Earth Observation and Geomatics Engineering 3(1) (2019) 112–119

EARTH OBSERVATION AND GEOMATICS ENGINEERING
website: https://eoge.ut.ac.ir

# Estimation of the degree of surface sealing with Sentinel 2 data using building indices

**Caleb Juma Masinde, Dorothy Chepkirui Rono, Michael Hahn***

Department of Geomatics, Computer Science and Mathematics, Hochschule für Technik Stuttgart, Germany

**Article history:**
Received: 15 January 2019, Received in revised form: 5 May 2019, Accepted: 10 May 2019

## ABSTRACT

Various building indices to identify and extract sealed surfaces have been developed and implemented by various authors. Previous research has shown that building indices are easy to implement since they do not use complex algorithms and therefore can be used as quick methods for monitoring impervious surfaces. The aim of this study is to assess the ability of selected indices to identify sealed surfaces. Also, previously, authors have posted results that building indices face difficulty in distinguishing between sealed surfaces and bare lands owing to the spectral similarity between these two land covers. Additionally, it has been concluded by some researches that the performance of building indices also depends on the time of image capture i.e. dry and wet seasons. In this study, we implement 6 selected indices using sentinel 2 data covering Nürtingen city in Stuttgart, investigate and compare their performance in different times of the year. Google earth engine is used to conduct these investigations.

**KEYWORDS**

Surface sealing
Sentinel Imagery
Building Indices
Otsu thresholding
Google Earth Engine

## 1. Introduction

### 1.1. Background

Earth monitoring satellites provide volumes of imagery which have a wide variety of uses such as in environmental conservation to monitor forest cover and wildfire, in agriculture to determine crop health and harvest outputs, in defense for monitoring of hotspots and in urban and rural planning.

As at now, there are several satellite data that are made available for free to the public within very few hours or days of capture. Common ones as per (USGS, 2015), are Landsat images which have been available since 1972, MODIS, according to (NASA, 2019) since 2000 and the most recent entrant Sentinel which has been available since 2015. (ESA, 2019) indicates that Sentinel offers a high temporal resolution of up to 2 days at high latitudes and a medium spatial resolution of 10 m and 20 m for commonly used bands, 60 m for water vapor and coastal aerosol bands.

Urban areas are important centers of economic growth and social mobilization. Thus, there is a need to scientifically identify and monitor them over time. Different remote sensing techniques exist to identify urban areas from satellite imagery. One of these methods is through the calculation of indices from different imagery band combination.

An index is a ratio obtained by carrying out numerical operations on pixels belonging to different bands but containing information from the same imaged area. (NASA, 2000) states that the most popular and commonly used index is the Normalized Difference Vegetation Index (NDVI) that is used for better visualization of vegetation in an area.

Similarly, several building indices have been developed to identify urban areas from remotely sensed imagery. This study aims at evaluating different indices with a case study of Nürtingen area, Stuttgart, Germany.

In this study, google earth engine (GEE) is used to implement selected building indices to obtain index images. Otsu binarization is employed on the index images to finally obtain the sealed surfaces. These surfaces are compared to a reference data from OpenStreetMap (OSM) to ascertain the accuracy of each indexing method.

This study reports that building indices have trouble in

* Corresponding author
E-mail addresses: 82maca1mpg@hft-stuttgart.de (C. J. Masinde); 82rodo1mpg@hft-stuttgart.de (D. C. Rono); michael.hahn@hft-stuttgart.de (M. Hahn)

DOI: 10.22059/eoge.2019.290064.1064

differentiating bare lands from sealed surfaces. However, this effect is reduced in wet and /or summer seasons.

### 1.2 Building Indices

The spectral properties of the detected light energy depend on the nature of the earth's surface. For instance, blue and red wavelengths are absorbed by plants while green and infrared are reflected. That is how we can perceive chlorophyll containing vegetation as being green in color. Infrared cannot be visibly seen by our eyes, but the wavelength can be assigned a false color by computer color guns and visualized. (Moreira & Galvão, 2010) concluded that short wave infrared on the other hand has a great potential for discriminating urban impervious surfaces from the general landscape. Thus, it is widely used when detecting concreted and any other surface sealed areas.

