ANALYZING THE SPRAWL OF BUILT-UP AREAS USING LANDSAT DATA AND THE NORMALIZED DIFFERENCE BUILT-UP INDEX

Iosif VOROVENCII, Prof. PhD. Eng., Transilvania University of Brașov, Romania, iosif.vorovencii@unitbv.ro

Abstract: *The sprawl of built-up areas is one of the most important anthropic changes which lead to the development of cities. This study analyses the sprawl of built-up areas around the city of Brașov, Romania, on the basis of two Landsat satellite images acquired in 1985 and 2019. The analysis was based on satellite images obtained through the difference between the NDBI and NDVI indices calculated for each satellite image. The results show that in the period 1985–2019 built-up areas around Brașov and its surroundings increased considerably (24.6%). Furthermore, it was noted that, although the reflectance of built-up areas varies very little from one season to the next, it can be influenced indirectly by the presence in the studied area of agricultural land and barren land whose reflectance is seasonal. This phenomenon led to confusion among built-up, agricultural land and barren land in mapping built-up areas.*

Keywords: *Landsat, NDBI, NDVI, built-up area, segmentation.*

1. Introduction

One of the main problems in mapping urban areas in order to evaluate changes is represented by the conversion of non-residential land to residential [1]. The changes concerning land use are usually caused by strong urbanization during certain intervals. Urbanization leads to the development of cities, but it can also result in temperature increases because of impervious areas, in the reduction of green areas, in the decrease of water quality [4].

During the last decades, fewer and fewer satellite images were used to solve problems related to urban sprawl [7]. In this respect, for mapping urban areas, several techniques were developed, applied and evaluated based on [6]: (1) the classification of input data, including pixel- and object-based classifications [5, 2] and (2) direct segmentation of the indices [10, 11]. Irrespective of the implemented technique, the spectral and spatial variability of the urban environment recorded on remote sensing images raises serious challenges in obtaining accurate results [8].

For the automatic mapping of built-up areas, specialized literature [10] suggested using the Normalized Difference Built-up Index (NDBI) and the Normalized Difference Vegetation Index (NDVI). The indices obtained from the Landsat Thematic Mapper (TM) images were recoded in order to create binary images. The approach suggested by Zha et al. (2003), namely to recode the NDBI and NDVI images, is based on the fact that the positive values of NDBI can indicate built-up areas, while the positive values of NDVI can show the presence of vegetation. The results obtained by Zha et al. (2003) showed that this method cannot separate clearly urban areas from barren and bare land. They suggested that the method in question should also be tested in other geographic areas because of the complicated models for the spectral answer of vegetation.

He et al. (2010) put forth an improvement of this method through the semi-automatic segmentation of the images obtained as the difference between NDBI and NDVI. This approach requires no recoding of the NDVI and NDBI images in order to obtain binary images, but it involves finding a threshold value to discriminate between built-up areas and non-built-up areas according to the image obtained as the difference between the two indices.

The main objective of this study is to map the sprawl of built-up areas in the city of Brașov, Romania, and its surroundings using two satellite images, namely Landsat 5 TM and Landsat 8 Operational Land Imager (OLI), acquired in 1985 and 2019. The study is based on the NDBI and NDVI indices and on the difference between them in determining a threshold value to discriminate between built-up areas and non-built-up areas.

2. Materials and Methods

2.1. Study area

The surface under analysis covers 11400 ha and includes the city of Brașov and its surroundings, which experienced strong economic development in the period 1990–2019 leading to the expansion of the built-up areas around the city (Figure 1). During this period new commercial, industrial, and residential areas were constructed, or the existing ones were extended, new roads were built, as well as the city ring road, the parking facilities were developed, etc.

