2021 International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea) | 978-1-6654-1458-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/MetroSea52177.2021.9611583

Unsupervised classification based approach for coastline extraction from Sentinel-2 imagery

Emanuele Alcaras
International PhD Programme "Environment, Resources and Sustainable Development", Department of Science and Technology, Parthenope University of Naples, Centro Direzionale, Isola C4, (80143) Naples, Italy
emanuele.alcaras@studenti.uniparthenope.it
[0000-0001-5982-9719]

Pier Paolo Amoroso
International PhD Programme "Environment, Resources and Sustainable Development", Department of Science and Technology, Parthenope University of Naples, Centro Direzionale, Isola C4, (80143) Naples, Italy
pierpaolo.amoroso@studenti.uniparthenope.it
[0000-0002-7882-4339]

Valerio Baiocchi
Department of Civil, Constructional and Environmental Engineering, Sapienza University of Rome, I-00184 Rome, Italy
valerio.baiocchi@uniroma1.it
[0000-0003-4491-7868]

Ugo Falchi
Department of Science and Technology, Parthenope University of Naples, Centro Direzionale, Isola C4, (80143) Naples, Italy
ugo.falchi@uniparthenope.it
[0000-0002-9410-8498]

Claudio Parente
Department of Science and Technology, Parthenope University of Naples, Centro Direzionale, Isola C4, (80143) Naples, Italy
claudio.parente@uniparthenope.it
[0000-0002-8328-9764]

***Abstract* — Coastline extraction techniques from multispectral satellite images are of great interest for protection and monitoring of coastal areas. In this regard, the Sentinel-2 satellites can give a great contribution thanks to their wide coverage of the earth's surface. These images can be processed by GIS software, so as to detect the sea from all the rest. However, the traditional supervised classification requires the involvement of the operator to create suitable training sites: this approach, in addition to being associated to the operator's skill, often takes a long time to be completed. This contribution presents a study carried out on Sentinel-2 dataset and proposes the application of an unsupervised classification method, the k-means, on four different classification indices. The coastlines extracted by unsupervised classification are therefore compared with the coastline manually vectorized from the RGB composition. The results demonstrate the effectiveness of k-means for distinguishing, in the images produced by the indices application, two clusters (water / no-water) in a reduced time lapse if compared with the traditional supervised techniques.**

Keywords — Unsupervised, K-means, Sentinel-2, Coastline Extraction, GIS, NDWI

I. Introduction

Today, more than half of the world's population lives in coastal regions [1]. In Italy, the population density on the coasts is more than double the national average [2]. It is therefore evident that coastal areas are particularly sensitive and require continuous monitoring. The high population density of the coastal areas makes the mapping of the coastal zones essential, such as for safe navigation [3], resource management [4], environmental protection [5], and sustainable coastal development [6] and planning [7]. In order to support these activities, it may be useful to know the position of the coastline. This operation can be carried out in several ways, including SAR [8], Lidar [9], UAV Survey [10], RTK-GPS positioning technology [11], Optical Satellite [12]. Nevertheless, the integration of present-day data with historical cartography permits to evaluate coastal changes [13, 14].

Many studies have been carried out on coastline extraction from optical satellites imagery, with both high-resolution satellite [15], and medium resolution [16]. However, not only the spatial resolution must be considered, but also other characteristics of the sensor, such as the spectral resolution [17], and of the satellite itself, such as the satellite revisit period [18]. Sentinel-2 satellites offer a good coverage of the Earth surface, due to their short revisiting time (5 days at the equator and 2-3 days at mid-latitudes) [19]. Giving an unambiguous definition of a coastline is complex and the subject is widely debated in the literature [20]; what can be observed from satellite imagery is the instantaneous shoreline defined as the line of separation between land and sea at the time of image acquisition. In the following we will refer to the instantaneous shoreline. By defining the coastline as the intersection of the land surface and the sea, it can be easily extracted using spectral information signed by satellite images due to the different nature (and signature) of the two neighbouring elements [20]. Some indices, such as the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI), allow to emphasize the presence of water with respect to everything else: the studies on this subject are innumerable [21-23] and present cases of both supervised [24] and unsupervised image classification [25]. The supervised classification requires the involvement of the operator to create suitable training sites: this task, in addition to being associated to the operator's skill, it often is very tedious and time-consuming [26]. Unsupervised classification can offer a viable alternative in this case, speeding up the operations and maintaining optimum efficiency.

This article presents a study carried out on Sentinel-2 dataset and proposes the application of an unsupervised classification method, the k-means, on seven different classification indices. The coastlines extracted by unsupervised classification are therefore compared with the coastline manually achieved from the RGB composition.