Creating indices improves on the overall identification of remotely sensed features on images. They make the investigated regions stand out from the rest of the imagery. In this regard, several indices have been developed to aid in the identification of urban spaces. Most of the indices make use of short-wave infrared. The rest use other criteria such as (Crist & Kauth, 1986).who utilized a biophysical combinational index (BCI) which makes use of tasseled cap components. These are composite values representing brightness, greenness and wetness generated through tasseled cap transformation and coefficients which are sensor specific. Table 1 gives a list of used building indices.

Table1. Selected building indices adopted from (Juan et al., 2018) and (Index Database, 2019)

| Remote Sensing Building Indices | Sentinel 2 band combination |
|---|---|
| UI (Urban Index) | $\frac{SWIR2 - NIR}{SWIR2 + NIR}$ |
| NDBI (Normalized Difference Built-Up Index) | $\frac{SWIR1 - NIR}{SWIR1 + NIR}$ |
| IBI (Index-based Built-Up Index) | $\frac{\frac{2*SWIR1}{SWIR1+NIR} - \left\{\frac{NIR}{NIR+RED} + \frac{GREEN}{GRN+SWIR1}\right\}}{\frac{2*SWIR1}{SWIR1+NIR} + \left\{\frac{NIR}{NIR+RED} + \frac{GREEN}{GRN+SWIR1}\right\}}$ |
| NBI (New Built up Index) | $\frac{R * SWIR1}{NIR}$ |
| BUI (Built up Index) | $\frac{2*((R*SWIR2) - (SWIR1*SWIR1))}{(R+SWIR2) - (SWIR1+SWIR2)}$ |
| BCI (Biophysical Combinational Index) | $\frac{(TC1 + TC3)/2 - TC2}{(TC1 + TC3)/2 + TC2}$<br>TC = Tasselled Cap Component<br>TC1 = brightness<br>TC2 =greenness<br>TC3 = wetness |

### 1.3 Study Area

This study area as shown in Figure 1 was conducted for the city of Nürtingen and its surrounding areas. Nürtingen is located between 9° 17' 31.07" E, 48° 34' 36.05" N and 9° 25' 30.63"E, 48° 40' 10.2" N to the East of city of Stuttgart, Baden-Württemberg, Germany. Its land use types that can be visually distinguished in a satellite imagery include urban zones, roads, forests, meadows, Neckar River and farmland

Figure 1: Study area, Nürtingen Stuttgart

Figure 2 shows polygons representing built up areas; residential, industrial and commercial zones obtained from OSM (Open Street Map) data accessed from (Geofabric, 2019).

## 2. Data and methods

### 2.1. Data

In this study, the data captured by Sentinel 2 satellite was used. Sentinel 2 satellites capture data in 13 bands with resolutions of 10, 20 and 60m at a temporal resolution of every 2 days at high latitudes but in other areas it can be up to 5 days depending on latitude. This means that a very large selection of recent data is available. However, its major limitation is the presence of clouds since Sentinel 2 uses an optical sensor. The wavelengths range from 443 to 2190 nanometers. Figure 3 shows some details of the Sentinel 2 spectral bands.

Google Earth Engine (GEE) and Quantum GIS (QGIS) were the software tools used in this exercise. GEE has recently gained popularity in satellite data analysis, for its cloud computing capabilities and multi petabyte catalog of open access satellite data. In addition, GEE provides ready to use scripts written in JavaScript for geospatial analysis which users can easily customize for their use. The results can also be visualized in a window with a base map overlay. All computations are carried out in GEE.

