
24 be the most effective methods and they are the most widely used (eg. Qian and Fowler
25 2007, Licciardi et al. 2011, Xu et al. 2012, Licciardi and Chanussot 2015). Their
26 application however might not be ideal if the endmembers have high correlation or the
27 goal of the study is to separate a group of spectrally similar endmembers from each
28 other in an otherwise spectrally heterogeneous area (Ifarragueri and Chang 2002, Farrell
29 and Mersereau 2005, Prasad and Bruce 2005).

30
31
32
33
34
35
36
37
38
39
40
41 In our study area, the vegetation is extremely mixed and spectrally similar.
42
43 Consequently, the band reduction and classification processes needed a slightly
44 different approach. Our goal was to create a band reduction method which has proved to
45 be successful using the Support Vector Machine (SVM) algorithm. We chose SVM
46 because earlier studies demonstrated that it produces better results in forest
47 classification than Spectral Angle Mapper, Maximum Likelihood or other algorithms
48 (Huang et al. 2002, Pal and Mather 2005, Vyas D. et al. 2011, Shao and Lunetta 2012)
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 and it is. As a verification, we compared our band reduction results to the ones produced
7
8 by the PCA method and by using the original full-dimensional data. We also designated
9
10 mixed vegetation classes by using an altered version of the V-I-S (Vegetation-
11
12 Impervious-Soil) model presented by Ridd (1995). We substituted the urban
13
14 endmembers with the commonly appearing tree species and created a vegetation map
15
16 showing mixed classes using sub-pixel information. This method is similar to some
17
18 spectral unmixing methods which can be linear or non-linear (Keshava and Mustard
19
20 2002, Keshava 2003), but instead of determining the ratio of every possible endmember
21
22 in one pixel we only tried to determine if given class is present in given pixel or not.
23
24

25 26 **2. Study area**

27
28 Our study area is located in the Budai Hills, which are topographically part of relatively
29
30 low mountains (the highest peak is 559 m) next to Hungary's capital city, Budapest
31
32 (*Figure 1*). The study field is relatively flat and is bordered by the Csillebérc,
33
34 Széchenyi-hegy and Farkas-árok morphological units. The vegetation is diverse and
35
36 mainly arboreal.
37
38

39
40 *Figure 1 here*
41
42
43
44

45 This site is a member of the Pannonicum Floristic province, within the Pannonicum
46
47 Floristic province which belongs to the Bakonyicum Floristic district. The macroclimate
48
49 is continental considerably affected by submediterranean climate, thus characteristically
50
51 Eurasian, continental and submediterranean elements determine the flora in this region.
52
53 Ideally, the region's typical plant community is the Pannonian - Balkanic turkey oak-
54
55 sessile oak forests, *but the study area is a protected territory, thus faster growing tree*
56
57
58
59
60

1
2
3
4
5
6 species are the dominant ones. In the past decades, the forestry responsible for
7
8 silviculture only rarely interfered in the natural succession of this woody area, therefore
9
10 newly planted patches are only sparsely found in the area. It follows that the prime
11
12 trees (sessile oak - *Quercus petraea*, common hornbeam - *Carpinus betulus*) are unable
13
14 to do substance. Nowadays, the plant community of the area represents a mixture of
15
16 *Tilia* (lime-tree), *Quercus* (oak) and *Sorbus* (sorb) species. The study area also includes
17
18 homogeneous plantings of *Pinus nigra* (European black pine) substance. *Fagus*
19
20 *sylvatica* (Beech) appears in high numbers in some gorges or on slopes with northerly
21
22 aspects, because the secondary tree species – such as lime-trees and sorbs – can't grow
23
24 fast enough where the number of sunny hours is too low.

25
26 In our study area the different species are rarely separated in homogeneous
27
28 groups. Instead, oak species are commonly mixed with black pine, maple and rarely
29
30 with beech, but we also found maple-beech mixed patches.

31 **3. Data**

32 **3.1 Pre-processing**

33
34
35 The Hyperion instrument is capable of resolving 220 spectral bands between 400 to
36
37 2500 nm. The images are covering a 7.5 km wide and 100 km long area on the surface.
38
39 Since it was originally designed for aerial applications the signal-to-noise ratio is
40
41 relatively low and there are many errors in the data (Datt et al. 2003, Khurshid et al.
42
43 2006). Most of these can be corrected using pre-processing methods or avoided by
44
45 using a selection of bands.
46
47

48
49 Data pre-processing steps were made according to Apan et al. (2004). We
50
51 corrected striping, marked some spectral ranges as “bad bands” in the atmospheric
52
53 regions and also ruled out the bands which had a very low signal-to-noise ratio. In the
54
55
56
57
58
59
60

1
2
3
4
5
6 end, we used only 141 bands. For the image pre-processing and data processing we used
7
8 EXELIS VIS's ENVI 5.2 software.
9

10 The image was captured on the 14th July 2013 (ID:
11
12 EO1H1880272013195110KA_SG1_01). The data was Level1Gst and was in radiance
13
14 units. We selected a mid-summer image to avoid seasonal, temporary spectral
15
16 differences which are caused by (for example) different vegetation stages so the method
17
18 can focus more on the constant, physiological properties of the trees. Cloud coverage
19
20 was low (less than 1%) but it affected some pixels in the study area. The clouds and
21
22 built-up areas were masked out during the process.
23

24 25 **3.2. Field survey**

26
27 Homogeneous and heterogeneous vegetation blocks were digitized using a Trimble
28
29 Nomad GPS (min. accuracy was 7 m). We selected vegetation blocks partly on species
30
31 level (European black pine; beech) and partly on genus level (oak: sessile oak and
32
33 turkey oak; maple: field maple, Norway maple). This way we classified altogether 588
34
35 pixels in the field as homogeneous groups and used 314 of them as training samples
36
37 (their average spectra was used for classification and band reduction) and 274 as control
38
39 points. We calculated their average spectra, using 38 points for ground vegetation, 114
40
41 for pine, 83 for beech, 24 for maple and 55 for oak. We also collected 203 pixels for
42
43 different mixed scenarios (62 pixels for maple-beech mixed, 5 for oak-beech mixed –
44
45 which was present only in a very few areas –, 71 for oak-pine mixed and 35 for oak-
46
47 maple mixed) from which we used 115 as training (to extract the average rule values)
48
49 and 58 as control.
50
51
52
53
54
55
56
57
58
59
60

4. Methods

4.1. Band reduction

Principal component analysis is a multivariate statistical method to orthogonally transform the data and reduce the number of dimensions (bands) by converting the points to uncorrelated data. It is considered as a standard application in many fields and is still the most popular band reduction method in (hyperspectral) remote sensing. One of its drawbacks is, when applied it does not discriminate between the training spectra and the rest. As a result, when two or more training spectra have high similarity related to the rest of the image, the small differences which make them different from each other will likely disappear (Farrell and Merserau 2005, Prasad and Bruce 2005) – which was also the case in this study.

