Urban land cover mapping, using open satellite data. Case study of the municipality of Thessaloniki.

Ioannis Tsolakidis and Marios Vafiadis
Division of Hydraulics and Environmental Engineering, School of Civil Engineering, A.U.Th, GR-54124 Thessaloniki, Macedonia, Greece
E-mail: tsolakidisioannis@gmail.com

Abstract. The use of open satellite data has become a valuable tool for monitoring the urban environment due to high temporal, spectral and spatial resolution. Planners and stakeholders can rely on these data since they offer an alternative point of view of a very complex urban scenery. The processing of satellite images with the combined use of different types of data (statistical, in situ etc) can extract information for the present situation, in urban and suburban areas. This study use image data from the Sentinel satellite mission, in order to extract urban land cover characteristics in Thessaloniki city. The Sentinel mission started the image acquisition on 2014. The data set are freely available for downloading. The use of high resolution satellite images (e.g IKONOS) in urban studies, has been proved very useful all these years. On the other hand the use of a medium size spatial resolution satellite, as Sentinel-2, with 10m pixel size, is a very promising optical earth-observation satellite, which is tested in this study. Image processing techniques used to produce classified map of urban characteristics and land cover classes (built up areas, vegetation etc). The contribution of a radar image of Sentinel-1, in the classification process, is also tested. All the above processing is done with an automatic procedure, giving additional value to the proposed methodology. The final classified map will give planners a management tool for decision-making concerning the sustainable urban development and will be the basis for implementing hydrology studies, planning, land use change detection and other.

1. Introduction
Urbanization is considered as a spatial transformation of the economy, where the population moves through migration from an agricultural, rural based existence to one where production occurs in cities of endogenous numbers and size [1]. This transformation of the landscape, due to urban expansion, alters the natural land cover types, to impervious urban land. As a consequence, this produce both short and long term changes in the urban environment. Long term changes deal with impacts of urbanization on climate conditions of urban and suburban areas [2]. Departments of urban planning need these information in order to take the right decisions and address all the arise problems [3]. Monitoring the urban environment, recording details about the spatial changes of land cover and land use in a level of a few square meters pixel size can be implemented with the use of remote sensing images and image processing techniques. Landsat satellite was one of the first satellite systems that was used for change detection in urban studies [4], [5]. The most critical factors in monitoring the urban environment from space are the cost of the satellite image and how easily these data can be found. The spatial analysis, the temporal and spectral resolution of the sensor are also very important parameters since they can detect different type of changes.
In recent years a term called open data has been widely used and the growth of “openness” movement has attracted the scientific community. Open data are data that can be freely used, re-used and redistributed by anyone. In this study, open satellite data from Sentinel-2 mission of Copernicus Earth Observation Programme are being used in order to extract urban land cover characteristics like green space, urban (impervious) land and other cover types, in the municipality of Thessaloniki, through supervised classification of the satellite image [6]. The contribution of a radar image of Sentinel-1 satellite, of the same acquisition date is being investigated in the classification accuracy. Finally, the NDVI indice of the presence of vegetation is compared with the similar class from the classifications. Qiu et al [7] studied the performance of Vegetation Red-Edge bands on improving land cover classification. Comparing classified satellite images of different dates, the loss and gain of different land covers can be monitored. The processing of the satellite images was done in an open-source software. Both the satellite images and the software were downloaded from the site of European Space Agency (ESA).

2. Methodology
2.1. Data
Two different satellite products were used in this study. A Sentinel-2 level 1C product, with acquisition date 18/02/2019 (Figure 1) and a Sentinel-1A, C-band synthetic aperture radar (SAR) level-1 GRD product, of the same date, in dual polarisation (VV+VH) and descending mode (Figure 2). The Sentinel-2 product is a multispectral image with 13 spectral bands and 3 different spatial resolutions (Table 1). Figure 1 shows part of the region of Central Macedonia of Greece, including the prefectures of Serres, Kilkis and Thessaloniki. The city of Thessaloniki and Thermaikos Gulf are at the south part. Figure 2 shows a broader area of Central Macedonia with no orientation, due to satellite passage. Data provision is available for downloading from Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/home).

Table 1. Spatial and spectral characteristics of Sentinel-2 sensor.

| spatial resolution | Band number (central wavelength) |
|--------------------|----------------------------------|
| 10m                | \(B_2(490\text{nm}), B_3(560\text{nm}), B_4(665\text{nm}), B_8(842\text{nm})\) |
| 20m                | \(B_5(705\text{nm}), B_6(740\text{nm}), B_7(783\text{nm}), B_{8a}(865\text{nm}), B_{11}(1610\text{nm}), B_{12}(2190\text{nm})\) |
| 60m                | \(B_1(443\text{nm}), B_9(945\text{nm}), B_{10}(1375\text{nm})\) |

Level-1C processing of the Sentinel-2 product is both radiometrically and geometrically corrected, including orthorectification and spatial registration on a global reference system, with sub-pixel accuracy [8]. The Sentinel-2 image is a top of atmosphere (ToA) image. On the other hand, the Sentinel-1 instrument is able to transmit horizontal (H) or vertical (V) linear polarizations. The instrument is able to receive, on two separate receiving channels, both H and V signals simultaneously. Dual-polarisation products are provided in the form of two images. The images have the same product characteristics and are co-registered [9]. The range pixel spacing is 10m.