QGIS is used to generate random points to aid in accuracy assessment.

| Name | Units | Min | Max | Scale | Resolution | Wavelength | Description |
|---|---|---|---|---|---|---|---|
| B1 | | | | 0.0001 | 60 meters | 443.9nm (S2A) / 442.3nm (S2B) | Aerosols |
| B2 | | | | 0.0001 | 10 meters | 496.6nm (S2A) / 492.1nm (S2B) | Blue |
| B3 | | | | 0.0001 | 10 meters | 560nm (S2A) / 559nm (S2B) | Green |
| B4 | | | | 0.0001 | 10 meters | 664.5nm (S2A) / 665nm (S2B) | Red |
| B5 | | | | 0.0001 | 20 meters | 703.9nm (S2A) / 703.8nm (S2B) | Red Edge 1 |
| B6 | | | | 0.0001 | 20 meters | 740.2nm (S2A) / 739.1nm (S2B) | Red Edge 2 |
| B7 | | | | 0.0001 | 20 meters | 782.5nm (S2A) / 779.7nm (S2B) | Red Edge 3 |
| B8 | | | | 0.0001 | 10 meters | 835.1nm (S2A) / 833nm (S2B) | NIR |
| B8A | | | | 0.0001 | 20 meters | 864.8nm (S2A) / 864nm (S2B) | Red Edge 4 |
| B9 | | | | 0.0001 | 60 meters | 945nm (S2A) / 943.2nm (S2B) | Water vapor |
| B11 | | | | 0.0001 | 20 meters | 1613.7nm (S2A) / 1610.4nm (S2B) | SWIR 1 |
| B12 | | | | 0.0001 | 20 meters | 2202.4nm (S2A) / 2185.7nm (S2B) | SWIR 2 |

Figure 3. Spectral bands for Sentinel 2 sensors adopted from Google Earth Engine data catalog

### 2.2. Atmospheric correction

At the beginning of the study, the only Sentinel 2 data available in GEE was top of atmosphere (TOA) data. Therefore, there was need to apply atmospheric corrections to the data to obtain bottom of atmosphere (BOA) data that can be used for further analysis.

Until now, GEE does not support any kind of atmospheric corrections. Amongst the alternatives available for atmospheric corrections, e.g. sen2cor, our best alternative was (PY6S, 2012), a python code available in (GitHub, 2019) which the creator utilized a py6s interface. The interface is a model that simulates a satellite signal in the solar spectrum and removes errors caused by atmospheric factors. Simply, the functionality is that the python code is run locally, the process generates atmospherically corrected images which are then uploaded directly to the asset folder in GEE. Although it can be time consuming when working with many images, the corrected images are directly accessed in GEE therefore cancelling out any need for downloads to local storage and uploads to GEE.

However, halfway through this study, Sentinel 2 BOA data covering the study area became available in GEE catalog. For more flexibility in choosing data sets for investigations, we opted to use BOA data available from GEE catalog.

### 2.3 Implementation using Sentinel 2 data

The building indices that are implemented mainly utilize the SWIR and NIR bands. To investigate the spectral signatures of each visible land cover class, several polygons each representing bare, forest water and urban land classes are created and the mean values in each band plotted in a scatter plot. This is done in GEE

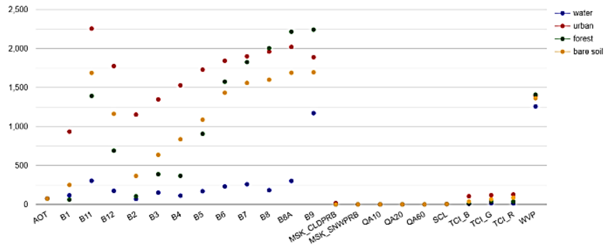


Figure 4. Band values for water, bare, urban, forest land cover classes.

From Figure 4, it is noted that urban surfaces have highest band values in SWIR 1 and SWIR2 bands. They also have high values in NIR bands, but so do forests and bare lands.

The indices were computed using the spectral band combinations as shown in Table 1. All the computations were done in GEE and scripts are available in GEE online repository. (Appendix 1). First, the image collection of images taken through 1 year from April 2018 to May 2019 with cloud coverage of less than 50 % was selected and this was a total of 39 images. Again, this shows that clouds could undermine the full utilization of all images captured by an optical sensor. The index for one image is calculated and then the formula is looped through the image collection to calculate indices for all images. The resulting index image is added to the original image as a band.