To avoid this, we propose a band reduction method which separates the spectra by selecting only the bands in case of which the minimum difference between the average training samples in one band is larger than average minimum difference value of all the training samples measured in all bands. This method will be done in respect of the bands where the probability of confusion is the highest, eliminating the bands where one or more training samples are more similar, naturally raising the class separability.

First, we calculated the average (radiance) spectra from the training areas and used them as input. Second, we calculated the differences between all of the above mentioned training spectra in each band, then the minimum value was assigned to the corresponding band (this became the *minimum intensity difference*). Third, we calculated the average of all minimum intensity differences. We called a band influential, if the smallest intensity difference was larger than the average minimum intensity difference calculated using all bands.

The formula seen in Equation (1) describes the method of finding the influential bands ($ib(b)$).

Equation (1)

$$mid(b_i) = \min_{1 \leq j < k \leq m} |I(j, b_i) - I(k, b_i)|$$

$$ib(b_i) = \begin{cases} 1, & \text{if } mid(b_i) > \frac{1}{n} \sum_{l=1}^n mid(b_l) \\ 0, & \text{if } mid(b_i) \leq \frac{1}{n} \sum_{l=1}^n mid(b_l) \end{cases}$$

Where b_i is the band number, $mid(b_i)$ is the “minimum intensity difference” at band i . j and k are the IDs of the classes – therefore $I(j, b_i)$ is the average intensity of class j at band i . $ib(b)$ is a Boolean-type variable. Therefore $j, k \in (1 \dots m)$ and $i \in (1 \dots n)$ where m is the number of classes and n is the number of bands.

Using the method we selected 51 bands which can be seen in *Table 1*. The original and band reduced spectra are in *Figure 2a and 2b*.

Figure 2 here

Table 1 here

The selected bands formed three different groups. The first group is between 752.43 nm and 884.7 nm (including 711.72 nm separately) in the NIR clearly corresponding to the chlorophyll reflectance ranges (Curran et al. 1991, Kumar et al. 2001). The second group is between 1013.3 nm and 1114.19 nm (including 993.17 nm separately), which in case of vegetation may be connected to starch, protein, oil and and H₂O absorption (Kumar et al. 2001). The third group is 1164.68 nm to 1305.96 nm,

1
2
3
4
5
6 which corresponds to lignin, cellulose, starch and H₂O content (Serrano et al. 2002,
7 Kumar et al. 2001). Using this method we selected 6 bands which were not present
8 using the SDA reduction and 1 which was only included there. The 6 bands only
9 selected by our method are: 711.72 nm, 884.7 nm, 1013.3 nm, 1114.19 nm, 1194.97 nm
10 and 1205.07 nm and the only band which wasn't selected by us but was included using
11 SDA was 742.25 nm.
12
13
14
15
16

17
18 Since the training areas for the groups we used for the average calculation (thus
19 for finding the influential bands as well) not only differ in the species composition but
20 also in many forest physiological properties (eg. closure, height, age), the selected
21 influential bands do not only show the difference between the physiochemical
22 difference between the different tree species, but also the physiological difference of the
23 forest segments. Thus the final training spectra are consisting of the mixture of
24 photosynthetic and non-photosynthetic vegetation parts, soil and reflective properties of
25 the ground vegetation (if present) as well, having relatively large intra-class variability.
26 This variability is expected to be lower if the forest segments have similar age, closure
27 and the ratio of the dominant species is closer to 100% (since then one pixel will
28 correspond to one type of surface cover), but because of the spectral differences
29 between photosynthetic and non-photosynthetic parts it's expected to be higher at low
30 closure (Guerschman et al. 2015).
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 **4.2. Classification**

46 Support Vector Machine is a supervised classification algorithm which performs well
47 compared to other methods if the dimensionality is high and the training spectra are
48 highly correlated (Pal and Mather 2006). SVM was performed firstly on the full-
49 dimensional, secondly on the PCA-reduced, thirdly on the SDA reduced, and fourthly
50 on the 51 influential bands data.
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 In order to achieve the highest accuracy with PCA-reduction, we used the first 3
7 principal components (only the first 3 principal components were selected because
8 increasing the number of PCA bands lowered classification accuracy).
9

10
11 We also compared our band selection and classification results to the bands
12 selected by the Stepwise Discriminant Analysis method as used by George et al. (2014)
13 and Vyas et al. (2011) and described by Green and Carroll (1978). Using this method we
14 selected 47 spectral bands which are included in *Table 1* as comparison to our selection
15 method.
16
17
18
19
20

21 The training areas were selected by taking into consideration the vegetation
22 patterns. Both homogeneous and heterogeneous patches were selected. The
23 homogeneous patches were used as training areas for all the classification methods and
24 the heterogeneous ones were utilized later, while introducing mixed vegetation classes.
25
26
27

28 Our first goal was to classify the image for the dominant species using the rule
29 images (which contained probability values of each pixel corresponding to a class) of
30 the SVM, classifying a pixel into the class with the highest probability (this is the
31 default setting in most image analysis software). In order to achieve the highest
32 accuracy, the grid search method described by Hsu et al. (2010) was used. Using this
33 method the user should try different γ and penalty parameter (C) values while using the
34 radial basis function (RBF) kernel and constantly cross-validate the results. According
35 to this in the final classification we used radial basis function kernel, gamma value was
36 0.25 and the penalty parameter was 100. The highest results were achieved by using
37 these settings.
38
39
40
41
42
43
44
45
46
47
48

49 The designated classes corresponded to three deciduous genera (beech, maple,
50 oak) and one coniferous (pine). Considering the typical surface cover types of the area
51 we had to introduce ground vegetation (opening areas without trees) as a new class.
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 There were also a lot of built-up areas near the study area but as stated earlier they were
7
8 masked out prior to analysis.
9

10 **4.3. Assigning mixed vegetation classes**

11
12 Our second goal was the application of the rule images created during the classification
13 process and the use of them to create mixed vegetation classes. As stated in section
14 5 Since in the previous step the highest accuracy was achieved by using our influential
15 band reduction method (51 bands 84.28%), the rule images created during that
16
17 classification process were further utilized.
18
19
20
21

22
23 We used an alternative version of the V-I-S model, modifying the method of Lu
24 and Weng (2004), who applied subpixel classification for their Indianapolis study area
25 by substituting the urban endmembers (vegetation, soil, impervious) with the dominant
26 species in the study area (beech, pine, maple, oak). We used the same idea as the above
27 mentioned scientists. The basis of this method is to classify the pixels by using the rule
28 image values in a way, that if a pixel reaches a given value of classification probability
29 it can be classified as a mixture of two or more classes. It is also possible to classify
30 each pixel to its dominant class (assigning the one where the rule layers have the highest
31 probability value). Since at the mixed vegetation areas the rule values were similar for
32 several classes and the rule value of the dominant class was only slightly higher, the
33 introduction of mixed classes was more realistic. Using the same idea, we classified the
34 image for the dominant species and classified each pixel regarding the certainty of the
35 membership to one or two classes.
36
37
38
39
40
41
42
43
44
45
46
47

48 As stated earlier, the commonly mixing species were oak with pine, maple and
49 beech as well as maple with beech, thus we introduced oak-pine, oak-maple, oak-beech
50 and maple-beech classes beside the original pure classes (oak, maple, pine, beech,
51 ground vegetation). Pine-maple and pine-beech mixtures were not present, so we
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 excluded them from the further steps. We extracted the mean rule values of each
7
8 significant (mixing) class according to the field survey data of the mixed training areas.
9
10 They can be seen in *Table 2*.

