EXTRACTION OF BUILT-UP AREA USING HIGH RESOLUTION SENTINEL-2A AND GOOGLE SATELLITE IMAGERY

S.Vigneshwaran, S.Vasantha Kumar *

School of Civil and Chemical Engineering, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India - svigneshwarana@outlook.in and svasanthakumar1982@gmail.com

KEY WORDS: Built-up area, Extraction, Normalized Difference Index, High Resolution Satellite Imagery, Sentinel-2A, Google Satellite Imagery

ABSTRACT:

Accurate information about the built-up area in a city or town is essential for urban planners for proper planning of urban infrastructure facilities and other basic amenities. The normalized difference indices available in literature for built-up area extraction are mostly based on moderate resolution images such as Landsat Thematic Mapper (TM) and enhanced TM (ETM+) and may not be applicable for high resolution images such as Sentinel-2A. In the present study, an attempt has been made to extract the built-up area from Sentinel-2A satellite data of Chennai, India using normalized difference index (NDI) with different band combinations. It was found that the built-up area was clearly distinguishable when the index value ranges between -0.29 and -0.09 in blue and near-infrared (NIR) band combination. Post extraction editing using Google satellite imagery was also attempted to improve the extraction results. The results showed an overall accuracy of 90% and Kappa value of 0.785. Same approach when applied for another area also yields good results with overall accuracy of 92% and Kappa value of 0.83. As the proposed approach is simple to understand, yields accurate results and requires only open source data, the same can be used for extracting the built-up area using Sentinel-2A and Google satellite imagery.

1. INTRODUCTION

Urbanization is a major concern in most of the metropolitan cities in the world as the cities are expanding in an unprecedented pace due to migration of people from rural to urban areas for better job opportunities and living conditions. In 1950, only 30 per cent of the world's population lived in cities but in 2050, it is projected that 66 per cent of the world's population will live in cities (Spence et al., 2013). This twofold rise may lead to many urban related problems. In Tamil Nadu state of India, 50% of the total population was living in major cities of the state, which is one of the highest in the country (State Planning Commission, 2012). This shows that most of the cities in the state are experiencing rapid urbanization in recent decades. Urbanization is unavoidable in a developing country like India. However if there is no proper control, it may lead to loss in agricultural land and productivity, cutting down of tress, crowded habitats, water distribution and sewage treatment problems, air pollution, traffic congestion, etc. Accurate information on urban built-up area is essential for preparation of master plan and detailed development plans, provision of basic amenities and urban infrastructure facilities, identifying suitable location for solid waste disposal, planning for smart cities, satellite towns, etc. In recent decades, the use of satellite data has replaced the traditional field survey methods for identification of the urban built-up land due to advancements in remote sensing technology and geographic information systems (GIS).

In general, the techniques for extraction of built-up area from satellite imagery can be grouped into two types: (1) techniques based on conventional multi-spectral image classification such as supervised, unsupervised, object based or deep learning classification (Ndehedehe et al., 2013; Zhang et al., 2014;

Rawat and Kumar, 2015; Hegazy and Kaloop, 2015; Forget et al., 2017; Bramhe et al., 2018) (2) techniques based on normalized difference indices such as normalized difference built-up index (NDBI), principal component analysis (PCA) based built-up area index (PCABI), enhanced built-up index (EBI), urban index (UI), etc. (Zha et al., 2003; Xu, 2007; He et al., 2010; Kumar et al., 2012; Varshney, 2013; Bhatti and Tripathi, 2014; Shahi et al., 2015; Bouzekri et al., 2015; Vecchi et al., 2015; Garg et al., 2016; Kapil and Pal, 2017; Zhang et al., 2017; Hidayati et al., 2018; Valdiviezo et al., 2018; Li et al., 2018).