To differentiate sealed surfaces from other areas, Otsu method is applied to get a threshold index value that alienates sealed surface. This process is described further in section 2.4. Finally, an accuracy assessment tool is run by using a set of sample random points generated with reference to open street map data, which is considered as ground truth. All these scripts to perform the above methods, are contained in one file which takes very little time, as little as few seconds to execute and can be repeated until desired results are obtained. This is one of the advantages of using GEE.

### 2.4. Thresholding

The generated Index images in section 2.3 are of type float with different range of values depending on firstly, the image and the index too. For example, in the image used in Figure 7, NDBI values range from -0.7 to 0.4, the ideal range of NDBI values is from -1 to 1. The other indices have different values ranges, but they are all float numbers.

Theoretically, it is expected that index images should highlight sealed surfaces, as areas with maximum values. To delineate these sealed surfaces from other land covers, one needs to know the minimum and maximum pixel values representing sealed surfaces in the index images. To do this, (Kaimaris & Patias, 2016) utilized manual method of comparing the index image with the original satellite image and noting down the range of values covered by sealed surfaces. Another possible method is to use classified image for comparison instead of the original satellite image. Both methods are time consuming and prone to under or overestimation. (Jungnickl, 2017) utilized the median and standard deviation of pixel as the lower and upper threshold.

In this study, however, to extract the sealed surfaces, we utilize Otsu thresholding method proposed by (Otsu, 1979). Otsu method assumes normal distributed data with two classes following a bimodal histogram. It calculates the optimal threshold value separating these two classes, background and foreground. The threshold value is then used to binarize the index image; values greater than the threshold represent sealed surfaces.

The Otsu's Binarization for thresholding is a quick method for automatic thresholding. However, it works best only if the data portrays a bimodal distribution (Jadwiga, 2006). From Figure 5, all other index images show almost a bimodal histogram except NBI and BCI. For this index images we still use Otsu method since it still gives a reasonable threshold value.

## 3. Experiments and results

### 3.1 Index Images

Figure 6 shows RGB image of Nürtingen city and its surrounding areas taken in May 7th, 2018.

Figure 6. May 2018 RGB image

From visual inspection, the locations of each land cover class can be easily identified. The rectangle with a yellow outline shows an area of bare farmlands, which will be important for comparison in subsequent discussions. The index images generated from section 2 are shown in Figure 7. The index images shown in this figure are the resultants of building indices computations. The red color shows high values (sealed surfaces) while the blue color shows low values (unsealed surfaces). BCI and NBI visually increased intensity contrast between urban areas and background.