II. Study Area and Dataset

ESA offers Sentinel-2 satellite imagery as a completely free data source. The Copernicus Sentinel-2 mission comprises a constellation of two satellites placed in the same sun-synchronous orbit, phased at 180° to each other, namely

Sentinel-2A and Sentinel 2-B [19]. In the framework of the geographical data and their storage [27], Sentinel images are part of Copernicus geodatabase as main focus of the Risk and Recovery Mapping Service [28]. The Sentinel-2 Multispectral Instrument (MSI) samples 13 spectral bands with different spatial resolution: four bands at 10 metres, six bands at 20 metres and three bands at 60 metres. The main characteristics of Sentinel-2A satellite sensor, which products are used in this article, are reported in Table 1 [29].

TABLE I. CHACTERISTIC OF SENTINEL-2A IMAGES

Bands	Central Wavelength (nm)	Resolution (m)
B1 - Coastal Aerosol	443	60
B2 - Blue	490	10
B3 - Green	560	10
B4 - Red	665	10
B5 - Red Edge	705	20
B6 - Red Edge	740	20
B7 - Red Edge	783	20
B8 - NIR	842	10
B8A - Narrow NIR	865	20
B9 - Water Vapour	945	60
B10 - SWIR Cirrus	1375	60
B11 - SWIR	1610	20
B12 - SWIR	2190	20

The study area covers a stretch of coast about 24 km long in south-eastern Sicily (Italy), east of the city of Pozzallo (Figure 1). This area has mainly sandy coasts which generally have a low vertical gradient (1° - 2°). Nearshore areas present smooth slopes with bathymetric contours running parallel to the coastline [30].

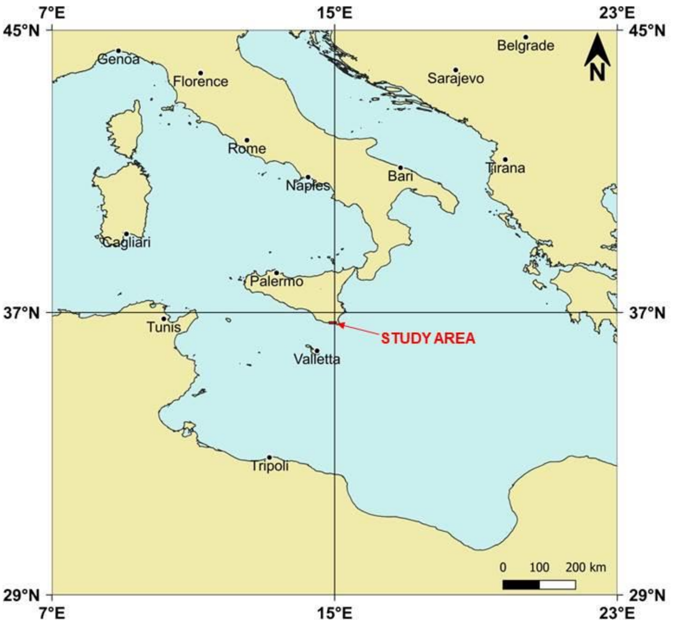

Fig. 1. Localization of the study area in the Mediterranean Sea in equirectangular projection and WGS84 geographic coordinates (EPSG: 4326).

The Sentinel-2A imagery used for the experiments were acquired in October - 29 - 2020. As reported in Figure 2, the study area is included in the following UTM/WGS84 zone 33N coordinate system: E_1 = 486,000 m; E_2 = 505,000 m; N_1 = 4,059,000 m; N_2 = 4,066,000 m.

Fig. 2. RGB composition of the Sentinel-2A images in UTM/WGS 84 plane coordinates (EPSG: 32633).

III. METHODS

A. Index Calculation

For this study 4 indices are experimented: Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Normalized Difference Water Index (NDWI), Modified NDWI (MNDWI).

NDVI is typically used for the identification of vegetated areas, however it allows to easily distinguish, in addition to vegetation, two other classes: bare soil and water [31]. NDVI formula can be expressed as follow:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

EVI is an optimized vegetation index, more responsive to canopy structural variation [32]:

$$EVI = 2.5 \cdot \frac{NIR - RED}{NIR + 6 \cdot RED - 7.5 \cdot BLUE + 1} \quad (2)$$

NDWI is designed to enhance the presence of water, by presenting a major contrast between water and land than the one that can be obtained from NDVI. NDWI can be achieved with the following formula [33]:

$$NDWI = \frac{GREEN - NIR}{GREEN + NIR} \quad (3)$$

Figure 3 shows the output obtained by the application of NDWI.