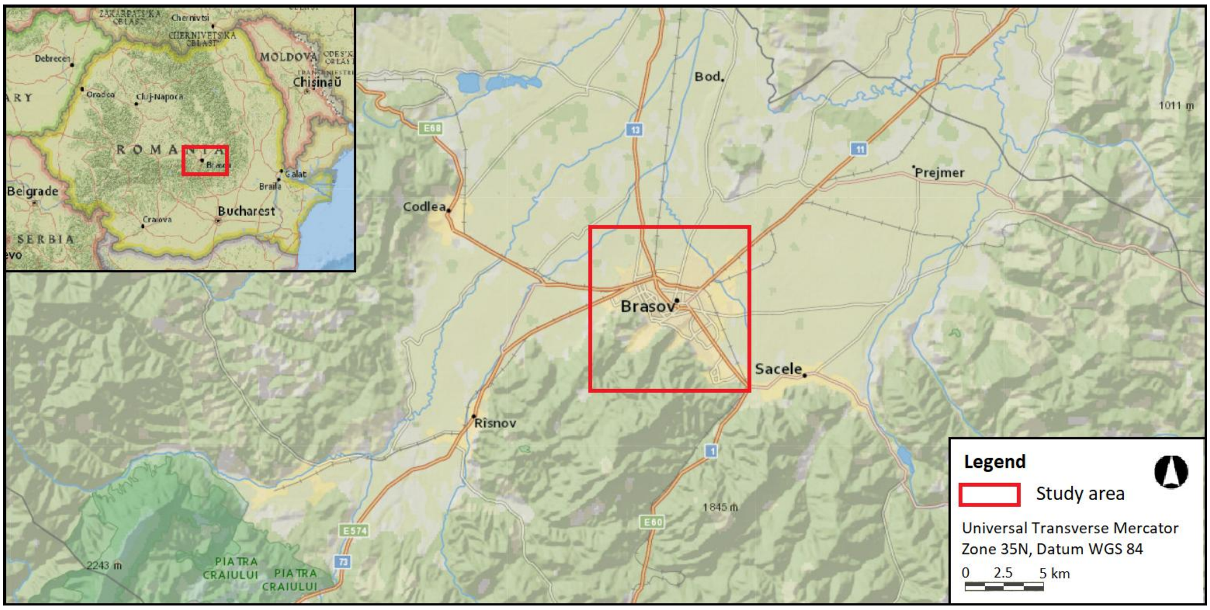

Fig. 1. Location of the studied area

The studied area encompasses built-up areas, cultivated and uncultivated agricultural land, pastures and hayfields, barren land and waters, spread to the east, north and west of Brașov. To the south there are coniferous and deciduous forests. All these land cover classes were affected by changes during the analysed period.

The studied area is characterized by average annual temperatures of 6°–8°C (7.8°C in Brașov), with an average temperature in July between 17° and 18°C. The area is characterized by thermal inversions. The average annual precipitation quantity varies quite broadly from around 600–800 mm to 1100–1200 mm. The altitude of the studied area ranges between 530 m and 1020 m, from flat agricultural land, situated around the city of Brașov, up to Poiana Brașov mountain resort.

2.2. Materials

The analysis used two satellite images, one Landsat 5 TM image acquired on 23.08.1985 and one Landsat 8 OLI image acquired on 22.09.2019, both downloaded free of charge at https://earthexplorer.usgs.gov/. The images have a spatial resolution of 30 m in multispectral and a radiometric resolution of 8 bits (Landsat 5 TM) and 12 bits (Landsat 8 OLI). The study used the spectral bands 3, 4 and 5 (Landsat 5 TM) and the bands 4, 5 and 6 (Landsat 8 OLI). The images were geo-referenced in Universal Transverse Mercator (UTM), zone 35N, datum WGS 84, and have a L1 level of correction. In order to validate the results we used the cadastral maps drafted before 1989 (1972–1986) and the orthophotos drafted in 2019.

2.3. Method

The analysis is based on the comparison between the spectral behaviour of built-up areas and that of the other land cover classes recorded on the Landsat images. Starting from this spectral behaviour we calculated the NDBI and NDVI indices and from the difference between them we obtained an image which was segmented semi-automatically by identifying built-up/non-built-up areas [6]. The NDVI and NDBI indices were calculated based on the acquired bands in red, near-infrared and middle-infrared, previously corrected atmospherically and radiometrically. The equations used are presented in Table 1. The images were processed using QGIS software.