2.2. Preprocessing of Sentinel-1 and Sentinel-2 images
The preprocessing of the Sentinel-1 and Sentinel-2 images was done in Sentinel Application Platform (SNAP) software. SNAP is a free software which can be downloaded from ESA
Figure 1. The downloaded Sentinel-2 product (composite color image R=4,G=3,B=2). The image shows part of the region of Central Macedonia of Greece. Thessaloniki is at the south part, with Thermaikos Gulf.

(a) Amplitude VH
(b) Intensity VH
(c) Amplitude VV
(d) Intensity VV

Figure 2. The 4 images of the downloaded Sentinel-1 product, with dual polarization (VV VH). The images shows a broader area of Central Macedonia with no orientation.

website and is composed of toolboxes that can manage, process, analyse and visualize both multispectral and radar images. For the Sentinel-2 image the processing steps, included an atmospheric correction with the AT2COR routine in SNAP, in order to reduce the top of atmosphere (ToA) image to Bottom of Atmosphere (BoA). The atmospheric correction is an effort made to reduce the noise from the interaction (diffuse, attenuation) of the radiance path with the atmosphere. Then, the resampling of the resolution of the different bands image pixels to 10m pixel size followed. The image was cropped to a smaller area of interest so as to have a
better management of the process. A reprojection was followed to the Greek Reference System of 1987. The final multispectral (MS) image is presented in Figure 4. As mentioned in § 2.1, Sentinel-2 level-1C includes radiometric correction and orthorectification and for that reason only the above processing steps were implemented. In the final multispectral (MS) image, a principal component transformation was applied. The second band (PC2) from the transformed image (Figure 6) distinguished the impervious land better. On the other hand, the vegetation areas have an important role in land surface temperature [10]. For the detection of the vegetation areas the NDVI indice was applied to the final MS image, using the Red and NIR Bands and the equation 1 (Figure 7).

\[
NDVI = \frac{NIR - R}{NIR + R}, \quad (-1 < NDVI < +1)
\]

The processing steps of Sentinel-1 image included a radar multilooking in order to produce square pixels, radiometric calibration and finally terrain correction with the use of global digital elevation model (SRTM). The terrain correction left blank the sea, due to DEM absence in that area. Finally, the image was reprojected to the Greek Reference System of 1987. The final Radar image is presented in Figure 5. A critical task, for the continuing of the study, was the classification of the final images and the extraction of land cover. More specifically there were tested two images, the final MS image and a synthetic one, that was composed of the bands of MS image, the PC2 Band and the Radar image bands [11]. Both images were classified through a supervised procedure, using the Random-Forest algorithm [12], in SNAP software. Random Forest classifier is a non parametric machine learning technique, capable of using continuous and categorical data sets. It is also, statistically unbiased that it can handle an increasing data sample size and has several advantages over traditional remote sensing classification algorithms. In Random Forests several decision trees are created (grown) and the response is calculated based on the outcome of all of the decision trees, [13], [14]. The processing chain of Sentinel-1,2 satellite images is presented in Figure 3.

![Figure 3. Processing chain of Sentinel images.](image-url)
2.3. Supervised classification

For the classification of the two images (LS and MS), the number and type of the classes had to be selected, taking into consideration the urban environment and the spatial analysis of 10m of the Sentinel images [15]. Nine classes were selected, one for urban/impervious surfaces (Urban), two for vegetation (Healthy vegetation and Shrub vegetation), one for soil (Bare soil), one for rooftops (Concrete/metal roof), one for tiles (Tile roof), one for shadow and two for water (Shallow and Deep water). After the determination of number and type, the selection of the training data for each class from the MS image followed, in order to proceed to the supervised classification with the Random Forest algorithm. The predictor variables for the classification of the MS image were the 13 bands and for the LS image were the 13 bands from MS, the PC2 band and the 4 bands from the radar image [11]. The final classified images are presented in Figures 8, 9. The accuracy of the classified images was estimated with the calculation of the Accuracy Matrix and the Kappa coefficient. For the Accuracy Matrix, 10 samples per class were chosen and the corresponding statistics were calculated (Tables 2, 3). Finally, the classified images were cut exactly at the boundaries of the municipality of Thessaloniki (Figures 10, 11). The digital boundaries of the Municipality exported from the OpenstreetMap dataset in shapefile format.

