LAND COVER ANALYSIS BASED ON DESCRIPTIVE STATISTICS OF SENTINEL-2 TIME SERIES DATA

ORSOLYA GYÖNGYI VARGA[1]*, ILDIKÓ GOMBOSNÉ NAGY[2], PÉTER BURAI[3], TAMÁS TOMOR[4], CSABA LÉNÁRT[5], SZILÁRD SZABÓ[6]

[1,6] University of Debrecen. Department of Physical Geography and Geoniformation Systems, Hungary
[2] Envirosense Hungary Kft., Hungary
[3, 4, 5] University of Debrecen, Remote Sensing Centre, Hungary
*Email: varga.orsolya.gyongyi@science.unideb.hu

Abstract
In our paper we examined the opportunities of a classification based on descriptive statistics of NDVI throughout a year's time series dataset. We used NDVI layers derived from cloud-free Sentinel-2 images in 2018. The NDVI layers were processed by object-based image analysis and classified into 5 classes, in accordance with Corine Land Cover (CLC) nomenclature. The result of classification had a 76.2% overall accuracy. We described the reasons for the disagreement in case of the most remarkable errors. .

1. Introduction

Time series analysis is a widely popular application of satellite data, because it gives an opportunity to monitor specific phenomena or general changes in land cover in details (Szabó et al. 2019). There are a wide range of examples in the literature for time-series analysis based on certain satellite sensors (Lunetta et al. 2006; Gulácsi and Kovács 2015; Kovács 2018; Macfarlane et al. 2017) or on combined usage of different sensors (Kinthada et al. 2014; Wu et al. 2012). This latter application partly solved the problem of the availability of continuous cloud-free data, since acquisition times of different satellites are usually not uniform. This increases the chance of setting up a time-series dataset of as frequent cloud-free acquisitions as

possible over the same area. In order to handle these datasets together, scientists use spectral indices which are calculated from different spectral bands (Balázs et al. 2018, Szabó et al. 2016). However, due to the various characteristics of band wavelenghts and other factors, e.g. differences in radiometric correction, this way of usage may be incorrect without further processing (Gan 2014; Rodriguez et al. 2013; Guyot and Gu 1994). There is even a NASA initiative for the harmonization of the data of the two most popular satellite systems, Landsat and Sentinel-2 (Martin et al. 2018). A simple way to gain a comprehensive time-series dataset without further processing is to use a single sensor with a relatively frequent revisit time. For this purpose Sentinel-2 is ideal with 5 days revisit time and an outstanding (10 m)

spatial resolution among freely available multispectral satellite data sources.

Application of spectral vegetation indices in order to determine land cover and even the changes in these spectral indices over a certain period is a popular research topic in a wide range of research fields, like forestry, agriculture and general land cover change detection (Bolanio et al. 2015; Banskota et al. 2014; Nath and Acharjee 2013; Sahebjalal and Dashtekian, 2013). However, many of these studies use pixel-based classification methods when analysing the spectral index values or the changing of spectral index values (Defries and Townshend 1994, Gandhi et al. 2015).

Our aim was to analyse a discrete time-series dataset of Sentinel-2 satellite imagery from 2018, focusing on the vegetation period. Solely based on the descriptive statistical parameters of the time-series dataset, we performed an object-based classification (OBIA) in order to create a land cover map which we compared to the newly released Corine Land Cover 2018 (CLC2018) database. This paper is one of the very first applications of CLC2018. We intended to

analyse the changes of different land cover classes throughout the year, furthermore, to reveal how the vegetation index changes of pixels and segmentation of similarly changing pixels can help in discriminating the classes.