Table 1. The equations used to calculate the NDVI and NDBI indices, and the difference between them

Index name	Acronym	Equation	Source
Normalized Difference Built-up Index	NDBI	$NDBI = \frac{R_{MIR} - R_{NIR}}{R_{MIR} + R_{NIR}}$	[9]
Normalized Difference Vegetation Index	NDVI	$NDVI = \frac{R_{NIR} - R_{RED}}{R_{NIR} + R_{RED}}$	[10]
Index difference	BU	*BU = NDBI - NDVI*	[6]
R_{RED} - the visible red radiation; R_{NIR} - the near-infrared radiation; R_{RED} - the middle-infrared radiation			

After the NDBI and NDVI indices were determined, we calculated the difference between these indices and obtained a new image (BU), which contains both non-built-up areas and built-up areas. Within this image we identified an optimal threshold value for its segmentation in a binary image in which to code the built-up areas with the value 1 and the non-built-up areas with the value 0. In order to determine an optimal threshold value we applied the method used by He et al. (2010). The basic idea of the semiautomatic approach is to select a threshold value in the training samples, starting from the fact that, if the threshold value leads to obtaining maximum accuracy in extracting the built-up area from the training samples, then it will lead to obtaining high accuracy at the level of the BU image as well. Basically, the process involves four steps.

The first step consisted in choosing, based on the visual interpretation of the image, certain typical built-up areas as training areas. The criteria for the selection of the sample areas were [6]: (1) the pixels included in the training samples represent only built-up areas and (2) the training samples represent islands surrounded by pixels representing non-built-up areas.

The second step involved setting a search range and pace by analysing the BU image histogram. The search range was set as the difference between the minimum value (a) and the maximum value (b) in the BU image. The first search interval P_1 was calculated based on the formula [6]:

$$P_1 = \frac{b-a}{m} \quad (1)$$

where *m* is a positive integer number which determines the number of potential thresholds in the search process and can be set manually. The level of the manually set *m* does not affect the efficiency of the search and the final results. A high *m* value increases the number of potential thresholds in a search but reduces the number of searches.

The third step consisted in defining the success rate related to the extraction of built-up in order to evaluate the performance of each value of the potential threshold during the search process meant to identify the pixels which represent built-up/non-built-up. The success rate L_k for a potential value *k* of the threshold is calculated as follows [6]:

$$L_k = \frac{(A_{k1}-A_{k2})}{A} \times 100\% \quad (2)$$

where A_{k1} is the number of built-up pixels detected inside all the training patches, A_{k2} is the number of built-up pixels detected incorrectly, and *A* is the number of pixels within all the patches. After calculating all the success rates L_k for a search, for all the threshold values *m*, we can obtain the maximum and minimum L_k values and attributed as L_{max} and L_{min} during the search process. If the two parameters fail to meet the input conditions presented in the four steps, a new search begins. The interval for a new search can be found in the interval ($k_{max} - P_1$, $k_{max}+P_1$) and a new small search pace is set based on the modification of the search interval with the help of equation (1). In the above mentioned interval, k_{max} represents the value of the potential threshold corresponding to L_{max} in the search process [6].

The fourth step consisted in repeating steps two and three, this being an iterative process. This iterative process is over when the following condition is fulfilled:

$$L_{max} - L_{min} \leq \delta \quad (3)$$

where L_{max} and L_{min} are the maximum and minimum values of the success rate in one of the search processes while δ is the constantly accepted error. The threshold corresponding to L_{max} is considered the optimal threshold to extract the built-up areas of the BU image [6].

Before implementing the four steps to determine the optimal threshold values, the NDBI, NDVI and BU values calculated for each image were normalized at the intervals (0–255) and (0–4095) respectively, using the following equations:

$$\text{for Landsat TM: } N = \frac{(V-V_{min})}{(V_{max}-V_{min})} \times 255 \quad (4)$$

$$\text{for Landsat OLI: } N = \frac{(V-V_{min})}{(V_{max}-V_{min})} \times 4095 \quad (5)$$

where N is the value of the pixel in the normalized image, V is the value of the pixel in the non-normalized image, while V_{max} and V_{min} are the maximum and the minimum values in the non-normalized image.

After identifying the optimal threshold, the BU images were transformed into binary images, the value 0 was attributed to non-built-up areas, and the value 1 to built-up areas. The accuracy assessment was achieved through the stratified random sampling method using 50 samples for each built-up and non-built-up area. On the basis of the values for the built-up areas estimated for each satellite image we calculated the built-up area for the studied period.