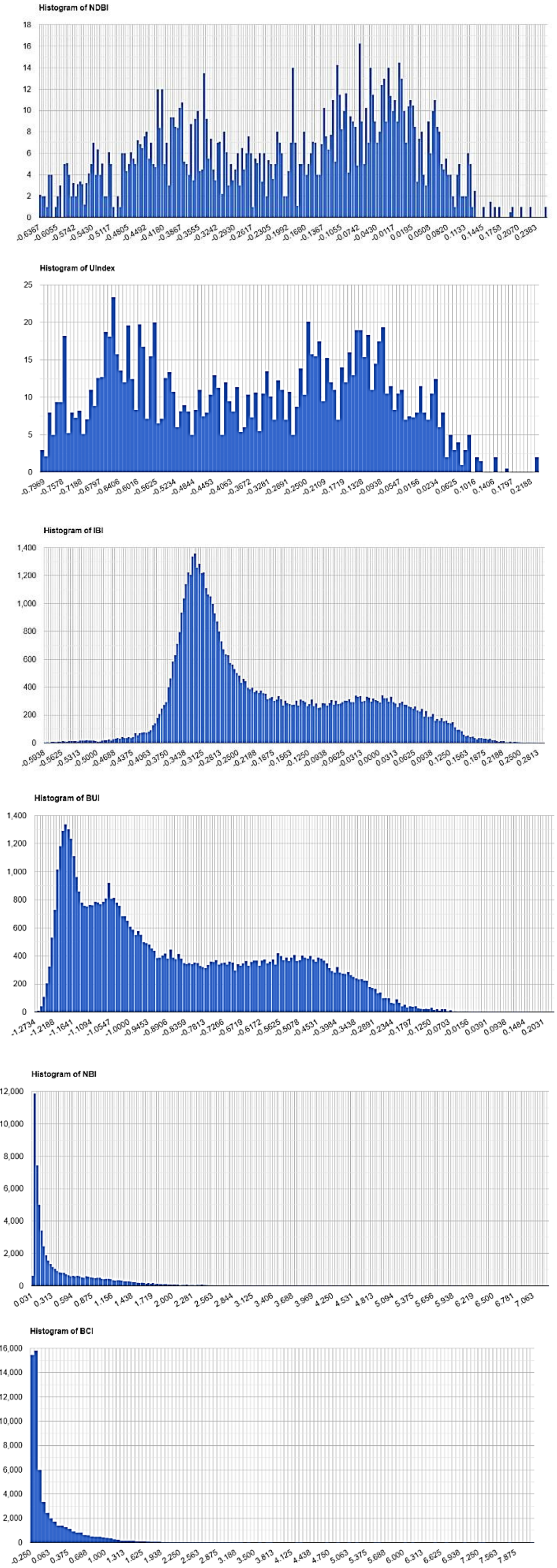


Figure 5. Histograms of index images

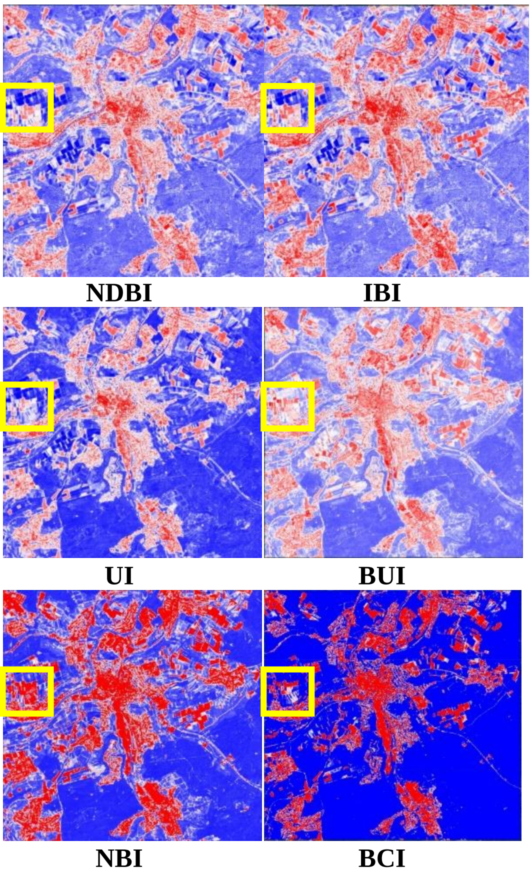

Figure 7. Index images

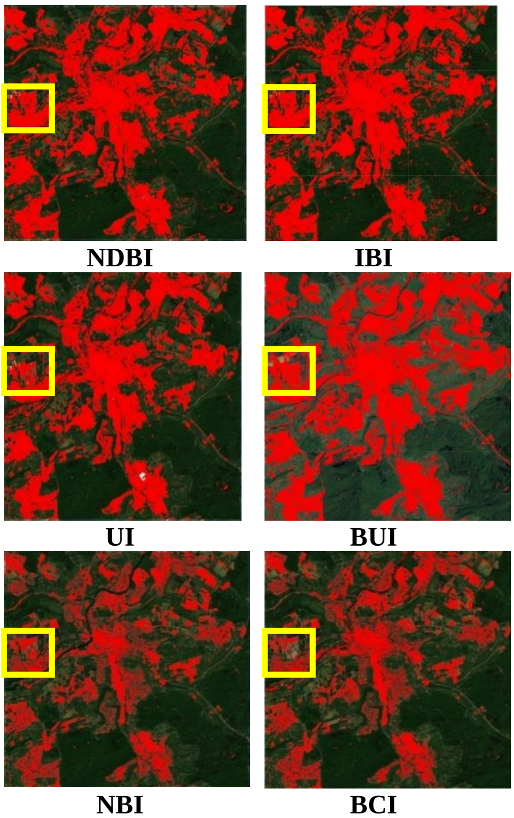

Figure 8. Binary index images after applying Otsu method

The red patches represent surfaces with index values above the threshold value determined by Otsu method.