2. Materials and Methods

Study site

The study site was a 25*25 km area around Tokaj, NE Hungary. The area is characterised by a quite inhomogenous land cover, because it lies on the boundary of 5 different microregions (Fig. 1). It is advantageous, because in a study site with inhomogenous land cover we could observe the behaviour of many land cover types. However, the area has a permanent low ratio of changing areas, which added up to 2% of the study site in all 6 year time intervals of 2000-2006, 2006-2012 and 2012-2018, respectively, according to CORINE land cover change layers describing the changes in these specific time intervals (Varga et al. 2019). This low ratio of changes makes it possible to analyse the behaviour of different land cover classes, without handling changes, thus the changes adds a maximum
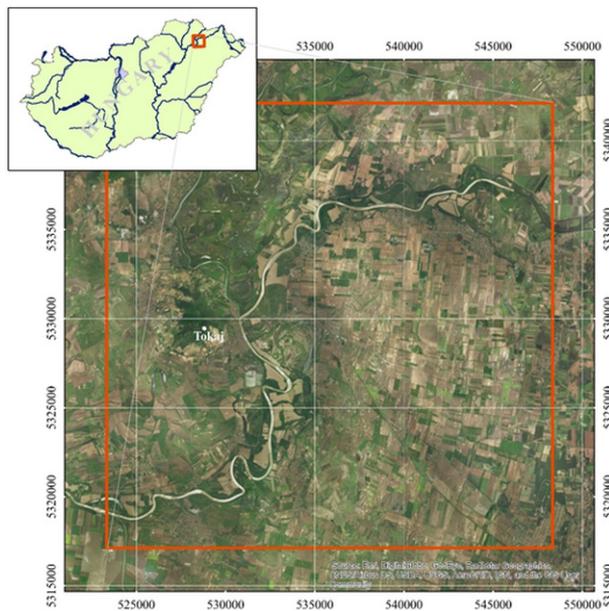

Fig. 1: Location of the study area, assigned by a red rectangle

of 2% error to the classification error in the area.

Datasets

We used data derived from Level-2A (atmospherically corrected) Sentinel-2 images. Sentinel-2A had been launched in 2015, Sentinel-2B had been launched in 2017. Sentinel-2 is a high-resolution multispectral imaging system, with 290 km wide swath, 10, 20 and 60 m resolution in 13 different bands and 5 days revisit time (achieved by Sentinel 2A and 2B together). We selected cloud-free images covering the study area throughout the year of 2018. Based on cloud conditions, it was not possible to set up an equidistant dataset, but one image per month most of the vegetation period was available, between May and November excluding July. The red (central wavelength: 665 nm) and infrared (central wavelength: 842 nm) bands have 10 m spatial resolution and these bands are the basis for calculating a widely used vegetation index for estimating biomass, called NDVI (Normalized Difference Vegetation Index). NDVI is generally calculated as in Equation 1 (Baret et al. 1989).

$$NDVI = (NIR-Red)/(NIR+Red) \quad (Eq.1.)$$

In Figure 2. we present the whole processing chain. We calculated NDVI for all the selected images, and then extracted the following descriptive statistical values in ArcMap on a pixel-by-pixel basis throughout the dates in 2018: minimum, maximum, mean, median, standard deviation and range. We performed multiresolution segmentation with a largest scale parameter of 500 in eCognition software (Blaschke 2010). Then we performed a supervised nearest neighbor classification, based on the above mentioned descriptive statistical layers, moreover considering various shape characteristics of the segments (asymmetry, compactness, rectangular fit, roundness, shape index). The shape characteristics may help to isolate the categories more successfully, due to the characteristic shape of segments of some classes, e.g. the rectangular shape of agricultural parcels. Object-based image analysis is a popular method in landscape-related analysis, there are many examples of its application in even Hungarian study areas. (Bertalan et al. 2016; Túri 2015; Varga et al. 2014).

The latest version of Copernicus Land Cover database has been released in January 2019 and is distributed by the European
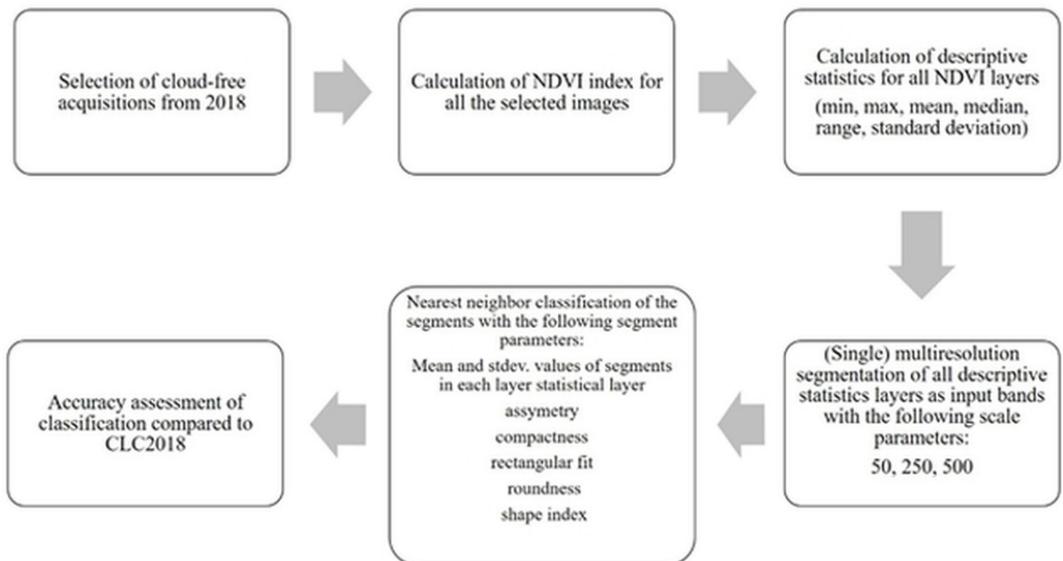

