UNIVERSITY OF SZEGED

Faculty of Science and Informatics

Doctoral School of Earth Sciences

Department of Physical Geography and Geoinformatics

Applicability of Hyperspectral Remote Sensing Methods for Landscape Analyses

Theses of PhD Dissertation

Bálint Csendes

Supervisors:

Erzsébet Merényi, PhD, Research Professor, Rice University

László Mucsi, PhD, Associate Professor, University of Szeged

Szeged

2016

Introduction

Current remote sensing technology has made it possible for researchers to map country-scale land cover so precisely that even smaller changes in surface material quality are identifiable.

Hyperspectral sensors installed on aircrafts and satellites are often used for environmental monitoring. They provide much greater amount of data than multispectral systems, while field surveys do not necessarily provide as much information as the classification-evaluation process requires.

As a result, one of the most important aims of recent remote-sensing-based Earth Science research is the development of automated data processing techniques as well as statistical image processing methods, which provide reliable results despite limited field data collection.

The topic of my doctoral dissertation is the landscape analysis application of AISA hyperspectral data. I aimed at applying internationally accepted image processing methods to environmental geographical phenomena typical of certain Hungarian areas. Also, my goal was to develop new statistical methods for the spectral discrimination of specific surface types such as oversaturated soils or floodplain plant species.

Aims

1. Hyperspectral remote sensing technology has been used world wide since the 1980s, while it has been used regularly for environmental analyses in Hungary since the new millennium. I intended to find out how hyperspectral imaging can be applied to vegetation and inland excess water mapping as well as to differentiating between urban and other artificial surfaces in Hungary. I also wanted to learn about its data processing and field analysis requirements, so besides aerial photos, I included the LiDAR elevation data in my analysis although they cover the area of the aerial photos only partially. In addition, I aimed at finding a solution to mapping shrub layer plants by applying the available hyperspectral data.

2. Atmospheric or sensor-induced noise often makes it difficult to identify spectral patterns, therefore it is useful to remove more affected data bands from the analysis. However, remote sensing research does not have a widely used method for doing so. Hence I intended to analyze the spectral discrimination techniques of reflectance curves as well as the mathematical-statistical background of their redundancy reduction, to introduce certain algorithms, then to apply them in order to find a solution for minimizing the image noise of hyperspectral data.

3. In order to prepare a high accuracy map of land cover we have to know the inner homogeneity of the classes and the standard deviation of the pixel reflectance values. Data normalization can be performed when the previously mentioned data are at our disposal, at the same time, classification efficiency is also influenced by the spectral distribution curves in a given mapping category. Therefore, I aimed at developing a homogeneity analysis method that can also be applied to hyperspectral data containing several hundreds of data bands.

4. The spectral values of the training areas often overlap to some extent, which may disturb the image classification algorithms, however, they may also provide important information concerning the studied surfaces. I applied spectral separability measurements to the transformed hyperspectral data samples in order to be able to differentiate between the classes used for mapping, then I compared these values to the extent of misclassifications.

5. Furthermore, I aimed at comparing certain supervised classification algorithms on the basis of their accuracy, reliability and required computing capacity. In order to prepare a suitable critical analysis I used a wide variety of sample sites from Tápai-rét because I considered it important to analyze the efficiency of the methods in a remote sensing study with specific requirements, which means that the field data collection of the sample points is either temporally or spatially limited, and the relatively high amount of classes makes it even more difficult to identify surface types.

6. Remote sensing technology has several methods of image classification. Unsupervised as well as parametric and non-parametric supervised classifications may be applied to data interpretation, although their efficiency depends on limiting factors such as the lack of training points, the presence of image noise or the lack of detailed field information. I intended to answer the following questions: what kind of remote sensing application requires unsupervised classification, and when is it more effective to use training areas? I compared the relative advantages and disadvantages of the two approaches by using images from Hungarian sites of both vegetated surfaces and sample areas covered by inland excess water; I also paid attention to the differences arising from the theoretical background of the algorithms as well as the spatial patterns of the analyzed geographical phenomena and environmental processes.

Results

1.