11
12
13
14 *Table 2 here*
15
16
17
18

19 Since the mean values of all significant species in the mixed classes were above 0.25 (as
20 it can be seen in *Table 2*), we selected this threshold value to create mixed classes. If a
21 pixel reached the above threshold value only in one class it was classified solely into
22 that class. But if the pixel reached it in two (or more) classes, it was classified as a
23 member of all. However, we note that triple (or higher) membership was very
24 uncommon – in these cases, the pixels were finally classified into the two classes with
25 the highest probability values.
26
27

28 This process was implemented by applying a decision tree classification method where
29 the input data sources were the rule images used for classification. This classification
30 method is widely used to aid other classification methods as well as it is used by its own
31 (eg. Friedl and Brodley 1997, Tooke et al. 2009). A pixel P is classified into class a , if
32 $P_a = \max \{P_i | i \in (1, n)\}$, where P_i is the corresponding rule image value of a given pixel in
33 class i , and n is the number of classes. In the present case the classes were oak, maple,
34 pine, beech and ground vegetation. This way, each pixel is assigned to the dominant
35 class present in the corresponding area. Thereafter, if $P_x \geq 0.25$, then P also becomes a
36 member of class x beside its dominant class. In case a pixel is assigned to more than two
37 classes, only the two highest probability classes are left.
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52
53
54 However pine-maple and pine-beech mixtures in the study area are not present there is
55
56
57
58
59
60

1
2
3
4
5
6 still a possibility of classifying some pixels as such (eg. at the boundaries of two
7 touching homogeneous pine and maple blocks). To avoid this, we introduced the
8 constraint: if $P_{pine} \geq 0.25$, then maple and beech classes were not investigated.
9
10
11

14 5. Results

15
16 Overall accuracy of assigning the dominant species was 76.67% when we used the full-
17 dimensional (141 bands) data. In case of the PCA reduced dataset, the overall accuracy
18 proved to be higher (81.92%) ~~and while using~~ SDA to select a subset of bands (47
19 bands) it was even higher (82.85%) but the highest accuracy was achieved when the
20 influential bands (51 bands) were selected by our method (84.28%). The introduction of
21 mixed vegetation classes raised the accuracy even further to 85.78%. The error matrices
22 in *Table 3* show the overall accuracy, kappa statistic values and confusion for each
23 classification result. We also indicated the user's accuracy values (User's accuracy is a
24 measure indicating the probability that a pixel is Class A given that the classifier has
25 labeled the pixel into Class A – Story and Congalton 1986).
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 *Table 3a and 3b here*

44 Since the accuracy of our method produced only a slight raise in accuracy compared to
45 other methods, we also made a comparison by using a different number of influential
46 and SDA reduced bands to eliminate the probability of random noise affecting the
47 results. While the number of bands is determined in both methods we checked the
48 overall classification accuracy using 46-56 influential bands (sorted by the minimum
49 intensity difference value), and on 42-52 SDA reduced bands (sorted by the individual
50 Wilk's lambda value of each bands) ~~as seen in Figure 3.~~ On Figure 3 it is visible,
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 that in case of both methods a significant raise in accuracy occurred just before the
7
8 optimal number of bands. It's also visible that the accuracy value remained relatively
9
10 constant (with small changes) around value produced by each method's proposed
11
12 combination of bands.

13
14
15
16 *Figure 3 here*
17
18
19

20
21 *Figure 4* shows the false colour composite image of the area with the classification
22
23 results. Altogether, 9 classes were assigned. 5 of them for the dominant species (beech,
24
25 maple, oak, pine and ground vegetation) and 4 of them, in the last image, for the mixed
26
27 classes (beech-maple, oak-beech, oak-maple and oak-pine).

28
29
30 *Figure 4 here*
31
32
33

34
35 In each classification image, oak is dominant on the southwestern slopes, while beech
36
37 and clearings appear commonly on the northeastern slopes, thus the impact of aspect on
38
39 species distribution is clearly recognizable. European black pine is present only in
40
41 planted patches or at the margins of the area, where it is commonly mixed with sessile
42
43 oak.

44
45 When we used only the dominant species for classification, the confusion
46
47 between beech and oak was the highest in the full dimensional and 51 bands data. In the
48
49 classification based on the full dimensional data 41.67% of the beech control points
50
51 were misclassified as oak, while in case of the 51 bands data this proportion was only
52
53 30.56%. By using the PCA, the results were more accurate regarding the beech (94.2%
54
55
56
57
58
59
60

1
2
3
4
5
6 was classified correctly), but 59.26% of the maple was also misclassified as oak,
7
8 making it the least accurate result (25.93%).
9

10 The introduction of mixed classes raised the overall accuracy compared to the
11
12 51 bands data, but the accuracy values of single-species classes dropped (except for
13
14 beech). However, the relatively good accuracy (>80%) of mixed classes raised both the
15
16 kappa and overall accuracy values. The accuracy of maple was fairly low (70.37 %),
17
18 and 22.22% of the maple control points were misclassified as maple-beech.
19

20 21 **6. Discussion**

22
23 We applied the SVM classification method for a small study area with a mixed forest
24
25 based on the same hyperspectral image using three different sets of bands. First, full
26
27 dataset (141 bands) was used as input to SVM, second, the PCA reduced spectra were
28
29 used, and third, we introduced a new band reduction method to select the influential
30
31 bands based on the training spectra. In this latter case we found 51 influential bands and
32
33 concluded that these “influential” bands resulted the highest overall accuracy (84.28%).
34
35 We also introduced mixed vegetation classes based on the V-I-S model and assigned
36
37 pixels with high rule values in more than one class into mixed classes using a decision
38
39 tree classifier. Accuracy was the highest (and confusion the lowest) by using this
40
41 method (85.79%).
42

43 We also compared the two optimal number of bands (47 for SDA and 51 for
44 influential band selection). By 47 bands SDA outperformed our influential band
45 selection method (82.85% for SDA and 62.48% for our method) but at 51 bands and
46 above influential band selection yielded somewhat better results (83.61% for SDA and
47 84.28% for our method). This difference is small, but remained persistent in case of
48 even more bands.
49
50
51
52
53
54
55
56
57
58
59
60

Formatted: New paragraph

1
2
3
4
5
6 As seen in Table 1, bands corresponding to 711.72 nm, 884.7 nm, 1013.3 nm,
7 1114.19 nm, 1205.07 nm were selected by our influential band selection method and
8 were not selected using SDA while the band corresponding to 742.25 nm was ruled out
9 by influential band selection but it was selected using SDA. In conclusion in this case
10 we managed to find an even better combination of bands for mixed vegetation classes
11 by using more bands in the NIR region and in the starch, protein, oil and H₂O
12 absorption bands (Kumar et al. 2001).