Fig. 2: Step-by-step processing workflow of Sentinel-2 images

Environment Agency and the Copernicus Land Monitoring Service. This database characterizes the status of land cover in 2017-2018 and was created on the basis of satellite images with 100 m positional accuracy. Minimum mapping unit was 25 ha, thematic accuracy was >85% and the features were classified into 44 categories in a 3-level hierarchical nomenclature system (Copernicus website, 2019). CLC2018 is an ideal database for analysing the land cover status based on a time series dataset covering the vegetation period of 2018. After processing, both of the maps (classification result and CLC2018) had the same classes according to the CLC2018 level-1 nomenclature with 5 classes (Feranec et al. 2016). The more diverse classes of the Sentinel-2 classification has also been aggregated into the CLC2018 level-1 nomenclature.

Accuracy Assessment

We resampled either the classification results map and the CLC2018 vector layer into 100 m resolution raster maps in order to create maps with similar resolution and class aggregation characteristics. These maps were compared to each other and overall accuracy, quantity disagreement, allocation disagreement and total disagreement were calculated. These indices give a comprehensive insight into the thematic agreement of the maps (Pontius and Millones, 2011), and show the ratio of pixels which belong to the same class in both maps and errors with different reasons.

3. Results

All the computed NDVI descriptive statistics layers have been inputs in the segmentation process, and they could support the isolation of different categories even visually. In Figure 3 we present examples of the NDVI-based descriptive statistics layers.

As a result of classification in eCognition, a land cover map had been created with five classes in accordance with the Corine 5 main land cover categories (Fig. 4). These categories are the followings: artificial surfaces (AR), agricultural areas (AG), forest and semi-natural areas (FT), wetlands (WL), water bodies (WT).

Table 1. includes the error matrix and the calculated accuracy assessment indices. The overall agreement of the two maps was 76.2%. Table 2 provides the ratio of total disagreement which was 23.8%. Disagreement can be analysed in details through quantity and allocation disagreement. Quantity disagreement, 4.1% in this case, is due to the imperfect match in the category proportions of the study site in the two maps. Allocation disagreement, 19.7% in this case, is due to the imperfect match in the spatial allocation of the categories between the two maps.
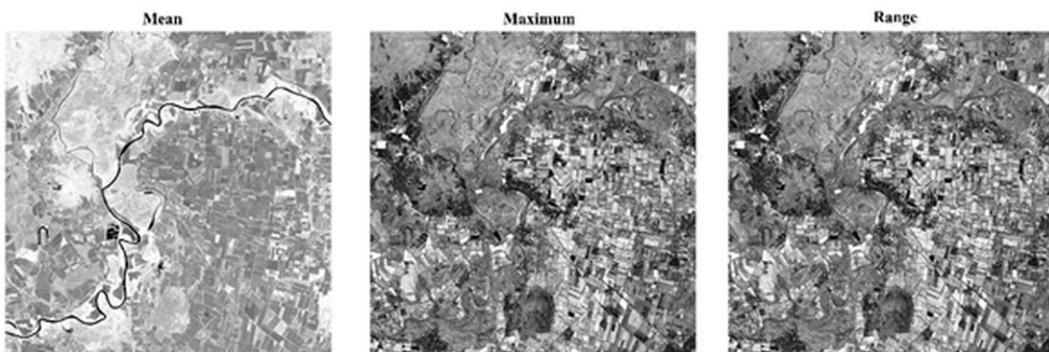