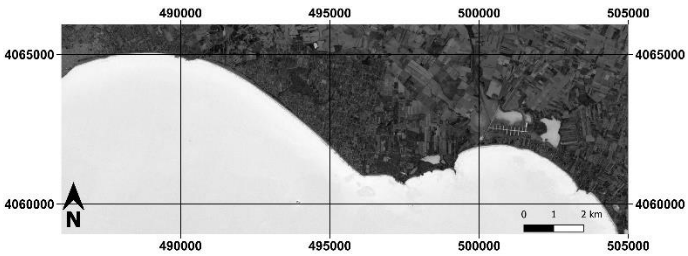

Fig. 3. NDWI obtained from the Sentinel-2A images in UTM/WGS 84 plane coordinates (EPSG: 32633).

An alternative version of NDWI is the MNDWI, which replace the Green band with the Blue one [34]:

$$MNDWI = \frac{BLUE - NIR}{BLUE + NIR} \quad (4)$$

Another possibility to identify the coastline is to calculate ratios between bands, for example the Red-Green ratio (RGR) and NIR-Red ratio (NRR) [35]:

$$RGR = \frac{RED}{GREEN} \quad (5)$$

$$NRR = \frac{NIR}{RED} \quad (6)$$

Another interesting approach is based on multiple steps starting from the difference between NDVI and NDWI [36]. In this paper we want to consider the first step only, as reported in Herndon et al. [37], and we call it Difference Index (DI):

$$DI = NDVI - NDWI \quad (7)$$

B. K-Means Clustering

K-means clustering is applied to each index, extracting two clusters (water/no-water) in order to classify the images. This procedure is carried out in SAGA GIS (Version 2.3.2).

K-means is a clustering algorithm, that partitions dataset into K number of clusters by standard Euclidean distance [38]; it is an iterative method that requires to repetitively define the centre of the clusters. It starts with arbitrary positions distributed on the data cloud, and then the positions are adjusted iteratively considering the results obtained each time. The process stops when there are no more variations in terms of allocation of centres and cluster boundaries [39].

In unsupervised classification, regardless of the adopted method, the resulting classes will be indicative of the natural spectral clusters present in the data, so clusters may also not correspond to actual covers or classes of materials [40].

In fact, there are pros and cons when adopting unsupervised classification. Pros: no prior knowledge of the investigated area is required; human error is greatly reduced; the classes are necessarily spectrally homogeneous; reduction of the work time. Cons: the obtained classes do not necessarily have a physical meaning; the user has limited control over the procedure and results.

However, in this case, the two resulting clusters effectively correspond to the researched classes: water/no water. The result of K-Means clustering applied to NDWI is shown in figure 4.

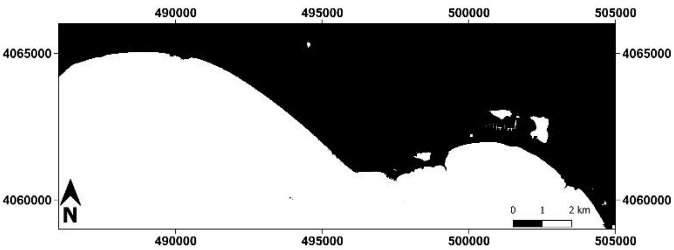

Figure 4 – Result of K-Means clustering applied to NDWI: land is represented in black, while water is represented in white.

C. Accuracy Tests

Once the classification is complete, the coastlines can be extracted. In this study seven coastlines are extracted. The products are compared with a reference one (Reference Coastline – RCL), achieved by manual vectorization of the RGB composition by means of visual analysis as shown in figure 5.

Figure 5 – RCL manually achieved (in red) on RGB composition of the Sentinel-2A images in UTM/WGS 84 plane coordinates (EPSG: 32633).

To assess the efficiency of the unsupervised classification two indices are calculated. The first one, namely Accuracy Index (AI), was proposed by Zangh et al., 2013 [41], and considers the relative shortening or lengthening of the extracted coastline with respect to the reference one:

$$AI = \frac{(L' - L)}{L} \cdot 100 \quad (5)$$

where L is the actual length of the coastline (or in this case the manually achieved coastline), L' is the length of the resulting coastline.

The second index, namely Ratio Index, taken into account was developed by Maglione et al., 2014 [17], considers the deviation between the two coastlines. Particularly, if the overlap between the two lines does not occur perfectly, polygons are generated: considering their area (A), and the actual length of the coastline (L), RI is defined as follow:

$$RI = \frac{A}{L} \quad (6)$$

IV. Results and Discussion

Table 2 reports the results obtained by applying AI and RI.