3. Results and discussion

As a result of applying the equations in Table 1, we obtained NDBI, NDVI and BU continuous images (Figure 2). Based on the analysis of the BU histogram for the Landsat 5 TM image from 1985, the search range was set to (0, 255) with an initial threshold of 25. The semi-automatic segmentation led to an optimal threshold value of 150 and the pixels with values over 150 were extracted from the BU image as representing built-up areas. For the Landsat 8 OLI image from 2019, the search range was set to (128, 4095), with an initial threshold of 50. The optimal threshold value obtained after the semi-automatic segmentation was 2400 which implied that the pixels with higher values were extracted from the BU image as representing built-up areas. The images with built-up/non-built-up areas are presented in Figure 3.

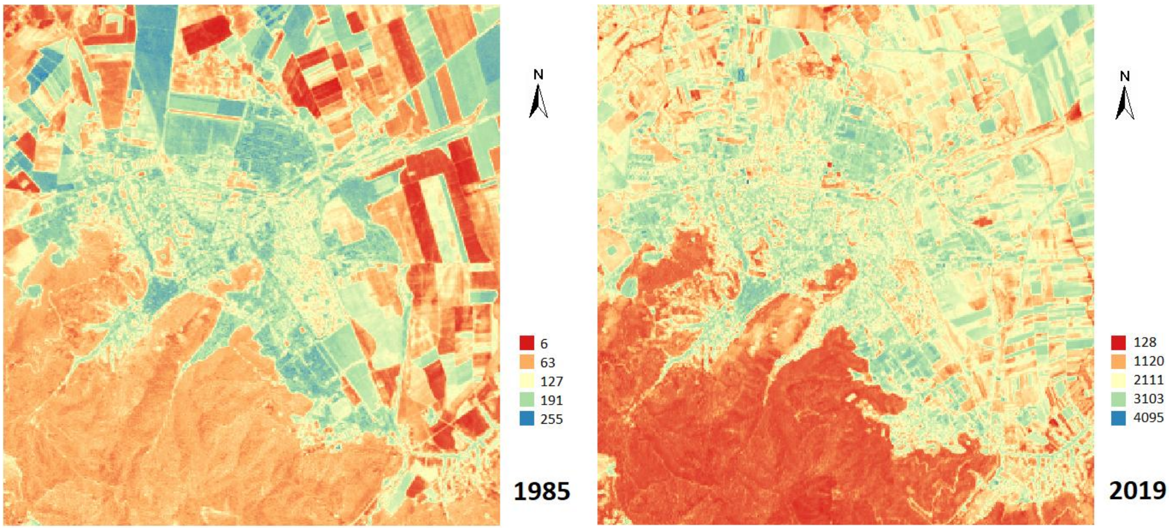

Fig. 2. The BU normalized images obtained after calculating the difference between NDBI and NDVI (also normalized)

The overall mapping accuracy of the image from 1985 was 81.25%, while for the image from 2019 it was 85.56%. For both images, several commission errors were noticed in the agricultural areas which remained uncultivated, but they were ploughed, situated to the north, north-east and north-west of the city of Brașov. On both images the north-west part was mapped as built-up area, although it consists only of agricultural land. It is possible that, when the two images were acquired, the land was uncultivated or, if it was cultivated, the cultures were very scarce. The poorer accuracy was identified for the image from 1985 even in built-up areas, in the place of the former Tractorul factory and in some locations in the old city of Brașov. Also, on the image from 1985 we identified locations (the neighborhoods of Noua, Tractorul) where the agricultural land was mapped as built-up. In the southern part of the studied area, covered by a matured and dense forest, the mapping accuracy was high. The overall accuracy can be considered good because it is close to the overall accuracy obtained in other studies (86.30%) [6].