Based on the results by the supervised classifications with the Random Forest algorithm, both positive and negative comments can be observed by the two classified images (MS and LS). The classification of MS image was better, with an overall accuracy equal to 95.06%. Especially, the green areas were better classified. The presence of "salt and pepper" effect was strong, due to pixel based approach and the heterogeneous urban environment. Finally, the remaining
Unclassified samples were 5. On the other hand, the overall accuracy of the LS classified image, was equal to 91.46%, the classification of urban areas had lower accuracy than the MS image, the effect of ”salt and pepper” was slightly improved and the unclassified samples were 2.

Figure 8. The Random-Forest Classified MS image.

Figure 9. The Random-Forest Classified LS image.

Figure 10. The Random-Forest Classified MS image of the Municipality of Thessaloniki.

Figure 11. The Random-Forest Classified LS image of the Municipality of Thessaloniki.
Table 2. Producers and Users Accuracy for MS classified image (Figure 8)

| Class                  | Producers Accuracy (%) | Users Accuracy (%) |
|------------------------|------------------------|--------------------|
| Urban                  | 88.89                  | 100.00             |
| Healthy vegetation     | 90.00                  | 90.00              |
| Concrete/metal         | 100.00                 | 100.00             |
| Tile roof              | 100.00                 | 90.00              |
| Shadow                 | 100.00                 | 100.00             |
| Shallow water          | 100.00                 | 100.00             |
| Shrub vegetation       | 100.00                 | 100.00             |
| Bare soil              | 60.00                  | 60.00              |
| Deep water             | 100.00                 | 100.00             |

Overall Accuracy 95.06% \( \hat{K} \) 0.94

Table 3. Producers and Users Accuracy for LS classified image (Figure 9)

| Class                  | Producers Accuracy (%) | Users Accuracy (%) |
|------------------------|------------------------|--------------------|
| Urban                  | 87.50                  | 77.78              |
| Healthy vegetation     | 66.67                  | 85.71              |
| Concrete/metal         | 100.00                 | 100.00             |
| Tile roof              | 100.00                 | 90.00              |
| Shadow                 | 100.00                 | 100.00             |
| Shallow water          | 100.00                 | 100.00             |
| Shrub vegetation       | 100.00                 | 100.00             |
| Bare soil              | 62.50                  | 55.56              |
| Deep water             | 100.00                 | 100.00             |

Overall Accuracy 91.46% \( \hat{K} \) 0.90

3. Land cover analysis of the classified images
From the classified images of the Municipality of Thessaloniki (Figures 10, 11), it was easy to compute the corresponding areas of classes, exporting values of land cover in square meters. These values can be used either as grouped, in order to discriminate the impervious and non impervious land of the municipality, or as standalone for the calculation of indices, like square meters of green areas per capita. More specifically, from the pie chart of Figure 12, it is clear that impervious surfaces that composed of the classes Urban, Tile roofs, Concrete/metal are equal to 18.1 \( Km^2 \) or 79% of the total land cover. Non impervious areas (Healthy vegetation, Shrub vegetation, Shallow water, Bare soil) are equal to 4.74 \( Km^2 \) or 21% of the total land cover areas. The population of the Municipality of Thessaloniki according the 2011 National
Census was 352182. With a simple division of the areas of Healthy and Shrub vegetation by the population, it appears that the green area per capita is equal to 7.4 m². It should be mentioned that as green areas are referred both the part of the suburban forest of Seich Sou, that is within the boundaries of the Municipality and the football stadiums. The exclusion of these areas will decrease the green areas per capita indice. The impervious areas and the land cover are also vital to hydrological studies. Bhaskar and Suribabu [16] estimate surface run-off for urban area with the use of remote sensing, while Guo et al [17] produced land cover maps and assess the impacts of land cover change with L-THIA model.

![Figure 12. The Random-Forest Classified MS image of the Municipality of Thessaloniki.](image12)

![Figure 13. The Random-Forest Classified LS image of the Municipality of Thessaloniki.](image13)

3.1. Vegetation areas from NDVI image

In order to evaluate the results of the classification and especially the vegetation areas a simple rule was applied on the NDVI image. The range values of the NDVI is between −1 to +1. As stated by Goncalves et al [18], values between 0.64 to +1 represent dense vegetation, 0.56 to 0.64 open vegetation, 0.26 to 0.56 herbaceous vegetation, −0.14 to 0.29 urban and below −0.14 water. For the present study values greater than 0.4 are considered as vegetation. Thus a threshold was applied on NDVI image of figure 7 (NDVI > 0.4). The resulting image is presented in Figure 14, where only the vegetation areas of Municipality of Thessaloniki appear. The vegetation area from Figure 14 is 3370649 m². The same area from the supervised MS image is 2620073 m². The difference between the two vegetation areas is 749637 m² or 7191 pixels. This difference expresses the ability of the Random Forests algorithm to classify the vegetation areas.