According to Xu (2007), classification based methods generally may not yield satisfactory accuracy, usually accuracy less than 80%, due to spectral confusion of the heterogeneous urban built-up land class when compared to other land use classes. Alternate to image classification, many researchers were trying out normalized difference indices utilizing specific spectral bands for automatic extraction of built-up land from satellite imagery. Zha et al. (2003) proposed a spectral index called normalized difference built-up index (NDBI) for automatic extraction of urban built-up land utilizing short wave infrared (SWIR) and near infrared (NIR) bands of TM imagery and reported an overall accuracy of 92.6%. Xu (2007) used three normalized difference indices, namely NDBI, modified normalized difference water index (MNDWI), and soil adjusted vegetation index (SAVI) to derive urban built-up areas using TM and ETM+ imagery. The study resulted in an overall accuracy ranging between 91.5 and 98.5%. It is to be noted here that in most of the studies, only TM and ETM+ imagery of moderate resolution were employed utilizing NIR and SWIR bands. Even though the indices perform well as seen in the above review, it may be difficult to apply NDBI based on SWIR and NIR bands directly to a high resolution imagery such as

---

* Corresponding author

Sentinel-2A. Because in Sentinel-2A, SWIR band has 20m. resolution whereas Blue (B2), Green (B3), Red (B4) and NIR (B8) has 10m. resolution. For urban related applications, the bands which are at 10m. resolution is generally preferred as the spatial resolution is high. However it is not known which band combination among the four 10m. resolution bands in Sentinel-2A will yield better results in extracting the built-up area. The objective of the present study is to identify the suitable band combination and the index range where the built-up area is clearly distinguishable in Sentinel-2A data. Once the suitable band combination and the index range are identified, the built-up area can then be easily extracted. The following section describes the materials and the methods used and in section 3, results were discussed followed by the concluding remarks in section 4.

2. MATERIALS AND METHODS USED

2.1 Study Area

The study area selected for the present work is located in Chennai, India. Chennai (formerly called as Madras) is the capital of the Tamil Nadu state and one of the largest metropolitan cities in the country. Figure 1 shows the areal extent of Chennai Corporation and its location in Tamil Nadu and India. The Chennai Corporation is headed by a Mayor and it comprises 15 zones and 200 wards. Until 2011, Chennai Corporation covered only an area of 176 sq.km. After that, several municipalities, town panchayats and panchayat unions adjacent to the city which were expanded along with the city were merged with the corporation and hence its areal extent has been increased from 176 sq.km to 430 sq.km. The Sholinganallur zone located in Southern part of Chennai city (zone number 15 in Figure 1) is taken as the study area to evaluate the performance of the proposed approach for extracting the built-up area using Sentinel-2A and Google satellite imagery. The areal extent of Sholinganallur zone is 41.525 sq.km and it contains many Information Technology (IT) companies and residential areas. The reason for selecting this particular zone in Chennai corporation is that it has a mix of different land covers like built-up area, open land, marsh land, water bodies, etc.
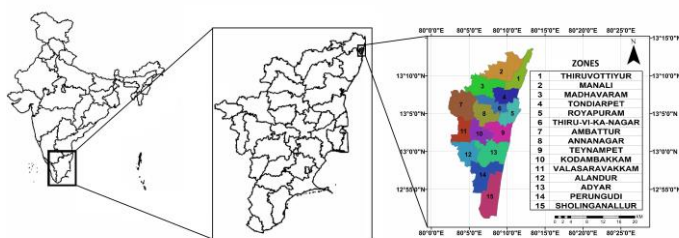

Figure 1. Map showing the location of Tamil Nadu state in India (left), location of Chennai Corporation in Tamil Nadu (middle) and Chennai Corporation (right).

2.2 Details of Satellite Data used

The Sentinel-2A satellite data acquired on August 24, 2016 was used in the present study for extraction of built-up area. The Sentinel-2A data were acquired, processed, and generated by the European Space Agency and repackaged by U.S. Geological Survey (USGS) into 100 km × 100 km tiles. Hence the satellite data was downloaded from USGS Earth Explorer (Sentinel-2A, 2018). As the revisit frequency is 5 days, large number of images acquired at different dates was available for download.

However, selecting an image without any cloud cover was difficult as percentage of cloud cover was high in most of the images. The satellite image taken on August 24, 2016 had a cloud cover of only 0.41% and hence the same was considered in the present study. After downloading Sentinel-2A data, clipping of the image within the study area was carried out as the downloaded image is of 100 km × 100 km size. The processing level of the downloaded image was Level-1C which includes radiometric and geometric correction, ortho rectification and spatial registration on a global reference system, namely, the WGS84 datum and Universal Transverse Mercator (UTM) Projection. There are 13 spectral bands in Sentinel-2A which extends from visible and near-infrared (VNIR) portion to Short-wave infrared (SWIR) portion. Out of 13 bands, four bands have 10m spatial resolution, six bands are at 20m resolution and the remaining three bands at 60m resolution. An image with high spatial resolution is generally preferred for urban related applications. Hence in the present study, only four bands which are at 10m. spatial resolution were considered. They are blue (Band 2 at 490 nm), green (Band 3 at 560 nm), red (Band 4 at 665 nm) and near-infrared (NIR) (Band 8 at 842 nm). The methodology to extract the built-up area using these four bands is explained in the following section.