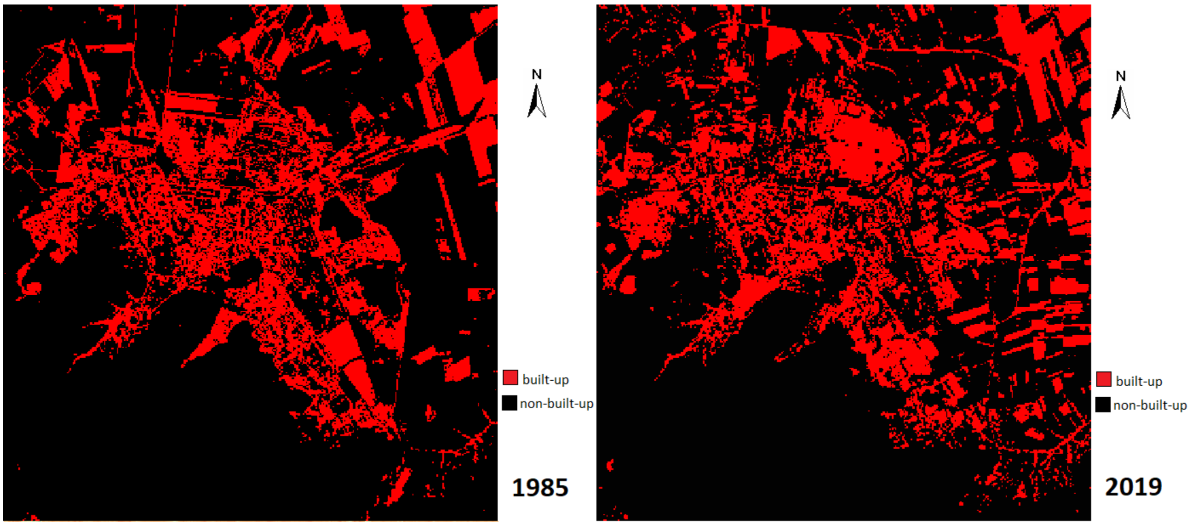

Fig. 3. The images obtained after the semi-automatic segmentation, which reflect the situation of the built-up and non-built-up areas for 1985 and 2019

The reflectance of the urban areas is very little influenced by seasonal variations. However, it can be indirectly affected by the presence of other vegetal covers whose reflectance is seasonal, such as forests or uncultivated or cultivated agricultural land which is in various developmental stages. The agricultural land around the city of Brașov is very fragmented, with small areas and with different owners. In this context, each owner cultivated the land with various cultures, at different times of the year. There is also uncultivated land which became pastures with different degrees of use and density of the vegetal cover. In these conditions, in the north of the city there is a very diverse mosaic of agricultural land with a spectral behaviour similar to built-up areas, and which caused confusion during mapping. In the case of the forest in the south of the city there were no problems in discriminating between built-up and forest. Thus, it is difficult to find a satellite image in which the agricultural land around the city is cultivated with plants which are at the same growth stage, and the vegetation is fully developed.

The best mapping results were obtained for urban areas, covering larger surfaces and without alternating with green areas. Such areas were few because of the nature of the city of Brașov which does not have a very dense street structure, it has very few large avenues, and in many palaces, the constructions are private houses with courtyards or gardens. This fringe structure, where constructions alternate with green areas or courtyards, in many locations led to commission errors. At the same time, we cannot discriminate between industrial, commercial and residential areas.

Although the specialized literature [10] recommends testing the method in other geographic areas as well, in this case the results are close to the ones obtained in other studies [6]. At any rate, the success of the method depends on the NDVI values of the vegetation around the built-up areas which must be much higher than zero. However, the spectral response of the vegetation can be different from one location to the next because of the species of plants and trees, the nature of the soil, the quantity of water in the soil, the vegetation conditions, the density of the vegetation and its health. In this context, the value of the NDVI did not exceed zero by much in the studied area around the city of Brașov, which caused the confusion between these areas and the built-up areas.

The results obtained based on the satellite images show that in the period 1985–2019, the increase of the built-up area was estimated at 642 ha (24,6%). The built-up areas were also measured manually, through vectorization, on the false colour composite images. The result 665 ha obtained manually is close to 642 ha, result estimated based on the applied method. The difference between the two sets of results is of 23 ha (3.5%). This difference can be explained by the fact that, by applying the semiautomatic segmentation method, within the city, certain pixels were identified as non-built-up area, whereas by applying the vectorization method the same pixels were included in the built-up area. Furthermore, we also mapped as built-up the pixels corresponding to the calcar queries close to the city and to different constructions isolated from the periphery of the city. The value obtained is in line with other studies concerned with urban sprawl in Brașov, which estimated, for the period 1981–2011, a urban sprawl of 545,66 ha (18.33%) [3]. The phenomenon of urban sprawl is specific to Romania after 1989 when the country passed from socialist property to private property. During this period both the private and the public sector developed, by expanding the existing residential areas or by constructing new ones, building new road, including the ring road of the city of Brașov, new industrial and commercial constructions around the city, etc. The sprawl of built-up area is part of the same increasing trend also noticed in other Romanian cities.