Hyperspectral airborne photographs were taken on 22 September 2010, at the end of the vegetation period. Their geometric accuracy being 1.5 m they proved to be especially suitable for detailed vegetation mapping. The 4.29 to 6.28 nanometer spectral resolution made it possible to separate the vegetation surfaces of Tápai-rét according to plant species, and to compile spectral libraries, which was preceded by the total survey and identification of the habitats of the sample area and the agricultural parcels as well as by the designation of the training points. I managed to trace the presence of the invasive desert false indigo (*Amorpha fruticosa*) stands in the shrub layer of the hybrid poplar planting by means of spectral angle mapping (SAM). My results showed that 1.5% of the pixels covering the studied forest patch displayed the average spectrum of the desert false indigo with a reflectance deviation of lower than 0.03 radian, and another 10.9% showed a spectral angle of deviation of 0.03 to 0.06 radians compared both to the spectral curves of the desert false indigo and the poplar stands. The hyperspectral data also proved to be reliable and accurate enough to separate and map artificial surfaces and areas affected by inland excess water, where the class of wet soil surfaces coincided with the higher pixel values of the Flow Accumulation Map prepared by using the data of the LiDAR elevation model. Classifications on soil patches disturbed by construction works had relatively poor results.

2.

When removing noise-affected bands, I considered the spatial autocorrelation existing between the values of the pixels, which I calculated with Geary's *C* formula, and I filtered the anomalies caused by different light intensity by means of normalization. I carried out principal component transformation to further reduce redundancy, and I depicted the main surface types in the spectral space of the first two principal component bands, free of the albedo effect, in a scatter plot diagram.

3.

I studied the spectral homogeneity of the classes by applying two approaches, one of which meant computing the standard deviation of the reflectance values of every fiftieth spectral channel, the other was depicting the distribution of data in a box-plot figure. I developed a method for the latter approach, a technique applicable to hyperspectral images: I determined homogeneity on the basis of the spectral angle deviation of the pixel values from the class average. Of the values calculated in radians, I plotted the maximum, the minimum and the average angles of deviation as well as the difference between the third and the first quartiles in a box-plot diagram.

4.

I calculated the separability of the training areas with both the Jeffries-Matusita and the Transformed Divergence algorithms. On a scale from 0 to 2, the lower values indicate the more significant spectral overlaps, while separability index 2 indicates total separability. I used the first five principal component bands for the analysis, as the AISA data contain a lot more spectral channels than the number of the surveyed sample points.

JM	**Jeffries-Matusita index**	**TD**	**Transformed Divergence index**
A	Green ash	K	White willow
B	Asphalt	M	Grey poplar
C	Concrete	N	Oak
D	Undisturbed soil surface	U	Vegetation in water
E	Disturbed soil surface	V	Open water
F	Red roof tiles	W	Medick
G	Desert false indigo	X	Corn
H	Grass(land) patches	Y	Garlic
J	Poplar hybrids	Z	Sugar beet