Formatted: Subscript

13
14
15
16
17
18
19
20
21 However, the results are good, they still show a certain degree of confusion
22 which can be explained by the high proportion of mixing trees in the field, resulting
23 high rule values for several classes. In case of the PCA reduced dataset, another
24 explanation for confusion might be the phenomenon described earlier: the PCA usually
25 removes the small differences between the spectrally similar classes.

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
As it can be seen on *Figure 34*, many of the pixels classified as oak based on the
full dataset and based on the PCA, were classified as maple based on the 51 bands data.
After the introduction of the mixed classes, most of those pixels were classified as into
the mixed oak-maple class. These points were verified via ground-truth points, and we
found that the PCA and the full dataset based processing classified the “unstable” pixels
(having high probability values in both maple and oak classes) as oak, whereas the 51
influential bands data resulted that these points were classified as maple, but in fact,
they are actually mixed areas.

When mixed classes were used, the accuracy was somewhat lower for the
homogeneous patches. This might be explained by the presence of other mixing species,
which remained unobserved during the field survey. This also might be caused by other
individual trees of non-listed species (eg. lime-tree, sorb) as well.

1
2
3
4
5
6 While our results were relatively good compared to most of the other studies,
7
8 there are even better ones in the literature. For example, George et al. (2014) used SDA
9
10 to reduce the bands for forest classification. The accuracy of their classification was
11
12 82.27% which is somewhat lower than the results produced by our influential band
13
14 selection method but it is about the same as what we've got while using the SDA
15
16 reduced dataset (82.85%) on our own sample area. Goodenough et al. (2002) produced
17
18 92.9% overall accuracy in Canada by using the full dataset and Maximum Likelihood
19
20 classification (however, they also had easily classifiable, non-vegetation classes that
21
22 have raised the overall accuracy value). In addition, we'd like to point out that none of
23
24 the above mentioned studies used mixed classes and the vegetation they studied was
25
26 fairly homogeneous.

27
28 Our relatively good results might also have been affected by different closure
29
30 parameters. In this case, soil and ground vegetation reflectance had different effects on
31
32 the different classes.

33
34 In general, our conclusion was that both the band reduction method and the
35
36 application of the V-I-S model in forest classification are suitable for characterizing and
37
38 classifying mixed vegetation areas.