Fig. 3: Layers including mean, maximum and range values of the time series dataset on a pixel-by-pixel basis
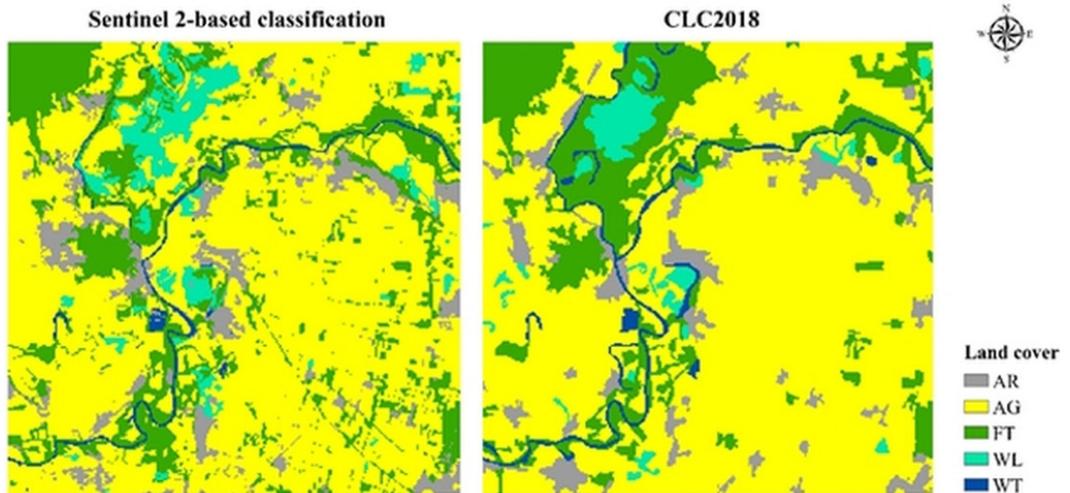
Fig. 4: Results of the classification of Sentinel-2 statistics layers (aggregated into CLC classes) and relevant clipped area of CLC2018 database (AR = artificial surfaces; AG = agricultural areas; FT = forest and semi-natural areas; WL = wetlands; WT = water bodies)

The largest errors were present in case of agricultural and forest areas (3995 and 2922 pixels, highlighted in Table 1). It means 3995 pixels which were forest according to CLC2018, were classified into agricultural areas in Sentinel-based classification. Moreover, 2922 pixels which were agricultural areas according to CLC2018, were classified into forest areas in Sentinel-based classification. Lower, but still important errors, around 1000 pixels, can be observed between wetland and agricultural

Table 1: Crosstabulation matrix of the Sentinel-based classification map and CLC2018 map. (OA = Overall accuracy, PA = Producer's accuracy, OE = Omission Error, UA= User's accuracy, CE = Commission Error)

		CLC2018						
	AR	AG	FR	WL	WT	Total	UA (%) / CE	
Sentinel-2 classification	AR	**2894**	1143	422	2	3	4464	65 / 35
	AG	1680	**36217**	3995*	1168	40	43100	84 / 16
	FT	72	2922*	**6770**	1341	20	11125	61 / 39
	WL	40	721	565	**832**	0	2158	39 / 61
	WT	122	219	342	52	**918**	1653	56 / 44
	Total	4808	41222	12094	3395	981	**62500**	
	PA (%) / OE	60 / 40	88 / 22	56 / 44	25 / 75	94 / 6		OA = 76.2

Table 2. Quantity disagreement, allocation disagreement and, calculated as their sum, the total disagreement of the two maps. These indices were derived from Table 1.