2.3 Methodology

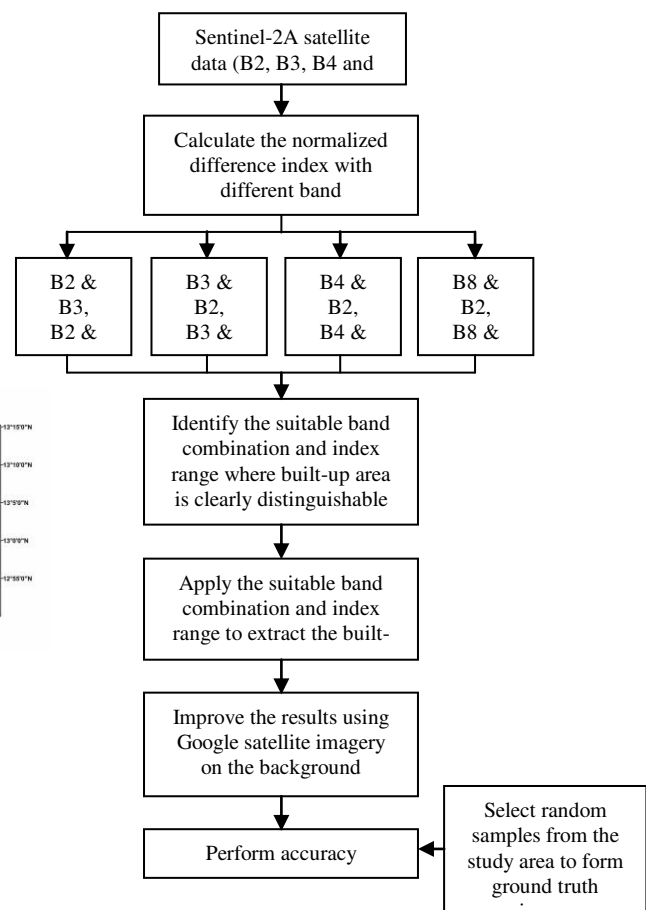The flowchart showing the proposed methodology is presented in Figure 2.

Figure 2. Flowchart showing the proposed methodology

The four bands of high resolution Sentinel-2A covering the study area is the input satellite data used in the present study.

The extraction of built-up area is based on the concept of normalized difference index which was originally proposed by Rouse et al. (1973) for the identification of vegetated areas by taking the ratio of the difference of the red and infrared radiances over their sum. It is well known that healthy vegetation absorbs most of the visible light and reflects a large portion of the NIR light and this is the reason why Rouse et al. (1973) considered these two bands in calculation of NDVI. However in case of a built-up area in a city, it is still unexplored, on which spectral band out of four high resolution bands of Sentienel-2A it absorbs more light and similarly the band in which it reflects more light. Hence the first step in the present work is to calculate the normalized difference index (NDI) using Eq. (1) with different band combinations to identify the suitable bands and the index range (minimum and maximum index value), where the built-up area is clearly distinguishable when compared to other land covers.

$$NDI = \frac{\beta_x - \beta_y}{\beta_x + \beta_y}, \qquad (1)$$

Where, β_x is the brightness value from band x, $x(band\ number) = 2,3,4,8$ and β_y is the brightness value from band y, $y(band\ number) = 2,3,4,8$; $x \neq y$. The raster calculator tool in ArcGIS 10.6 software was used for calculation of the normalized difference index.

A total of 12 band combinations as shown in Figure 2 was tried out to identify the suitable bands and the index range. This has been done by visual interpretation, i.e., visually looking at each of the 12 NDI outputs to identify in which band combination the built-up area is clearly distinguishable when compared to other land cover classes. Once the suitable band combination and index range is identified, the next step is to extract the built-up area. Even though the term "built-up area" includes all manmade structures like buildings, roads and impervious surfaces, however, in the present study our focus is on extracting only buildings. As buildings and roads (bitumen or concrete) exhibit different reflectance pattern, it may not be possible to capture both with a single NDI.