4. Conclusions

In this study we used two Landsat satellite images to establish the sprawl of the built-up areas in the period 1985–2019 for the city of Brașov and its surroundings. Although we used the improved version [6] of the original method [10], we noticed that this method has its limitations too. The results show that these limitations can be caused by the types of land cover which exist around the built-up areas or by the semiautomatic approach, which depends on the manual selection of the training samples. As for the first group of limitations, the problem can be overcome if we choose the acquired images when the defoliation of the vegetation around the built-up areas is minimum or inexistent. The mixture of built-up and barren farmland can be highlighted on the satellite images by using images acquired when the vegetal cover is at the maximum. As for the second group of limitations, we noticed that by using NDBI we cannot separate clearly the built-up areas from the agricultural land and the barren land because both types of land cover have the same spectral behaviour in the Landsat bands. In this context, the accuracy of the method is predictably low in mapping urban peripheral areas, where there are arid or agricultural areas, especially uncultivated ones.

Despite all these limitations, the NDBI method has also advantages in the sense that it does not require a complex mathematical formula and it can be used as a worthwhile alternative, efficient and advantageous, to quickly map built-up areas. The most important aspect in obtaining good results is to choose an adequate threshold to best discriminate between built-up areas and non-built-up areas.

5. References

1. *As-syakur, A. R., Adnyana, W. S., Arthana, W., Nuarsa, W. – Enhanced built-up and bareness index (EBBI) for mapping built-up and bare land in an urban area. Remote Sensing, 4, p. 2957-2970, 2012.*

2. *Cleve, C., Kelly, M., Kearns, F., Moritz, M. – Classification of the wildland–urban interface: a comparison of pixel- and object-based classifications using high-resolution aerial photography. Computers, Environment and Urban Systems, 32, p. 317–326, 2008.*
3. *Cocheci, M. – Environmental impact assessment of urban sprawl in the Braşov metropolitan area. Urbanism. Arhitectură. Construcţii, 5(2), 21-37, 2014.*
4. *Greşiţă, C. I. - Expert system used for monitoring the behavior of hydrotechnical constructions. REVCAD-Journal of Geodesy and Cadastre, 11, p. 75–84, 2011.*
5. *Guindon, B., Zhang, Y., Dillabaugh, C. – Landsat urban mapping based on a combined spectral–spatial methodology. Remote Sensing of Environment, 92, p. 218–232, 2004.*
6. *He, C., Shi, P., Xie, D., Zhao, Y. – Improving the normalized difference built-up index to map urban built-upareas using a semiautomatic segmentation approach. Remote Sensing Letters, 1(4), p. 213–221, 2010.*
7. *Jacquin, A., Misaova, L., Gay, M. – A hybrid object-based classification approach for mapping urban sprawl in periurban environment. Landscape and Urban Planning, 84, p. 152–165, 2008.*
8. *Powell, R., Roberts, D., Dennison, P., Hess, L. – Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. Remote Sensing of Environment, 106, p. 253–267, 2007.*
9. *Rouse, J. W. Jr., Hoas, R. H., Deering, D. W., Schell, J. A., Harlon, J. C. – Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation; NASA/GSFC Type III Final report, Greenbelt, MD, p. 371, 1974.*
10. *Zha, Y., Gao, J., Ni, S. – Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. International Journal of Remote Sensing, 24(3), 583-594, 2003.*
11. *Zhang, Q., Pavlic, G., Chen, W., Fraser, R., Leblanc, S., Cihlar, J. – A semiautomatic segmentation procedure for feature extraction in remotely sensed imagery. Computers and Geosciences, 31, p. 289–296, 2005.*