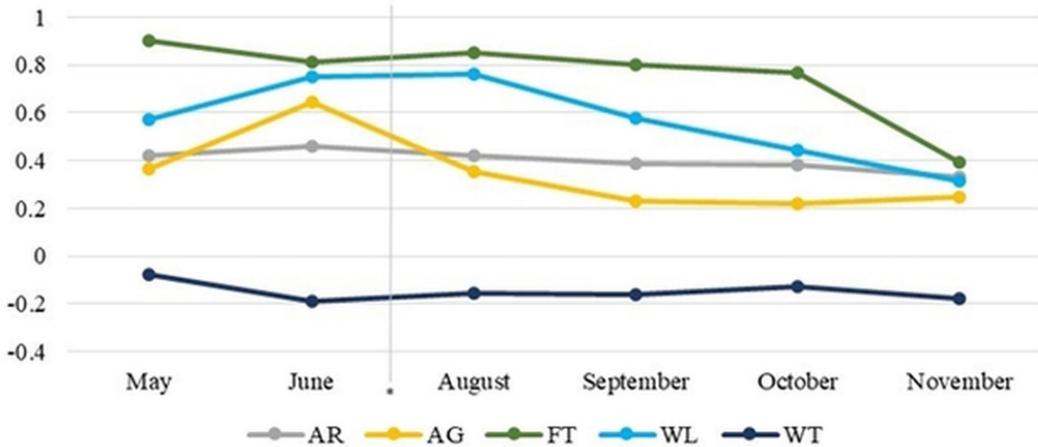|	
Quantity Disagreement (Q)	4.1
Allocation Disagreement (A)	19.7
Q+A=Total Disagreement (T)	23.8

Fig. 5. Changes of NDVI values in samples of different classes throughout specific months of 2018 (AR = artificial surfaces; AG = agricultural areas; FT = forest and semi-natural areas; WL = wetlands; WT = water bodies). Vertical grey line assigns that data was not available in July

areas, between wetland and forest areas, and between agricultural areas and artificial surfaces.

Fig. 5. depicts the changes in NDVI values throughout 2018 vegetation period. There was no data available in July due to disadvantageous cloud conditions, so we excluded the data of that month during the analysis. One sample of each category was used for setting up the diagram. Each sample was used as a training segment in classification process. Artificial, forest and water samples showed values throughout the examined period, but with remarkably different intensity of NDVI values. Agricultural and wetland areas showed larger variety with similar curves, but with different intensity of NDVI values.

4. Discussion

In the ratios of land cover classes of the area between the two maps, there is a low disagreement, as it is shown by quantity disagreement. It means that the differences in allocation of classes (Allocation Disagreement = 19.7%) causes more error than differences in quantity (Quantity Disagreement = 4.1%), because circa 80% of total disagreement was caused by allocation disagreement.

It is important to consider which classes showed higher omission or comission errors. Errors between wetland and agricultural areas, between wetland and forest areas, and between agricultural areas and artificial surfaces were also characteristic. The main reason for these phenomena is that all of these classes involve green areas or herbaceous vegetation which are hard to distinguish even in visual or spectral way. Forest areas, according to the CLC nomenclature, involve semi-natural areas (natural grasslands or transitional woodlands), agricultural areas involve pastures, and wetlands involve green areas with vegetation and partly woody areas. These types of land cover may have similar changes throughout the year and as reflected in NDVI statistics. Figure 5 showed the changes of classes within the examined time period and makes it clear which classes are likely to distinguish (forest, water, artificial), since they have remarkably different dynamics of changes, consequently they could be described by remarkably different statictics, like maximum or range. Samples used in Figure 5 assign typical segments of the classes which not contain mixed area. For example forest class sample contains an area covered by forest, but wetland sample does not contain wetland with a large amount of trees which would modify the NDVI change

dynamics of the typical sample.

The errors between agricultural and artificial areas are partly caused by the texture of these areas. As the classification was performed on segments created with a 500 scale parameter, some agricultural parcels had quite heterogenous segments, meaning a group of pixels with quite various spectral characteristics and statistics. These parcels contained small parcels with different vegetation which showed various changes troughout the year. Artificial areas, mainly urban areas, have the same characteristics due to the fact they are a mixture of a wide range of land cover types (artificial objects like buildings and roads, trees and parks, and gardens with various land cover) (Deák et al. 2016; Hüse et al. 2016). Due to this heterogeneity, confusion of these two classes is possible as well.

5. Conclusions

We analysed the possibilities of land cover classification based on object-based image analysis of NDVI-derived statistics layers. We intended to create classes based on the similar spectral behaviour of pixels which belong to the same land cover class. The classification based on Sentinel-2 images had 76.2% overall agreement and 23.8% total disagreement with the CLC 2018 dataset. The disagreement originated mainly from allocation disagreement. The error of classification of certain classes can be explained by two main factors: (1) the similar spectral characteristics of different land cover (agricultural area, forest and semi-natural area, wetland) types which belong to various classes in CLC nomenclature; (2) the heterogeneity characteristics of urban and small-parcel croplands. In certain classification procedures, like object-based image analysis, it is a possibility to define classes based on thresholds that originates from observing the changes of pixel values in different bands or in layers derived from bands. Our results showed how the pixels of different classes may vary through time

and how descriptive statistical layers can contribute to isolate them. This may help to determine new factors, combination of factors or a tighter group of factors along which the success of classification can be improved in the future.