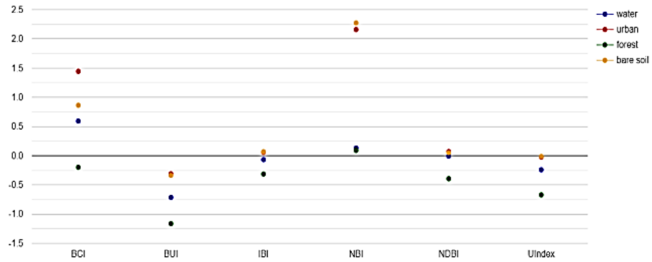

Figure 9: Average index values per land cover.

Figure 9 is a scatter plot showing average values for each index for selected land covers. Polygons representing each land cover were created and used as input to generate this scatter plot. The assumption is that the created polygons are perfectly representative of the whole image.

Analyzing the scatter plot, NDBI and IBI emphasize water, bare and urban almost simultaneously. UIndex (UI) emphasizes both urban and bare simultaneously but also the values of water are very close. On the other hand, BUI and NBI highlights both bare and urban surfaces simultaneously while BCI highlights more the urban surfaces, but values for bare lands are a little bit lower but very close to urban. This already shows the impossibility of the indices to alienate only sealed surfaces. In all cases, values for forests stay low indicating that building indices easily eliminates forests.

Analyzing Figure 8 and Figure 9, the result show that in all indices, it is not possible to alienate only sealed surfaces. The supposedly sealed surfaces still contain traces of other land cover types especially water and bare areas. Assessing the performance of individual indices:

**Water**: NBI performs very well in eliminating water areas. River Neckar is eliminated from sealed surfaces. In all other indices traces of water have been misclassified as sealed with BCI having the least effect. This shows that to achieve better results with most indices, water areas needs to be masked out.

**Roads**: NDBI, BUI, IBI, UI classify almost all visible roads as sealed. In NBI and BCI, roads are left out.

**Forests**: All indices eliminate forest areas effectively.

### 3.2. Confusion between Bare soil and sealed surfaces

(Valdiviezo-N et al., 2018) found out that in as much as building indices are easy to implement, they are faced with a difficulty of differentiating between bare land and sealed surfaces. In remotely sensed images, there exists a spectral similarity between these two land covers which makes their classification somewhat impossible. Investigations of the effectiveness of each index to differentiate these two land cover classes lead to the following discussion:

From Figure 8 the selected area (within the rectangle with a yellow outline) should be bare land but is falsely represented by all indices as sealed surfaces with BUI having the largest misrepresented area.

In NBI and BCI, the area falsified as sealed surfaces is greatly reduced as compared to other indices with BCI having the least misrepresented area.

### 3.3. Seasonal dependency

As mentioned above, Stuttgart region experiences different seasons throughout the year. (Valdiviezo-N et al., 2018), reveals that performance of indices to alienate sealed surfaces increases during wet months. This can be attributed to the fact that during wet months plants have already grown leaves therefore covering up ground bare surfaces, or it could be attributed to spectral response of wet and dry soil. During dry months, leaves fall off the vegetation and hence expose bare ground. To help evaluate the differences, the indices were computed on all images throughout the year, May 2018 to May 2019. For purposes of comparison, two images are selected; 07 May 2018 (Figure 6) taken as a wet month and 14 November 2018, as a dry month (Figure 10). Comparing results from the two images (Figure 8 and Figure 11) leads to the following discussion of the November image (Figure 11):

- Performance of all indices generally decreases
- NDBI, UI, IBI leave out shiny highly reflective surfaces
- Increased misclassified bare lands in NBI and BCI as compared to Figure 8.
- BUI: Minimal seasonal dependence. BUI is consistent. But still there is significant confusion of bare soil and urban in both months.