	A	B	C	D	E	F	G	H	J	K	M	N	U	V	W	X	Y	Z	**TD**
A	**0**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1.9**	**1.19**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	A
B	**2**	**0**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	B
C	**2**	**2**	**0**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	C
D	**2**	**2**	**2**	**0**	**1.93**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	D
E	**2**	**2**	**2**	**1.6**	**0**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	E
F	**1.99**	**1.85**	**1.83**	**2**	**2**	**0**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	F
G	**1.78**	**2**	**2**	**2**	**2**	**2**	**0**	**2**	**1.92**	**1.99**	**1.96**	**1.99**	**2**	**2**	**2**	**2**	**2**	**2**	G
H	**1.81**	**2**	**2**	**2**	**2**	**2**	**2**	**0**	**2**	**1.99**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	H
J	**1.64**	**2**	**2**	**2**	**2**	**2**	**1.8**	**2**	**0**	**1.54**	**1.04**	**0.8**	**2**	**2**	**2**	**2**	**2**	**2**	J
K	**1**	**2**	**2**	**2**	**2**	**2**	**1.89**	**1.87**	**1.45**	**0**	**1.89**	**1.94**	**2**	**2**	**2**	**2**	**2**	**2**	K
M	**1.85**	**2**	**2**	**2**	**2**	**2**	**1.85**	**2**	**0.87**	**1.56**	**0**	**1.29**	**2**	**2**	**2**	**2**	**2**	**2**	M
N	**1.79**	**2**	**2**	**2**	**2**	**2**	**1.92**	**2**	**0.7**	**1.82**	**1.05**	**0**	**2**	**2**	**2**	**2**	**2**	**2**	N
U	**1.9**	**2**	**2**	**2**	**2**	**2**	**1.98**	**2**	**1.7**	**1.98**	**1.93**	**1.92**	**0**	**2**	**2**	**2**	**2**	**2**	U
V	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**0**	**2**	**2**	**2**	**2**	V
W	**2**	**2**	**2**	**2**	**2**	**2**	**1.99**	**2**	**2**	**2**	**1.96**	**2**	**2**	**2**	**0**	**2**	**2**	**2**	W
X	**1.98**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1.99**	**2**	**2**	**2**	**2**	**2**	**0**	**1.96**	**1.68**	X
Y	**1.95**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1.98**	**2**	**2**	**2**	**2**	**2**	**1.89**	**0**	**1.11**	Y
Z	**1.97**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**1.99**	**2**	**2**	**2**	**2**	**2**	**1.56**	**1.02**	**0**	Z
JM	A	B	C	D	E	F	G	H	J	K	M	N	U	V	W	X	Y	Z	

Tables 1 and 2. Separability indexes of the studied classes

The resulting index values show that the Jeffries -Matusita method is more sensitive to the spectral overlaps of the classes than the Transformed Divergence method (Tables 1 and 2), however, both separability studies predicted which class may have a higher extent of misclassification. For example, green ash (A, *Fraxinus pennsylvanica*) and grey poplar (M, *Populus x canscens*) both demonstrate the relationship between low (< 1.5) separability values and poorer classification accuracy.

5.

In order to map the typical mosaic-like landscape of Tápai-rét, which includes natural, adventive, and agricultural plant species as well as concrete platforms used for petroleum extraction, I used image classification methods that provide reliable results even when only a small number of sample points are available. The SAM method had a more modest accuracy (64%), so I performed further classifications, among which the non-parametric Support Vector Machine (SVM) machine learning algorithm provided the smallest extent of misclassification (overall accuracy: 97%, kappa index: 0.97). On the basis of former studies I applied the Maximum Likelihood classifier as a parametric method, but I had to limit the number of input data bands in accordance with the available sample points. Despite the information loss that occurred during principal component transformation, the overall accuracy of this method was also over 90%. Taking the limited number of training points into consideration, I combined unsupervised clustering with supervised machine learning in a so-called hybrid SOM (Self-Organizing Map) system in order to eliminate overfitting.

P. A.	**Producer's Accuracy**	**U. A.**	**User's Accuracy**
A	Green ash	K	White willow
B	Asphalt	M	Grey poplar
C	Concrete	N	Oak
D	Undisturbed soil surface	U	Vegetation in water
E	Disturbed soil surface	V	Open water
F	Red roof tiles	W	Medick
G	Desert false indigo	X	Corn
H	Grass(land) patches	Y	Garlic
J	Poplar hybrids	Z	Sugar beet

	A	B	C	D	E	F	G	H	J	K	M	N	U	V	W	X	Y	Z	U.A.
A	**9**																		**1**
B		**14**						**1**											**93%**
C		**1**	**15**																**94%**
D				**10**	**1**														**91%**
E				**5**	**14**														**74%**
F						**14**													**1**
G							**13**				**1**								**93%**
H							**2**	**14**	**1**										**82%**
J									**9**		**1**	**2**							**75%**
K	**2**									**14**	**1**	**1**	**1**						**74%**
M	**1**						**1**		**3**	**1**	**12**	**1**							**63%**
N	**1**								**2**			**12**							**80%**
U	**1**												**14**						**93%**
V														**17**					**1**
W															**15**				**1**
X																**13**	**1**		**93%**
Y																**1**	**13**	**3**	**77%**
Z																**1**	**1**	**12**	**86%**
P. A.	**64 %**	**93 %**	**1**	**67 %**	**93 %**	**1**	**81 %**	**93 %**	**60 %**	**93 %**	**80 %**	**75 %**	**93 %**	**1**	**1**	**87 %**	**87 %**	**80 %**	**86 %**