6. References

Balázs, B. – Bíró, T. – Dyke, G. – Singh, S.K.; Szabó, Sz. (2018): Extracting water-related features using reflectance data and principal component analysis of Landsat images. Hydrological Sciences Journal 63: 2 pp. 269-284., DOI: 10.1080/02626667.2018.1425802

Banskota, A. – Kayastha, N. – Falkowski, M.J. – Wulder, M.A. –Frose, R.E: – White, J.C. (2014): Forest Monitoring Using Landsat Time Series Data: A Review. Canadian Journal of Remote Sensing, 40:362–384., DOI: 10.1080/07038992.2014.987376

Baret, F. – Guyot, G. – Major, D. J. (1989): Crop biomass evaluation using radiometric measurements. Photogrammetria. 43(5): 241-256, DOI: 10.1016/0031-8663(89)90001-x

Bertalan, L. ; Túri, Z. – Szabó, G. (2016): UAS photogrammetry and object-based image analysis (GEOBIA): erosion monitoring at the Kazár badland, Hungary. Acta Geographica Debrecina Landscape end Environment 10(3-4):169-178., DOI: 10.21120/LE/10/3-4/10

Blaschke, T. (2010): Object based image analysis for remote sensing. Journal of Photogrammetry and Remote Sensing, 65(1):2-16., DOI: 10.1016/j.isprsjprs.2009.06.004

Bolanio, K. – Makinano-Santillan, M. – Santillan, J. – Daguil, R.C. (2015): Using Normalized Difference Vegetation Index (NDVI) to Assess Vegetation Cover Change in Mining Areas of Tubay, Agusan del Norte. Conference: 36th Asian Conference on Remote Sensing, Quezon City, Metro Manila, Philippines

Claverie, M. –Ju, J. – Masek, J. G. – Dungan, J. L. – Vermote, E. F. – Roger, J. – Skakun, S. V. – Justice, C.(2018): The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote

Sensing of Environment, 219:145-161., DOI: 10.1016/j.rse.2018.09.002

Deák, B., Hüse, B., Tóthmérész, B. (2016): Grassland vegetation in urban habitats – Testing ecological theories. Tuexenia 36: 379-393., DOI: 10.14471/2016.36.017

Defries, R.S. –Townshend, J. R. G.(1994): NDVI-derived land cover classifications at a global scale. International Journal of Remote Sensing 15(17):3567-3586., DOI:10.1080/01431169408954345

Feranec, J. – Soukup, T. – Hazeu, G. et al. (2016): European Landscape Dynamics: CORINE Land Cover Data. CRC Press.

Gan, W. – Shen, H. – Zhang, G. L. – Gong, W. (2014): Normalization of NDVI from Different Sensor System using MODIS Products as Reference. IOP Conf. Series: Earth and Environmental Science 17, DOI:10.1088/1755-1315/17/1/012225

Gulácsi, A. – Kovács, F. (2018): Drought monitoring of forest vegetation using MODIS-based normalized difference drought index in Hungary. Hungarian Geographical Bulletin. 67(1):29-42., DOI: 10.15201/hungeobull.67.1.3

Gulácsi, A – Kovács, F (2015): Drought monitoring with spectral indices calculated from MODIS satellite images in Hungary. Journal of Environmental Geography 8(3-4):11-20, DOI: 10.1515/jengeo-2015-0008

Guyot, G. – Gu, X.F. (1994): Effect of radiometric corrections on NDVI-determined from SPOTHRV and Landsat-TM data. Remote Sensing of Environment, 49(3):169-180., DOI: 10.1016/0034-4257(94)90012-4

Hüse, B. – Szabó, Sz. – Deák, B. – Tóthmérész, B. (2016): Mapping ecological network of green habitat patches and their role in maintaining urban biodiversity in and around Debrecen city (Eastern Hungary). Land Use Policy 57: 574-581.,