TABLE II. AI AND RI VALUES FOR THE EXTRACTED COASTLINES

Method	AI (%)	RI (m)
NDVI	28.79	6.434
EVI	29.37	8.688
NDWI	27.55	2.565
MNDWI	28.21	3.398
RGR	28.13	26.006
NRR	31.61	9.887
DI	25.98	3.453

The results show that in each case the extracted coastline presents an elongation, since AI values are always positive. The coastline that suffered less elongation is DI, with the lowest AI value (25.98 %), while NRR did provide the highest value (31.61 %). However, all the results provided by AI are very similar and no consistent difference can be seen between the coastlines.

In order to evaluate RI results the pixel size is used as the spatial unit for the accuracy assessment [42], which in this case is 10 m. All the results provided by RI are within 10 m, so they can be considered consistent with spatial resolution of the imagery. Even in this case EVI provided the worst results, labelling it as the least suitable of the indices proposed in this work for the coastline extraction. Very good results are achieved by the application of NDWI (2.565 m), as expected.

Above all we can say that the proposed approach based on unsupervised classification develops very good results.

The unsupervised classification with just two classes provides good results, as good as other techniques described in literature [12, 43, 44]. In particular, k-means clustering is very suitable when applied to the indices investigated in this article. Nevertheless, unsupervised classification allows to achieve good results faster than supervised classification, since training sites are not required.

V. Conclusion

This article aims to analyse the results that can be achieved with unsupervised classification for coastline extraction from Sentinel-2 imagery. The attention is focused on the outputs provided by k-means clustering applied on seven different indices, which can be seen as synthetic bands. To test the products of the unsupervised classification, two indices are used to evaluate the coastlines in terms of elongation (AI) and shift (RI) from the reference coastline.

While there is no reference to evaluate AI results, the values obtained with RI can be compared with the pixel size of the Sentinel-2 bands 2, 3, 4 and 8 (10 m): in these terms the results are very encouraging, since every coastline provided RI values less than 10 m.

Finally, Sentinel-2 images provide a good support, however as a future goal, these techniques require to be tested on other kind of satellite multispectral imagery, such as the PRISMAs recently released by the Italian Space Agency (ASI).

References

[1] L. Creel, “Ripple effects: Population and coastal regions” Washington, DC: Population reference bureau, 2003.

[2] F. D’Ascola, "Urbanizzazione in aree costiere. - Focus su acque e ambiente urbano”, ISPRA – Istituto Superiore per la Protezione e la Ricerca Ambientale, Roma, 2013, pp. 285.

[3] J. M. Wozencraft and W. J. Lillycrop, "JALBTCX coastal mapping for the USACE", The International hydrographic review, 2006.

[4] J. C. Brock and S. J. Purkis, "The emerging role of lidar remote sensing in coastal research and resource management", Journal of Coastal Research, vol. 10053, 2009, pp. 1-5. https://doi.org/10.2112/SI53-001.1

[5] M. Atik, "Environmental protection in coastal recreation sites in Antalya, Turkey" Coastal Management, vol. 38.6, 2010, pp. 598-616. https://doi.org/10.1080/08920753.2010.519433

[6] A. Sebastiani, E. Buonocore, P. P. Franzese, A. Riccio, E. Chianese, L. Nardella, F. Manes, “Modeling air quality regulation by green infrastructure in a Mediterranean coastal urban area: The removal of PM10 in the Metropolitan City of Naples (Italy)”, Ecological Modelling, vol. 440, 2021. https://doi.org/10.1016/j.ecolmodel.2020.109383

[7] S. S. Han, and Z. Yan, "China’s coastal cities: development, planning and challenges", Habitat international, vol. 23.2, 1999, pp. 217-229.

[8] S. Dellepiane, R. De Laurentiis, F. Giordano, “Coastline extraction from SAR images and a method for the evaluation of the coastline precision”, Pattern Recognition Letters, vol. 25.13, 2004, pp.1461-1470. https://doi.org/10.1016/j.patrec.2004.05.022

[9] S. J. Kim, I. P. Lee, Y. C. Kim, H. Cheong, “Coastline Extraction from Airborne LiDAR Data” In Proceedings of the Korean Society of Surveying, Geodesy, Photogrammetry, and Cartography Conference, 2006, pp. 457-462.