Figure 10. November 2018 RGB image

### 3.4. Accuracy assessment

To validate the results and assess the performance of these indices to accurately distinguish sealed surfaces from non-sealed surfaces, accuracy assessment is done. Confusion matrix is one of the methods of accuracy assessment. Using freely available land use vector data from OpenStreetMap, 300 points are randomly sampled from each of the two generalized land cover types: sealed and non-sealed. Thus, a total of 600 pixels are sampled from both sealed surfaces (urban zones such as residential, commercial and industrial areas), and non-sealed surfaces such vegetation and water (appendix 2). This process of sampling of points is done in QGIS. With the sample points available, a confusion matrix is created by comparing, for each point, its land use type on the OSM land use and the binary index map. The degree at which the index map agrees with the reference data is determined by several types of accuracies as shown in Table 2. Accuracy assessment is implemented in GEE.

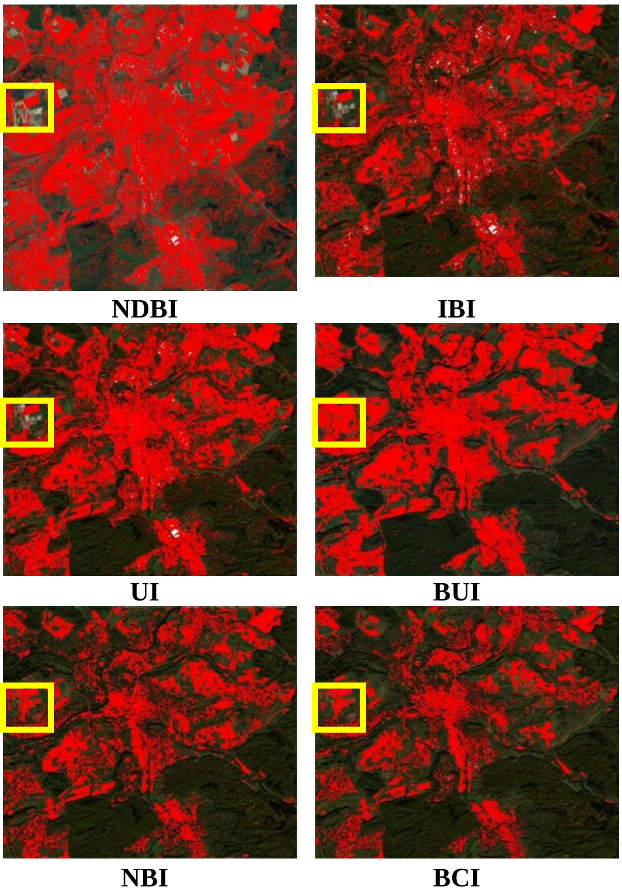


Figure 11. Sealed surfaces binary image, 14th Nov 2018

Table 2: Accuracies achieved by indices for the May 2018 image

| | Overall accuracy | Consumers accuracy | Producers accuracy | Kappa |
|---|---|---|---|---|
| BCI | 72 | 88 | 51 | 0.44 |
| BUI | 82 | 80 | 86 | 0.64 |
| NBI | 74 | 87 | 56 | 0.47 |
| UI | 86 | 87 | 85 | 0.72 |
| NDBI | 84 | 83 | 85 | 0.67 |
| IBI | 84 | 83 | 86 | 0.69 |

Table 2 shows that UI, NDBI, IBI, BUI achieved higher accuracies of 86%, 84%, 84%, 82% respectively while BCI and NBI having lower accuracies of 72% and 74%.

## 4. Conclusion

In this study we have implemented building indices using Sentinel 2 data. Although indices are easy to implement, there are several limitations that undermine the ability of these methods to extract sealed surfaces. Such factors include seasonality, sensor characteristics, location of area of interest, and diverse nature of materials used for sealed surfaces e.g. buildings can have metal, tiles, green roofs etc.