DOI: 10.1016/j.landusepol.2016.06.026

Jia, K. – Liang, S. – Wei, X. – Yao, Y. – Su, Y. – Jiang, B. – Wang, X. (2014): Land Cover Classification of Landsat Data with Phenological Features Extracted from Time Series MODIS NDVI Data. Remote Sensing, 6:11518-11532.,DOI: 10.3390/rs61111518

Kinthada, N. R. – Gurram, M.K. – Eadara, A. – Velagala, V.R. (2014): Land Use/Land Cover and NDVI Analysis for Monitoring the Health of Micro-watersheds of Sarada River Basin, Visakhapatnam District, India. J Geol Geosci 3:146. DOI:10.4172/2329-6755.1000146

Kovács, F. (2018): NDVI/EVI monitoring in forest areas to assessment the climate change effects in Hungarian Great Plain from 2000. In (Eds) Neale, C.M.U – Maltese, A.(2018): Remote Sensing for Agriculture, Ecosystems, and Hydrology XX, p. 56., ISBN: 9781510621497

Lunetta, R.S. – Knight, J.F. – Ediriwickrema, J. – Lyon, J. G. – Worthy, L. D. (2006): Land-cover change detection using multi-temporal MODIS NDVI data. Remote Sensing of Environment 105(2):142-154., DOI: 10.1016/j.rse.2006.06.018

Nath, B. – Acharjee, S. (2013): Forest Cover Change Detection using Normalized Difference Vegetation Index (NDVI): A Study of Reingkhyongkine Lake's Adjoining Areas, Rangamati, Bangladesh. Indian Cartographer, 33:348-353.

Macfarlane, C. – Grigg, A.H. – Daws, M.I. (2017): A standardised Landsat time series (1973–2016) of forest leaf area index using pseudoinvariant features and spectral vegetation index isolines and a catchment hydrology application. Remote Sensing Applications: Society and Environment. 6:1-14., DOI: 10.1016/j.rsase.2017.01.006

Meeragandhi, G – Parthiban, S. – Thummalu, N. – Christy, A. (2015): NDVI: Vegetation change detection using remote sensing and gis – A case study of Vellore District. Procedia Computer Science 57:1199 – 1210., DOI: 10.1016/j.procs.2015.07.415

Mingquan Wu, Zheng Niu, Changyao Wang, Chaoyang Wu, Li Wang (2012): Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion Model. Journal of Applied Remote Sensing 6(1):063507, DOI: 10.1117/1.JRS.6.063507

Pontius Jr., R.G. – Millones, M. (2011): Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing 32(15), DOI: 10.1080/01431161.2011.552923

Rodriguez, J. E. – Alonso, C. – Tarquis, A. M. – Benito, R. M. –Hernandez Díaz-Ambrona, C. (2013): Comparison of NDVI fields obtained from different remote sensors. Geophysical Research Abstracts. Vol. 15, EGU2013-14153, EGU General Assembly 2013

Sahebjalal, E. – Dashtekian, K. (2013): Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods. 8(37):4614-4622. DOI:10.5897/AJAR11.1825

Szabó, S. – László, E. – Kovács, Z. – Püspöki, Z. – Kertész, Á. – Singh, S. K. – Balázs, B. (2019): NDVI dynamics as reflected in climatic variables: spatial and temporal trends – a case study of Hungary. Giscience and Remote Sensing 57: Article In Press pp. 1-21.

Szabó, S. – Gácsi, Z. – Balázs, B. (2016). Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories. Acta Geographica Debrecina Landscape and Environment 10 (3-4): 194-202., DOI: 10.21120/LE/10/3-4/13

Túri, Z. K. (2015): A tájszerkezet-kutatás geoinformatikai módszereinek elemzése alföldi mintaterületeken. Debrecen, Magyarország: Debreceni Egyetemi Kiadó (2015) , 168 p., ISBN: 9789633185131

Varga, O. G. – Szabó, S. – Túri, Z. (2014): Efficiency assessments of GEOBIA in land cover analysis, NE Hungary. Bulletin of Environmental and Scientific Research. 3(4):1-9.

Varga, O. – Pontius, R.G. – Singh, S.K. – Szabó, S. 2019. Intensity Analysis and the Figure of Merit's Components for assessment of a Cellular Automata - Markov simulation model. Ecological Indicators, Article in Press