41 **67. References:**

42
43 | Abrams, M., D. Pieri, V. Realmuto and R. Wright. 2013. "Using EO-1 Hyperion Data
44
45 as HypIRI Preparatory Data Sets for Volcanology Applied to Mt Etna, Italy."
46
47 IEEE Journal of Selected Topics in Applied Earth Observations in Remote
48
49 Sensing 6 (2): 375-385. doi: 10.1109/JSTARS.2012.2224095
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6 Apan, A., A. Held, S. Phinn and J. Markley. 2004. "Detecting sugarcane 'orange rust'
7 disease using EO-1 Hyperion hyperspectral imagery." *International Journal of*
8 *Remote Sensing* 25 (2): 489-498. doi: 10.1080/01431160310001618031
9
10
11
12 Atzberger, C., R. Darvishzadeh, M. Immitzer, M. Schlerf, A. Skidmore and G. le Marie.
13
14 2015. "Comparative analysis of different retrieval methods for mapping grassland
15 leaf area index using airborne imaging spectroscopy". *International Journal of*
16 *Applied Earth Observation and Geoinformation* 43: 19-31. doi:
17 10.1016/j.jag.2015.01.009
18
19
20
21
22 Boitt, M., C. Ndegwa and P. Pellikka. 2014. "Using Hyperspectral Data to Identify
23 Crops in a Cultivated Agricultural Landscape - A Case Study of Taita Hills,
24 Kenya." *Earth Science and Climatic Change* 5 (9): 232-236. doi: 10.4172/2157-
25 7617.1000232
26
27
28
29
30 Burai, P., B. Deák, O. Valkó and Tomor T. 2015. "Classification of Herbaceous
31 Vegetation Using Airborne Hyperspectral Imagery." *Remote Sensing* 7 (2): 2046-
32 2066. doi: 10.3390/rs70202046
33
34
35
36 Caiyun, Z. 2014. "Combining Hyperspectral and Lidar Data for Vegetation Mapping in
37 the Florida Everglades." *Photogrammetric Engineering and Remote Sensing* 8
38 (80): 733-743. doi: 10.14358/PERS.80.8.733
39
40
41
42 Clark, M.L., D.A. Roberts and D.B. Clark. 2005. "Hyperspectral discrimination of
43 tropical rain forest tree species at leaf to crown scales." *Remote Sensing of*
44 *Environment* 96 (3-4): 375-395. doi: 10.1016/j.rse.2005.03.009
45
46
47
48 Curran, P.J., J.L. Dungan, B.A. Macler, S.E. Plummer 1991. "The effect of a red leaf
49 pigment on the relationship between red edge and chlorophyll concentration."
50 *Remote Sensing of Environment* 35 (1), 69-76. doi:10.1016/0034-
51 4257(91)90066-F
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6 Dalponte, M., L. Bruzzone and D. Gianelle. 2008. „Fusion of Hyperspectral and LIDAR
7 Remote Sensing Data for Classification of Complex Forest Areas.” IEEE
8 Geoscience and Remote Sensing 46 (5): 1416-1427. doi:
9 10.1109/TGRS.2008.916480
- 10
11 Dalponte, M., H.O. Orka, T. Gobakken, D. Gianelle and E. Naeset. 2012. “Tree
12 Species Classification in Boreal Forests With Hyperspectral Data.” IEEE
13 Geoscience and Remote Sensing 51 (5): 2632-2645. doi:
14 10.1109/TGRS.2012.2216272
- 15
16 Datt, B., T.R. McVicar, T.G. Van Niel and D.L.B. Jupp. 2003. “Preprocessing EO-1
17 Hyperion Hyperspectral Data to Support the Application of Agricultural Indexes.”
18 IEEE Geoscience and Remote Sensing 41 (6): 1246-1259. doi:
19 10.1109/TGRS.2003.813206
- 20
21 Farrel, M.D. and R.M. Mersereau. 2005. “On the impact of PCA dimension reduction
22 for hyperspectral detection of difficult targets.” IEEE Geoscience and Remote
23 Sensing 2 (2): 192-195. doi: 10.1109/LGRS.2005.846011
- 24
25 Friedl, M.A. and C.E. Brodley. 1997. “Decision tree classification of land cover from
26 remotely sensed data.” Remote Sensing of Environment 61 (3): 399-409. doi:
27 10.1016/S0034-4257(97)00049-7
- 28
29 Galidaki, G. and I. Gitas. 2015. “Mediterranean forest species mapping using
30 classification of Hyperion imagery.” Geocarto International 30 (1): 48-61. doi:
31 10.1080/10106049.2014.883439
- 32
33 George, R., H. Padalia and S.P.S. Kushawaha. 2014. “Forest tree species discrimination
34 in western Himalaya using EO-1 Hyperion.” International Journal of Applied
35 Earth Observation and Geoinformation 28: 140-149.
36 doi:10.1016/j.jag.2013.11.011
- 37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6 Gong, P., R. Pu and B. Yu. 1997. "Conifer species recognition: An exploratory analysis
7 of in situ hyperspectral data." *Remote Sensing of Environment* 62 (2): 189-200.
8 doi: 10.1016/S0034-4257(97)00094-1
9
- 10
11
12 Goodenough, D.G., A.S. Bhogal, A. Dyk, A. Hollinger, Z. Mah, K.O. Niemann, J.
13 Pearlman et al. 2002. "Monitoring forests with Hyperion and ALI", *Geoscience*
14 and *Remote Sensing Symposium*, 2002. IGARSS '02. 2002 IEEE International,
15 2002, 2, 882-885.
16
17
18
19
- 20 Goodenough, D.G., A. Dyk, K.O. Niemann, J.S. Perman, H. Chen and Han T. et al.
21 2003. "Processing Hyperion and ALI for forest classification". *Geoscience and*
22 *Remote Sensing*, 2003, 41, 1321-1331
23
24
- 25
26 Green, P.E., J.D. Caroll. 1978. "Analyzing Multivariate Data". The Dryden Press,
27 Illinois
28
- 29
30 Guerschman J.P., P.F. Scarth, T.R. McVicar, Renzullo L.J., Malthus T.J., Stewart J.B.,
31 Rickards J.E., Trevithick R. 2015. "Assessing the effects of site heterogeneity and
32 soil properties when unmixing photosynthetic vegetation, non-photosynthetic
33 vegetation and bare soil fractions from Landsat and MODIS data." *Remote*
34 *Sensing of Environment* 161: 12-26. doi: 10.1016/j.rse.2015.01.021
35
36
37
38
- 39 Hsu, C.W., C.C. Chang and C.J. Lin. 2010. "A Practical Guide to Support Vector
40 Classification"
41
42 <https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf>
43
44 (accessed October 22 2015.)
45
46
- 47 Huang, C., L.S. Davis and J.R.G. Townshend. 2002. "An assessment of support vector
48 machines for land cover classification." *International Journal of Remote Sensing*
49 23 (4): 725-749. doi: 10.1080/01431160110040323
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6 Hufkens, K, M.A. Friedl, T.F. Keenan, O. Sonnentag, A. Bailey, J. O'Keefe and A.D.
7
8 Richardson. 2012. "Ecological impacts of a widespread frost event following
9
10 early spring leaf-out". *Global Change Biology* 18 (7): 2365-2377. doi:
11
12 10.1111/j.1365-2486.2012.02712.x
13
14 Ifarragueri, A. and C.I. Chang. 2002. "Unsupervised hyperspectral image analysis with
15
16 projection pursuit." *IEEE Geoscience and Remote Sensing* 38 (6): 2529-2538.
17
18 doi: 10.1109/36.885200
19
20 Ines, A.V.M., N.N. Das, J.W. Hansen and E.G. Njoku. 2013. „Assimilation of remotely
21
22 sensed soil moisture and vegetation with a crop simulation model for maize yield
23