Tables 3 and 4. Confusion matrix of the SOM classification

The artificial neural network method provided 86% accuracy in the first point distribution fold, while the mean of the cross-validation folds was 82.4%, a more significant misclassification occurred only in the cases of green ash (A, *Fraxinus pennsylvanica*) and hybrid poplar (J, *Populus sp.*) (Tables 3 and 4). Supervised, non-parametric methods (SAM and SVM) required significantly less computing capacity concerning both memory use and processing time than iterative, unsupervised classifications, including the clustering method of SOM. Beside the previously mentioned processes, data transformation and conversion operations may also cause further difficulties and delays.

6.

The extent of inland excess water shows significant temporal and spatial variations, as a result, the accurate indication of the often a few-centimeter-deep water patches and oversaturated soil surfaces was not always possible. However, I managed to define training areas during the inland excess water periods of the years from 2010 to 2013, both on oversaturated soil surfaces and intact, dry soil surfaces, the sample points of which were used in the supervised classification of hyperspectral data. Dry and oversaturated soil surfaces were separated with an 84% accuracy by the SAM method, and with a 90% accuracy with binary spectrum coding. At the same time, the limitations of the training point classifications became clear, so I also carried out ISODATA clustering on the subset image, the result map of which provided a 100% accuracy, and it also followed the spatial pattern of inland excess water a lot more closely. Supervised classifications provided better results when mapping vegetation species (SVM: 97% overall accuracy, Maximum Likelihood: 94% overall accuracy), because these algorithms, though susceptible to overfitting, are more sensitive to the subtle spectral differences of the plant species which are difficult to separate even in the field, especially when the data is noisy.

List of Publications Related to the Theses

Csendes, B. (2012): Detection of invasive plants on the flood plain of river Tisza, using hyperspectral airborne imagery. In: Tomislav Malvić, János Geiger, & Marko Cvetković (eds.): Conference book *"Geomathematics as geoscience":* 4th Croatian-Hungarian and 15th Hungarian geomathematical congress, pp. 187-194.

Csendes, B. (2013): Invazív növények spektrális tulajdonságainak vizsgálata légifotókon. In: Gábor Keresztes (ed.): *Tavaszi Szél, 2013: Spring wind*, pp. 408-414.

Csendes, B. & Mucsi, L. (2016): Inland excess water mapping using hyperspectral imagery. *Geographica Pannonica* Vol. 20 Issue 4, in print.

Tobak, Z., **Csendes, B.**, Henits, L., van Leeuwen, B., & Mucsi, L. (2012): A városi felszín spektrális tulajdonságainak vizsgálata légifelvételek alapján. In: Diána Nyári (ed.): *Kockázat - Konfliktus - Kihívás: A VI. Magyar Földrajzi Konferencia, a MERIEXWA nyitókonferencia és a Geográfus Doktoranduszok Országos Konferenciájának Tanulmánykötete*, pp. 1088-1097.

Tobak, Z., **Csendes, B.**, Henits, L., van Leeuwen, B., Szatmári, J., & Mucsi, L. (2012): Városi felszínek spektrális tulajdonságainak vizsgálata légifelvételek alapján. In: József Lóki (ed.): *Az elmélet és gyakorlat találkozása a térinformatikában III.* - Térinformatikai Konferencia és Szakkiállítás, Debrecen, 2012.05.24.-2012.05.25. pp. 413-420.

Tobak, Z., **Csendes, B.**, Henits, L., van Leeuwen, B., & Mucsi, L. (2013): Légifelvételek spektrális és térbeli információtartalmának felhasználása a városi felszínborítás térképezésében. In: József Lóki (ed.): *Az elmélet és a gyakorlat találkozása a térinformatikában IV.* - Térinformatika Konferencia és Szakkiállítás, Debrecen, 2013.05.23.-2013.05.24. pp. 441-450.