The misclassification of vegetation, according to the chosen NDVI threshold results, is due to the complex and heterogeneous urban environment and the resolution of 10m pixel size. Small areas of vegetation were not correctly classified, because of mixing with other urban classes. Larger areas had better performance. Other parameters of the discrepancy, between NDVI and classification, is because of the chosen threshold value of 0.4, that has to be tested with in situ data. Perhaps another value would give results, closer to the classification. Also, the Random Forests classification parameters have to change, then should run the classification and compare the results again.
Figure 14. Vegetation areas of the Municipality of Thessaloniki from thresholded NDVI image (values > 0.4).

4. Conclusions
In this study, a methodology was presented, where with no cost and with the use of open satellite data, in a free software environment, advanced spatial analysis of the land cover of an urban area, was performed. More specifically, the open satellite data of Sentinel-1,2 mission from Copernicus Earth Observation Programme under the European Commission and the European Space Agency were used. Both multispectral and radar images, which were processed in a free software and supervised classifications of images with the Random Forests classifier, were produced. All the processing chain was done in the SNAP environment, in a short time, with great reliability, giving the opportunity to non expert users to implement such analysis. With the current methodology different indices and maps can be extracted such as the impervious surfaces of a city and the green areas. Monitoring of these parameters in short and long time basis and correlating with in situ data (meteorological etc), might lead to valuable information about climate change in an urban and suburban environment. The accuracy of the proposed methodology, can be improved. More precisely, the supervised classification of the satellite images, with the Random Forests algorithm, can be done using more training areas with more careful choice. Also, in the design of the accuracy matrix, the samples in each class has to be 40 – 50, in order to be statistical significant and a random sampling approach should be chosen. From the current results, the classification of MS image was better, with an overall accuracy equal to 95.06% than 91.46% of the LS image. The green areas of MS image were better classified than LS, and the presence of ”salt and pepper” effect was slightly improved in LS image. The NDVI threshold approach showed that in both classified MS and LS images the mixed urban green areas weren’t classified correctly. This was due to the complex and heterogeneous urban environment, where small areas of vegetation were mixed with the neighbouring land cover, as a result of 10m pixel size. Larger green areas were classified better. However that problem can be overcome with additional data input in the classification process (digital surface model etc) this methodology can help understand the impervious and non impervious surfaces, which are extremely important for the rise of the temperature and other problems in an urban environment.
Acknowledgements
We would like to thank Aggeliki Dimaki, environmentalist, for helping us, with the grammatical errors of our manuscript.

References
1] Henderson J V 2005 Handbook of Economic Growth 1 1543–1591 ISSN 1574-0684
2] Xian G and Crane M 2005 ISPRS Arch. XXXVI-8/W27
3] Yinghui X and Qingming Z 2009 JURSE 1–5
4] Hu T, Yang J, Li X and Gong P 2016 REMOTE SENS-BASEL 8 ISSN 20724292
5] Ma L, Zhao H and Li J 2016 ISPRS Arch. XLII-B8 965–972 ISSN 2194-9034
6] Topalolu R H, Sertel E and Musaolu N 2016 ISPRS Arch. 41 1055–1059 ISSN 16821750
7] Qiu S, He B, Yin C and Liao Z 2017 ISPRS Arch. 42 871–874 ISSN 16821750
8] ESA 2015 SENTINEL-2 User Handbook Tech. Rep. 1
9] Bourbigot M, Johnsen H, Piantanida R, Hajduch G and Poullaouec J 2016 Sentinel-1 ESA Tech. rep.
10] Chaithanya V V, Binoy B V and Vinod T R 2017 IJARSG 6 2088–2096
11] Clerici N, Valbuena C, Cesar A and Posada J M 2017 J. Maps 13 718–726 ISSN 17445647
12] Breiman L 2001 Mach Learn 45 5–32
13] Horning N 2010 Int. Conf. GIS-IDEAS 1–6 ISSN 1550-5499 (Preprint 1609–3631)
14] Kulkarni A D and Lowe B 2016 LIRITCC 4 58–63 ISSN 2321-8169
15] Bussios N, Tsolakidis Y, Tsakiri-Strati M and Goergoula O 2004 Proc. of the XXth ISPRS Cong. 215–220
16] Bhaskar J and Suribabu C R 2014 JJCE 8 70–80 ISSN 2225157X
17] Guo H, Li X, Huang Q and Zhang D 2011 ASPRS 2011 An. Conf.
18] Gonçalves A B, Godoi R d F, Filho A C P, Filho M T and Pistori H 2018 Anuario IEGEO 41 24–36 ISSN 01019759