24 prediction." *Remote Sensing of Environment* 138: 149-164. doi:
25
26 doi:10.1016/j.rse.2013.07.018
27
28 Keshava N. and J.F. Mustard. 2002. "Spectral Unmixing". *IEEE Signal Processing*
29
30 *Magazine* 19 (1): 44-57. doi: 10.1109/79.974727
31
32 Keshava N. 2003. "A survey of spectral unmixing algorithms". *Lincoln Laboratory*
33
34 *Journal* 14 (1): 55-78.
35
36 Khurshid, K.S., K. Staenz, L. Sun, R. Neville, P.H. White, A. Bannari, C.M.
37
38 Champagne and R. Hitchcock. 2006. "Preprocessing of EO-1 Hyperion data."
39
40 *Canadian Journal of Remote Sensing* 32 (2): 84-97. doi: 10.5589/m06-014
41
42 Kumar, L., K.S. Schmidt, S. Dury, A.K. Skidmore 2001. "Imaging spectrometry and
43
44 vegetation science." *Remote Sensing and Digital Image Processing* 4: 111-155.
45
46 doi: 10.1007/978-0-306-47578-8_5
47
48 Licciardi, G.A. and J. Chanussot. 2015. "Nonlinear PCA for Visible and Thermal
49
50 Hyperspectral Images Quality Enhancement." *IEEE Geoscience and Remote*
51
52 *Sensing* 4 (2): 201-205. doi: 10.1109/LGRS.2015.2389269
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6 Licciardi, G.A., P.R. Marpu, J. Chanussot and J.A. Benediktsson. 2011. "Linear Versus
7
8 Nonlinear PCA for the Classification of Hyperspectral Data Based on the
9
10 Extended Morphological Profiles." *IEEE Geoscience and Remote Sensing* 9 (3):
11
12 447-451. doi: 10.1109/LGRS.2011.2172185
13
- 14 Lu, D. and Q. Weng. 2004. "Spectral Mixture Analysis of the Urban Landscape in
15
16 Indianapolis with Landsat ETM+ Imagery." *Photogrammetric Engineering and*
17
18 *Remote Sensing* 70 (4): 1053-1062.
19
- 20 Martín-Queller, E., J.M. Diez, I. Ibanez and S. Saura. 2012. "Effects of silviculture on
21
22 native tree species richness: interactions between management, landscape context
23
24 and regional climate." *Journal of Applied Ecology* 50 (3): 775-785. doi:
25
26 10.1111/1365-2664.12064
27
- 28 Mitri, G.H. and I.Z. Gitas. 2010. "Mapping Postfire Vegetation Recovery Using EO-1
29
30 Hyperion Imagery." *IEEE Geoscience and Remote Sensing* 48 (3): 1613-1618.
31
32 doi: 10.1109/TGRS.2009.2031557
33
- 34 Nelson, M.D., R.E. McRoberts, G.R. Holden and M.E. Bauer. 2009. "Effects of satellite
35
36 image spatial aggregation and resolution on estimates of forest land area."
37
38 *International Journal of Remote Sensing* 30 (8): 1913-1940. doi:
39
40 10.1080/01431160802545631
41
- 42 Pal M. and Mather P.M. 2005. "Support vector machines for classification in remote
43
44 sensing." *International Journal of Remote Sensing* 26 (5): 1007-1011. doi:
45
46 10.1080/01431160512331314083
47
- 48 Pal, M. and P.M. Mather. 2006. "Some issues in the classification of DAIS
49
50 hyperspectral data." *International Journal of Remote Sensing* 27 (14): 2895-2916.
51
52 doi: 10.1080/01431160500185227
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6 Prasad, S. and L.M. Bruce. 2003. "Limitations of Principal Components Analysis for
7
8 Hyperspectral Target Recognition." IEEE Geoscience and Remote Sensing 5 (4):
9
10 625-629. doi: 10.1109/LGRS.2008.2001282
- 11
12 Pu, R., S. Bell, C. Meyer, L. Baggett and Y. Zhao. 2012. "Mapping and assessing
13
14 seagrass along the western coast of Florida using Landsat TM and EO-1
15
16 ALI/Hyperion imagery." Estuarine, Coastal and Shelf Science 115: 234-245. doi:
17
18 10.1016/j.ecss.2012.09.006
- 19
20 Pu, R., Q. Yu, P. Gong and G.S. Biging. 2005. "EO-1 Hyperion, ALI and Landsat 7
21
22 ETM+ data comparison for estimating forest crown closure and leaf area index."
23
24 International Journal of Remote Sensing 26 (3): 457-474. doi:
25
26 10.1080/01431160512331299324
- 27
28 Qian, D. and J.E. Fowler. 2007. "Hyperspectral Image Compression Using JPEG2000
29
30 and Principal Component Analysis." IEEE Geoscience and Remote Sensing 4 (2):
31
32 201-205. doi: 10.1109/LGRS.2006.888109
- 33
34 Ridd, M.K. 1995. "Exploring a V-I-S (Vegetation-Impervious-Soil) model for urban
35
36 ecosystem analysis through remote sensing: Comparative anatomy for cities."
37
38 International Journal of Remote Sensing 16 (12): 2165-2185. doi:
39
40 10.1080/01431169508954549
- 41
42 Serrano, L., J. Penuelas, S.L. Ustin 2002. "Remote sensing of nitrogen and lignin in
43
44 Mediterranean vegetation from AVIRIS data: decomposing biochemical from
45
46 structural signals." Remote Sensing of Environment 81 (2-3): 355-364.
47
48 doi:10.1016/S0034-4257(02)00011-1
- 49
50 Shao, Y. and R.S. Lunetta. 2012. "Comparison of support vector machine, neural
51
52 network, and CART algorithms for the land-cover classification using limited
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6 training data points.” *Journal of Photogrammetry and Remote Sensing* 70: 78-87.
7
8 doi: 10.1016/j.isprsjprs.2012.04.001
9
- 10 Story, M. and Congalton R.G. 1986. “Accuracy assessment - A user's perspective.”
11
12 *Photogrammetric Engineering and Remote Sensing* 52 (3): 397-399.
13
- 14 Tarabalka, Y., J.A. Benediktsson, J. Chanussot 2009. “Spectral–Spatial Classification of
15
16 Hyperspectral Imagery Based on Partitional Clustering Techniques” *IEEE*
17
18 *Geoscience and Remote Sensing* 47 (8): 2973-2981.
19
- 20 Thinkabail, P. S., E.A. Enclona, M.S. Ashton, C. Legg and M.J. De Dieu. 2003.
21
22 “Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African
23
24 rainforests.” *Remote Sensing of Environment* 90: 23-43. doi:
25
26 10.1016/j.rse.2003.11.018
27
- 28 Thonicke, K., A. Spessa, I.C. Prentice, S.P. Harrison, L. Dong and C. Carmona-
29
30 Moreno. 2010. “The influence of vegetation, fire spread and fire behavior on
31
32 biomass burning and trace gas emissions: results from a process-based model.”,
33
34 *Biogeosciences* 7: 1991-2011. doi:10.5194/bg-7-1991-2010
35
- 36 Tooke, R.T., N.C. Coops, N.R. Goodwin, J.A. Voogt. 2009. “Extracting urban
37
38 vegetation characteristics using spectral mixture analysis and decision tree
39
40 classifications.” *Remote Sensing of Environment* 113 (2): 398-407. doi:
41
42 10.1016/j.rse.2008.10.005
43
- 44 Varga, K., Sz. Szabó, G. Szabó, Gy. Dévai and B. Tóthmérész. 2015. “Improved land
45
46 cover mapping using aerial photographs and satellite images.” *Open Geosciences*
47
48 7 (1): 15-26. doi: 10.1515/geo-2015-0002
49
- 50 Vyas, D., N.S.R. Krishnayya, K.R. Manjunath, S.S. Ray and S. Panigrahy. 2011.
51
52 “Evaluation of classifiers for processing Hyperion (EO-1) data of tropical
53
54
55
56
57
58
59
60