It has been shown that the selected indices cannot separate bare land from sealed surfaces completely, though BCI and NBI reduce this effect significantly (appendix 3). We also discussed that all the selected indices effectively eliminate forests because of the apparent difference in spectral characteristics. Of the selected indices, NBI is the only index that eliminates water areas. When working with building indices, it is important to consider the time of image capture, because results show that performance reduces during dry months. Better results can be achieved during wet and humid months when the area covered by bare ground has reduced (i.e. then having vegetation).

## Acknowledgement

This publication was produced in the project „Neue Möglichkeiten der Datenerhebung, -analyse und Auswirkungsabschätzung im Wasser-Energie Nexus – NeMDa“, which is part of the Baden-Württemberg-STIPENDIUM for University Students – BWS plus, a programme of the Baden-Württemberg Stiftung. The authors would like to thank the Baden-Württemberg Stiftung for its support.

## References

Crist, E.P.,& Kauth, R.J., (1986). The Tassled Cap de-mystified. Photogramm. Eng. Remote Sens.

ESA ,(2019). Sentinel Online at https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial (10/05/2019).

Jadwiga, R., (2006). DIGITAL IMAGE PROCESSING TECHNIQUES FOR SPECKLE REDUCTION, ENHANCEMENT, AND SEGMENTATION OF OPTICAL COHERENCE TOMOGRAPHY (OCT) IMAGES, in: Optical Coherence Tomography. https://doi.org/10.1016/b978-012133570-0/50013-5

Geofabrik, (2019). Download Open Street map data at http://download.geofabrik.de/europe/germany/baden-wuerttemberg.html (11/06/2019)

GitHub, (2019). ee-atmcorr-timeseries at https://github.com/samsammurphy/ee-atmcorr-timeseries/tree/master/atmcorr (10/05/2019).

Google earth engine at https://earthengine.google.com/ (17 June 2019)

Index Database, (2019). A database for remote sensing indices. www.indexdatabase.de/db/i-single.php?id=93 (10 June 2019)

Jungnickl, C.,& Bill, R., (2017). Abschätzung des Versiegelungsgrades mit Sentinel-2-Daten durch Anwendung von Bebauungsindizes. In: Meinel, G, Schumacher, U., Schwarz, S., Richter, B. (Hrsg.). Flächennutzungsmonitoring IX. Nachhaltigkeit der Siedlungs- und Verkehrsentwicklung? Berlin:Rhombos, IÖR Schriften 73, S. 21-29. ISBN: 978-3-944101-73-6

Kaimaris, D., & Patias, P., (2016). Identification and Area Measurement of the Built-up Area with the Built-up Index (BUI). Int. J. Adv. Remote Sens. GIS. https://doi.org/10.23953/cloud.ijarsg.64

Moreira, R.C., & Galvão, L.S., (2010). Variation in spectral shape of urban materials. Remote Sens. Lett. https://doi.org/10.1080/01431161003692032

NASA, (2019). MODIS. modis.gsfc.nasa.gov/about/components.php (10 May 2019).

NASA ,(2000). Earth Observatory. earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php (10 May 2019).

Otsu, N., (1979). THRESHOLD SELECTION METHOD FROM GRAY-LEVEL HISTOGRAMS. IEEE Trans Syst Man Cybern.

PY6S, 2012. A python interface to 6S. py6s.readthedocs.io/en/latest/ (10 May 2019).

USGS ,2015. Landsat Missions. www.usgs.gov/land-resources/nli/landsat (10 May 2019).

Valdiviezo-N, J.C., Téllez-Quiñones, A., Salazar-Garibay, A., & López-Caloca, A.A., (2018). Built-up index methods and their applications for urban extraction from Sentinel 2A satellite data: discussion. J. Opt. Soc. Am. A. https://doi.org/10.1364/josaa.35.000035

**Appendix 1**:

Google earth engine scripts: https://code.earthengine.google.com/?accept_repo=users/dorothyrono/RSS
Or clone its Git repository by running the following command in a terminal:
https://earthengine.googlesource.com/users/dorothyrono/RSS

**Appendix 2**: Sample reference points for accuracy assessment

**Appendix 3**: Built up area polygons overlaid with sealed surfaces generated from BCI index