The very high resolution Google satellite imagery was used in the present study to improve the results of built-up area extraction. As like Sentinel-2A, Google satellite imagery is also open source which is one of its major advantages. Other advantage of Google imagery is we can see individual buildings as the spatial resolution is very high in the order of less than or equal to 1m. The only disadvantage of Google imagery is we cannot perform any classification or automatic extraction as it does not contain the original reflectance values or digital numbers (DN). However using the "Add Basemap" tool of ArcGIS 10.6 software, it is possible to open the Google satellite imagery on the background provided the system is connected to internet. In the present study, Google imagery was opened on the background of the extraction output which contains the two land use classes, namely, "Built-up area" and "Others" for the study area chosen. Now by manually comparing the extraction output with the background Google imagery, it is possible to check whether the extraction results are correct or not. For example, if any polygon is assigned "built-up area" during extraction and if it is found from Google image that it is not built-up area, then we can change the class to "Others". Once the post extraction editing using Google satellite imagery is done, the last step of accuracy assessment was carried out using 100 sample points generated randomly using 'Create Accuracy Assessment Points' tool in ArcGIS software. For each of the random point generated, the type of land cover ("Built-up area" or "Others") was compared with that of the actual land cover identified from Google satellite image to get the overall accuracy and Kappa statistic. The results are discussed in the following section.

3. RESULTS AND DISCUSSION

The analysis of different band combinations revealed that band-2 (blue) and band-8 (NIR) are good in extracting the built-up area when compared to other band combinations. It was found that when the index value ranges between -0.29 and -0.09, the built-up areas are clearly distinguishable when compared to other land covers. The Sentinel-2A satellite image (true color composite), extracted built-up area before post extraction editing using Google satellite imagery and built-up area after post extraction editing are shown in Figure 3 (a), (b), (c) respectively. The results of accuracy assessment are shown in Table 1.
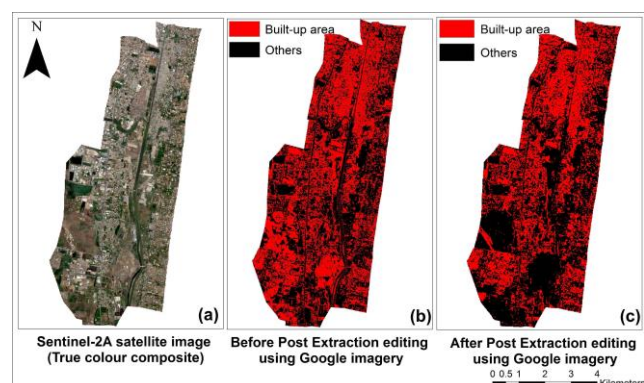

Figure 3 (a). True color composite of Sentinel-2A, (b) Before Post extraction editing using Google image, (c) After Post extraction editing using Google image

Table 1. Results of accuracy assessment
(a) Confusion matrix before post extraction editing using Google satellite imagery

Class	Built-up Area	Others	Total	User's Accuracy
Built-up Area	26	20	46	56.52%
Others	2	52	54	96.29%
Total	28	72	100	
Producer's accuracy	92.85%	72.22%	Overall accuracy = 78% Kappa coefficient = 0.54	

(b) Confusion matrix after post extraction editing using Google satellite imagery

Class	Built-up Area	Others	Total	User's Accuracy
Built-up Area	31	10	41	75.6%
Others	0	59	59	100%
Total	31	69	100	
Producer's accuracy	100%	85.5%	Overall accuracy = 90% Kappa coefficient = 0.785	

The overall accuracy before and after Google image correction was found to be 78% and 90% respectively. The reason for less

overall accuracy initially is that in south western part of the study area, there are some open lands which also had index values in the range of -0.29 to -0.09. Hence they have been classified by mistake as built-up but they are not actually the built-up lands. A zoomed view of that portion from Sentinel-2A is shown in Figure 4. It can be seen from Figure 4 that the building roofs and open land has same tone (brown colour) and hence the NDI could not differentiate between open lands and buildings. In such cases, the Google image can be used on the background to check whether the NDI output is correct or not because one can see very clearly the buildings in Google satellite image. Polygons which are wrongly assigned as "Built-up" by NDI were corrected using Google satellite image and after this post extraction editing, the accuracy has been improved from 78% to 90%. According to Congalton and Green (2009), an overall accuracy of 85% is the cut-off between acceptable and unacceptable results. In the present study, as the obtained accuracy is more than 85%, the results of built-up area extraction can be considered acceptable.