1
2
3
4
5
6 vegetation.” International Journal of Applied Earth Observation and
7
8 Geoinformation 13 (2): 228-235. doi: doi:10.1016/j.jag.2010.11.007
9

10 Wang, X., Y. Zeng, S. Wang and T. Zhao. 2015. “Identification of Conifer Species
11
12 Based on Laboratory Spectroscopy and an Artificial Neural Network.” Journal of
13
14 Software Engineering 9 (2): 362-372. doi: 10.3923/jse.2015.362.372
15

16 Williams, D.J., N.B. Rybicki, A.V. Lombana, T.M. O’Brien and R.B. Gomez. 2003.
17
18 “Preliminary Investigation of Submerged Aquatic Vegetation Mapping using
19
20 Hyperspectral Remote Sensing.” Environmental Monitoring and Assessment 81
21
22 (1): 383-392. doi: 10.1023/A:1021318217654
23

24 Xu, Y., P. Yu, X. Guo, B. Gao, X. Gao. 2012. “On hyperspectral remotely sensed image
25
26 classification based on MNF and AdaBoosting.” International Conference on
27
28 Audio, Language and Image Processing, Shanghai, China, 2012, 605-609. doi:
29
30 10.1109/ICALIP.2012.6376688
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Tables and table captions

Table 1. Selected influential bands

Selected influential bands (51 bands)					
Band no.	Wavelength (nm)	Band no.	Wavelength (nm)	Band no.	Wavelength (nm)
36	711.72	57	884.7	102	1164.68
40	752.43	78	993.17	103	1174.77
41	762.6	82	1013.3	104	1184.87
42	772.78	83	1016.98	105	1194.97
43	782.95	84	1023.4	106	1205.07
44	793.13	85	1027.16	107	1215.17
45	803.3	86	1033.49	108	1225.17
46	813.48	87	1037.33	109	1235.27
47	823.65	88	1043.59	110	1245.36
48	833.83	89	1047.51	111	1255.46
49	844	90	1053.69	112	1265.56
50	851.92	91	1057.68	113	1275.66
51	854.18	92	1063.79	114	1285.76
52	862.01	93	1073.89	115	1295.86
53	864.35	94	1083.99	116	1305.96
54	872.1	95	1094.09		
55	874.53	96	1104.19		
56	882.19	97	1114.19		
Bands selected by SDA (46-47 bands)					
Band no.	Wavelength (nm)	Band no.	Wavelength (nm)	Band no.	Wavelength (nm)
39	742.25	78	993.17	105 407	1194.97 1215.17
40	752.43	83	1016.98	107 408	1215.17 1225.17
41	762.6	84	1023.4	108 409	1225.17 1235.27
42	772.78	85	1027.16	109 410	1235.27 1245.36
43	782.95	86	1033.49	110 411	1245.36 1255.46
44	793.13	87	1037.33	111 412	1255.46 1265.56
45	803.3	88	1043.59	112 413	1265.56 1275.66
46	813.48	89	1047.51	113 414	1275.66 1285.76
47	823.65	90	1053.69	114 415	1285.76 1295.86
48	833.83	91	1057.68	115 416	1295.86 1305.96
49	844	92	1063.79	116	1305.96
50	851.92	93	1073.89		
51	854.18	94	1083.99		
52	862.01	95	1094.09		
53	864.35	96	1104.19		
54	872.1	102	1164.68		
55	874.53	103	1174.77		
56	882.19	104	1184.87		

Table 2. Mean rule values on the mixed areas

	Beech-Maple	Oak-Beech	Oak-Maple	Oak-Pine
Beech	0.364	0.251	0.081	0.007
Maple	0.297	0.141	0.334	0.056
Oak	0.124	0.401	0.424	0.489
Pine	0.066	0.097	0.117	0.442
Ground vegetation	0.147	0.112	0.042	0.004

Table 3a. Accuracy and confusion matrices of the classification results for single species classes

Full dataset (Kappa Coefficient = 0.6871; Overall accuracy = 76.67%)						
	Oak	Ground veg.	Maple	Pine	Beech	
Oak	76.67	3.45	14.81	1.14	41.67	
Ground veg.	3.33	96.55	3.70	0	11.11	
Maple	6.67	0	55.56	5.68	2.78	
Pine	13.33	0	3.70	89.77	0	
Beech	0	0	22.22	3.41	44.44	
User's accuracy	52.27	82.35	65.22	94.05	64	
PCA (Kappa Coefficient = 0.7611; Overall accuracy = 81.92%)						
	Oak	Ground veg.	Maple	Pine	Beech	
Oak	93.33	0	59.26	0	4.35	
Ground veg.	0	93.10	0	13.64	0	
Maple	0	0	25.93	4.55	0	
Pine	3.33	6.90	14.81	81.82	1.45	
Beech	3.33	0	0	0	94.20	
User's accuracy	59.97	90	87.5	69.23	98.48	
SDA (Kappa Coefficient = 0.7702; Overall accuracy = 82.85%)						
	Oak	Ground veg.	Maple	Pine	Beech	
Oak	90	0	0	5.68	30.56	
Ground veg.	0	93.1	0	0	8.33	
Maple	0	0	85.19	4.55	5.56	
Pine	6.67	0	3.7	89.77	5.55	
Beech	3.33	6.9	11.11	0	50	
User's accuracy	62,79	90	79	94	75	
Influential bands (Kappa Coefficient = 0.7890; Overall accuracy = 84.28%)						
	Oak	Ground veg.	Maple	Pine	Beech	
Oak	90	3.45	0	2.27	30.56	
Ground veg.	0	93.10	0	0	8.33	
Maple	0	0	85.19	4.55	5.56	
Pine	6.67	0	3.70	92.05	2.78	
Beech	3.33	3.45	11.11	1.14	52.78	
User's accuracy	65.85	90	79.31	95.29	76	

Table 3b. Accuracy and confusion matrix of the classification results for mixed classes

Mixed Classes (Kappa Coefficient = 0.8329; Overall accuracy = 85.79%)									
	Oak	Ground veg.	Maple	Pine	Beech	Oak-Maple	Oak-Pine	Oak-Beech	Beech-Maple
Oak	83.33	0	0	0	0	4.35	2.41	20	0
Ground veg.	0	82.76	0	0	9.09	0	0	0	0
Maple	0	0	70.37	0	0	2.17	1.20	0	1.45
Pine	6.67	0	3.70	88.41	0	0	3.61	0	0
Beech	3.33	10.34	3.70	0	81.82	0	0	0	4.35
Oak-Maple	0	0	0	1.45	0	89.13	4.82	0	4.35
Oak-Pine	6.67	0	0	10.14	0	2.17	86.75	0	0
Oak-Beech	0	3.45	0	0	0	0	0	80	0
Maple-Beech	0	3.45	22.22	0	9.09	2.17	1.20	0	89.86
User's accuracy	83.33	92.31	86.36	91.04	69.23	83.67	87.8	80	84.93

Figure captions

Figure 1. The study area

Figure 2. Radiance spectra – *a*: Full dataset; *b*: 51 selected influential bands

Figure 3. Classification accuracy in case of influential bands selection and SDA reduction in relation to the optimal number of bands according to each method

Figure 4. Classification results – *a*: Image composite; *b*: Full dataset; *c*: PCA reduced; *d*: SDA reduced, *e*: 51 influential bands; *f*: 51 bands, mixed classes