Figure 4. Similar tone for built-up area and open land in south western part of the study area

In order to check whether the proposed approach of using blue and NIR bands for built-up area extraction and further improving the extraction results using Google image, Sentinel-2A image of Adyar zone (zone 13) was taken into account and analyzed. The built-up area was extracted using blue and NIR bands and pixels in the range of -0.29 to -0.09 was classified as "Built-up area" and pixels outside this range were declared as "Others". The same band combination and index range as used before for zone 15 was applied here also for zone 13. The results are shown in Figure 5 and Table 2. The overall accuracy was found to be 92% with Kappa value of 0.83. The high accuracy and Kappa statistic shows the suitability of the proposed approach for extraction of built-up area using Sentinel-2A optical image.
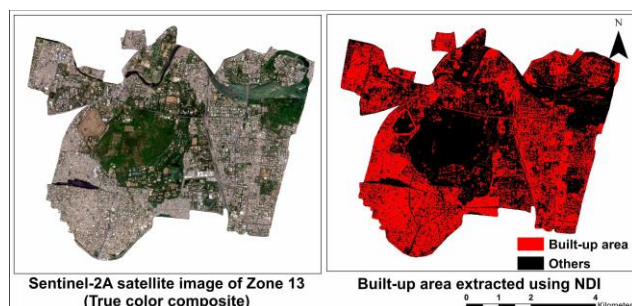

Figure 5 Sentinel-2A image of Zone 13 (left) and Built-up area extraction using NDI (right)

Table 2. Results of accuracy assessment for Zone 13

Class	Built-up Area	Others	Total	User's Accuracy
Built-up Area	39	6	45	86.66%
Others	2	53	55	96.36%
Total	41	59	100	
Producer's accuracy	95.12%	89.83%	Overall accuracy = 92% Kappa coefficient = 0.83	

4. CONCLUDING REMARKS

Information about the built-up area and its extent is essential for city planners to understand the urban growth pattern, direction of growth, urban sprawl type, etc. In recent years, researchers are paying more attention in using normalized difference indices for extraction of built-up land as it yields more accurate results when compared to traditional image classification methods. The normalized difference indices available in literature for built-up area extraction are mostly based on moderate resolution images such as TM and ETM+. Studies on use of open source high resolution imagery such as Sentinel-2A are very limited. Hence the present study focused on extracting built-up area from Sentinel-2A satellite imagery. When blue and NIR bands are used, the built-up areas are clearly distinguishable, especially when the index value ranges between -0.29 and -0.09. As the building roofs and open ground exhibits a similar tone, open ground are sometimes misclassified as built-up land. To correct this, Google satellite imagery are used on the background after the extraction process. The results are promising and the proposed approach can be used for extracting the built-up area from Sentinel-2A satellite data. The advantages of the proposed approach are simple, accurate, easy to understand and requires only open source data as input. The future scope of the work is to find suitable band combinations and index range for extracting other land covers such as vegetation, water bodies, wet land, etc. from Sentinel-2A image.

REFERENCES

Bhatti, S.S., Tripathi, N.K., 2014. Built-up area extraction using Landsat 8 OLI imagery. In: *GIScience & Remote Sensing*, 51(4), pp. 445-467.

Bouzekri, S., Lasbet, A.A., Lachehab, A., 2015. A new spectral index for extraction of built-up area using Landsat-8 data. In: *Journal of Indian Society of Remote Sensing*, 43(4), pp. 867-873.

Bramhe, V.S., Ghosh, S.K., Garg, P.K., 2018. Extraction of built-up areas using convolutional neural networks and transfer learning from Sentinel-2 satellite images. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-3, pp. 79-85.

Congalton, R.G., Green, K., 2009. *Assessing the accuracy of remotely sensed data*. 2nd Edition. CRC Press, New York.

Forget, Y., Linard, C., Gilbert, M., 2017. Automated supervised classification of Ouagadougou built-up areas in Landsat scenes using OpenStreetMap. In: *Proc. 2017 Joint Urban Remote Sensing Event (JURSE)*, Dubai, pp.1-4.

Garg, A., Pal, D., Singh, H., Pandey, D.C., 2016. A comparative study of NDBI, NDISI and NDII for extraction of urban

impervious surface of Dehradun [Uttarakhand, India] using Landsat 8 imagery. In: *2016 International Conference on Emerging Trends in Communication Technologies (ETCT)*, Uttarakhand, India, pp.1-5.