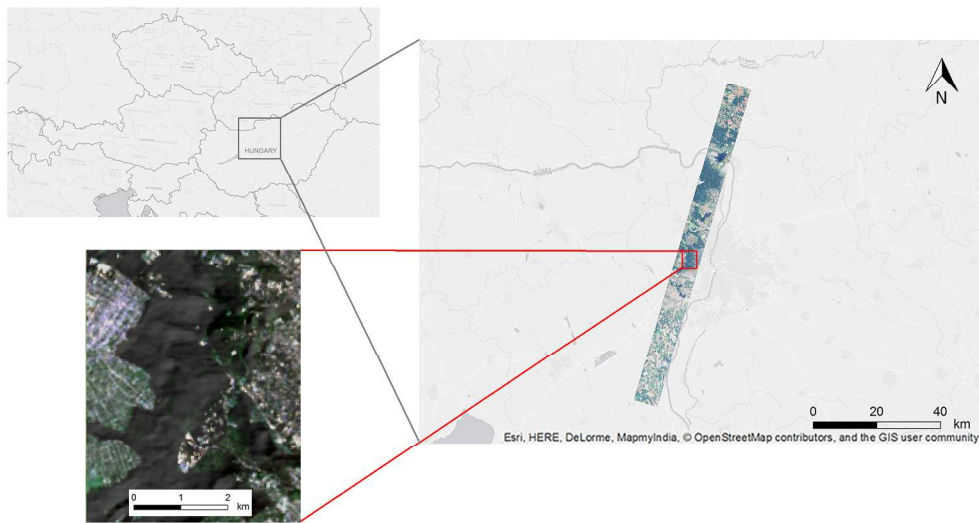


Figure 1. The study area

889x482mm (72 x 72 DPI)

Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

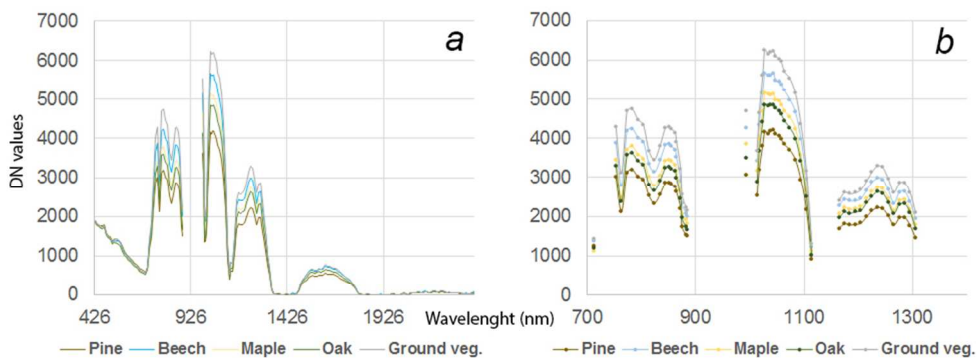


Figure 2. Radiance spectra – a: Full dataset; b: 51 selected influential bands

352x137mm (72 x 72 DPI)

Peer Review Only

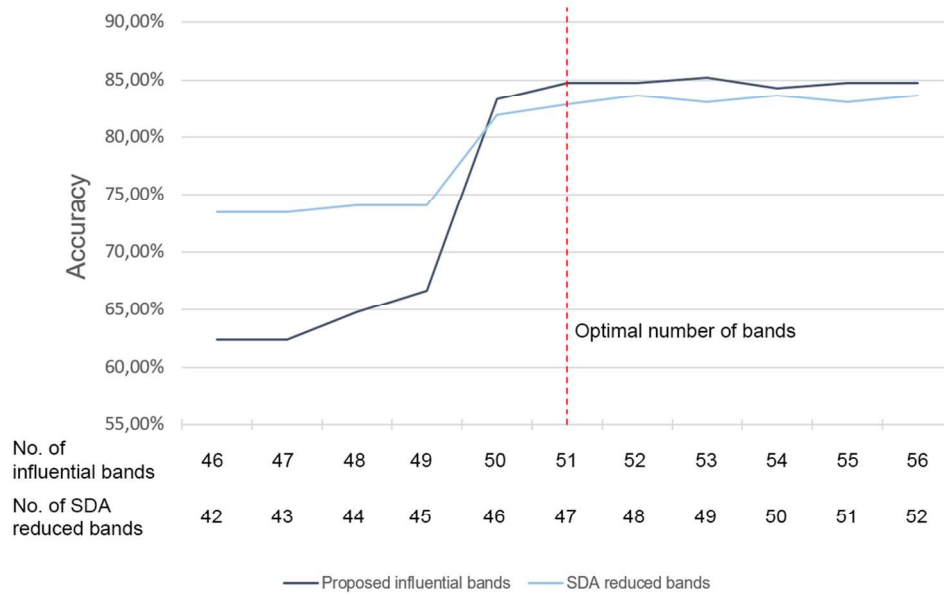


Figure 3. Classification accuracy in case of influential bands selection and SDA reduction in relation to the optimal number of bands according to each method

430x269mm (72 x 72 DPI)

Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

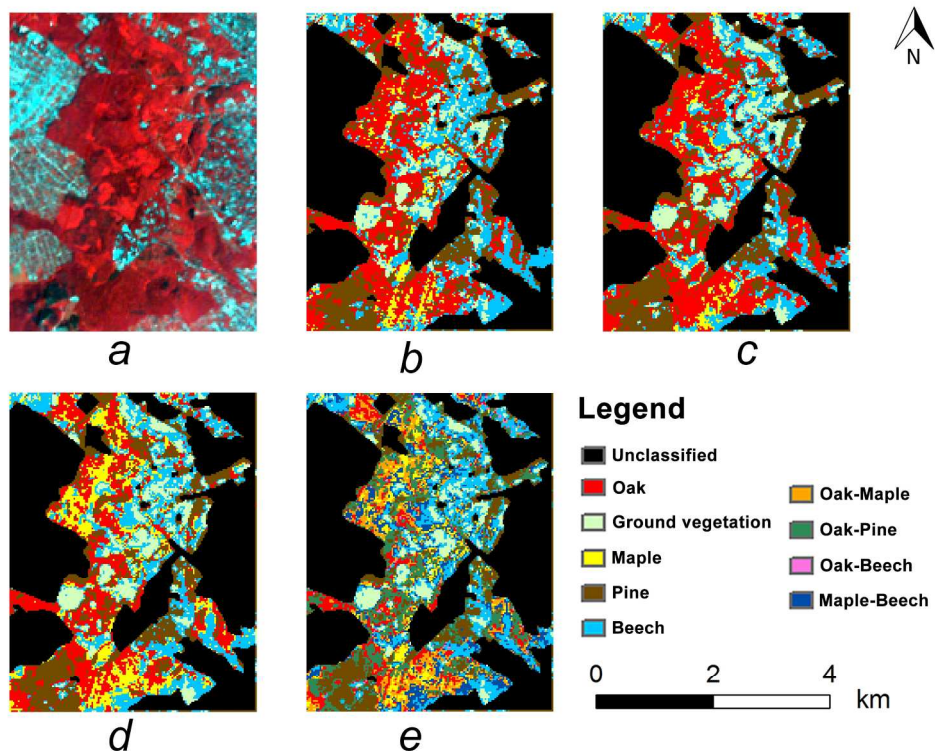


Figure 4. Classification results – a: Image composite; b: Full dataset; c: PCA reduced; d: SDA reduced, e: 51 influential bands; f: 51 bands, mixed classes

289x228mm (200 x 200 DPI)

www Only