He, C., Shi, P., Xie, D., Zhao, Y., 2010. Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach. In: *Remote Sensing Letters*, 1 (4), pp. 213–221.

Hegazy, I.R., Kaloop, M.R., 2015. Monitoring urban growth and land use change detection with GIS and remote sensing techniques in Daqahlia governorate Egypt. In: *International Journal of Sustainable Built Environment*, 4, pp. 117-124.

Hidayati, I.N., Suharyadi, R., Danoedoro, P., 2018. Developing an extraction method of urban built-up area based on remote sensing imagery transformation index. In: *Forum Geografi*, 32(1), DOI: https://doi.org/10.23917/forgeo.v32i1.5907.

Kapil, Pal, M., 2017. Comparison of Landsat 8 and Sentinel 2 data for accurate mapping of built-up area and bare soil. In: *The 38th Asian Conference on Remote Sensing*, New Delhi, India, pp. 1-4.

Kumar, A., Pandey, A.C., Jeyaseelan, A.T., 2012. Built-up and vegetation extraction and density mapping using WorldView-II. In: *Geocarto International*, 27 (7), pp. 557-568.

Li, L., Zhou, H., Wen, Q., Chen, T., Guan, F., Ren, B., Yu, H., Wang, Z., 2018. Automatic extraction of urban built-up area based on object-oriented method and remote sensing data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-3, pp. 877-883.

Ndehedehe, C.E., Oludiji, S.M., Asuquo, I.M., 2013. Supervised learning methods in the mapping of built up areas from Landsat-based satellite imagery in part of Uyo Metropolis. In: *New York Science Journal*, 6(9), pp. 45-52.

Rawat, J.S., Kumar, M., 2015. Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. In: *The Egyptian Journal of Remote Sensing and Space Sciences*, 18, pp. 77-84.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. In: *3rd ERTS Symposium NASA SP-351 I*, pp. 309–317.

Sentinel-2A (European Space Agency (ESA)) image courtesy of the U.S. Geological Survey. https://earthexplorer.usgs.gov/

Shahi, K., Shafri, H.Z.M., Taherzadeh, E., Mansor, S., Muniand, R., 2015. A novel spectral index to automatically extract road networks from WorldView-2 satellite imagery. In: *The Egyptian Journal of Remote Sensing and Space Sciences*, 18, pp. 27-33.

Spence, M., Annez, P.C., Buckley, R.M., 2013. *Urbanization and growth, The International Bank for Reconstruction and Development*. The World Bank on behalf of the Commission on Growth and Development, Washington D.C., USA.

State Planning Commission, 2012. *Twelfth five year plan Tamilnadu 2012-2017*, Government of Tamil Nadu, Chennai, India.

Valdiviezo, J.C., Quiñones, A.T., Garibay, A.S., Caloca, A.L., 2018. Built-up index methods and their applications for urban extraction from Sentinel 2A satellite data: discussion. In: *Journal of the Optical Society of America A,* 35(1), pp. 35-44.

Varshney, A., 2013. Improved NDBI differencing algorithm for built-up regions change detection from remote sensing data: an automated approach. In: *Remote Sensing Letters*, 4 (5), pp. 504–512.

Vecchi, D.D., Harb, M., Dell'Acqua, F., 2015. A PCA-based hybrid approach for built-up area extraction from Landsat 5, 7 and 8 datasets. In: *International Geoscience and Remote Sensing Symposium* (IGARSS 2015), Italy, pp. 1152-1154.

Xu, H., 2007. Extraction of urban built-up land features from Landsat Imagery using a thematic-oriented index combination technique. In: *Photogrammetric Engineering & Remote Sensing*, 73(12), pp. 1381-1391.

Zha, Y., Gao, Y., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. In: *International Journal of Remote Sensing*, 24, pp. 583–594.

Zhang, J., Li, P., Wang, J., 2014. Urban built-up area extraction from Landsat TM/ETM+ images using spectral information and multivariate texture. In: *Remote sensing*, 6, pp. 7339-7359.

Zhang, P., Sun, Q., Liu, M., Li, J., Sun, D., 2017. A strategy of rapid extraction of built-up area using multi-seasonal Landsat-8 thermal infrared band 10 images. In: *Remote Sensing*, 9(11), pp. 1-18.

Revised August 2018