[10] A. Papakonstantinou, K. Topouzelis, G. Pavlogeorgatos, “Coastline zones identification and 3D coastal mapping using UAV spatial data” ISPRS International Journal of Geo-Information, vol. 5.6, 2016, pp. 75. https://doi.org/10.3390/ijgi5060075

[11] W. Nugraha, A. D. Parapat, D. S. Arum, F. Istighfarini, “GNSS RTK Application to Determine Coastline Case Study at Northen Area of Sulawesi and Gorontalo” In E3S Web of Conferences, vol. 94, 2019, pp. 1016. https://doi.org/10.1051/e3sconf/20199401016

[12] E. Alcaras, A. Errico, U. Falchi, C. Parente, A. Vallario, “Coastline Extraction from Optical Satellite Imagery and Accuracy Evaluation” In International Workshop on R3 in Geomatics: Research, Results and Review, Springer, Cham, October 2019, pp. 336-349. https://doi.org/10.1007/978-3-030-62800-0_26

[13] M. B. Giannini, P. Maglione, C. Parente, R. Santamaria, “Cartography and remote sensing for coastal erosion analysis”, WIT Transactions on Ecology and the Environment, vol. 149, 2011, pp. 65-76.

[14] M. Fabris, "Monitoring the Coastal Changes of the Po River Delta (Northern Italy) since 1911 Using Archival Cartography, Multi-Temporal Aerial Photogrammetry and LiDAR Data: Implications for Coastline Changes in 2100 AD", Remote Sensing, vol. 13.3, 2021, pp. 529. https://doi.org/10.3390/rs13030529

[15] D. Dominici, S. Zollini, M. Alicandro, F. Della Torre, P. M. Buscema, V. Baiocchi, “High resolution satellite images for instantaneous shoreline extraction using new enhancement algorithms”, Geosciences, vol. 9.3, 2019, pp. 123. https://doi.org/10.3390/geosciences9030123

[16] M. Specht, C. Specht, O. Lewicka, A. Makar, P. Burdziakowski, P. Dąbrowski, “Study on the Coastline Evolution in Sopot (2008–2018) Based on Landsat Satellite Imagery”, Journal of Marine Science and Engineering, vol. 8.6, 2020, pp. 464. https://doi.org/10.3390/jmse8060464

[17] P. Maglione, C. Parente, A. Vallario, "Coastline extraction using high resolution WorldView-2 satellite imagery", European Journal of Remote Sensing, vol. 47.1, 2014, pp. 685-699. https://doi.org/10.5721/EuJRS20144739

[18] M. F. Bruno, M.G. Molfetta, L. Pratola, M. Mossa, R. Nutricato, A. Morea, …, M. T. Chiaradia, “A combined approach of field data and earth observation for coastal risk assessment”, Sensors, vol. 19.6, 2019, pp. 1399. https://doi.org/10.3390/s19061399

[19] ESA, Missiones, Sentinel-2, 2021. https://sentinel.esa.int/web/sentinel/missions/sentinel-2#:~:text=It%20aims%20at%20monitoring%20variability,monitoring%20of%20Earth's%20surface%20changes. [Accessed on 28/05/2021]

[20] E. H. Boak, I. L. Turner, “Shoreline Definition and Detection: A Review”, Journal of Coastal Research, vol. 21.4, 2005, pp. 688-703. http://dx.doi.org/10.2112/03-0071.1

[21] A. M. Saeed, A. M. Fatima, "Coastline extraction using satellite imagery and image processing techniques", International Journal of Current Engineering and Technology , vol. 6.4, 2016.

[22] B. T. Pham, I. Prakash, "Application of simple remote sensing techniques for the detection and extraction of coastline-a case study of diu island, India", Indian Journal of Ecology, vol. 45.4, 2018, pp. 778-784.

[23] C. Dai, I. M. Howat, E. Larour, E. Husby, “Coastline extraction from repeat high resolution satellite imagery”, Remote Sensing of Environment, vol. 229, 2019, pp. 260-270. https://doi.org/10.1016/j.rse.2019.04.010

[24] J. Yin, F. He, "Researching the method of coastline extracted by remote sensing image", 2011 International Conference on Remote Sensing, Environment and Transportation Engineering, IEEE, 2011. https://doi.org/10.1109/RSETE.2011.5965052

[25] R. D. Viana, G. N. L. dos Reis, V. M. G: Velame, T. S. Körting, “Shoreline extraction using unsupervised classification on Sentinel-2 imagery”, Proceedings of XIX Brazilian Symposium on Remote Sensing, Vol 19, 2019.

[26] W. Liu, R. Qin, F. Su, "Weakly supervised classification of time-series of very high resolution remote sensing images by transfer learning", Remote Sensing Letters, vol. 10.7, 2019, pp. 689-698. https://doi.org/10.1080/2150704X.2019.1597295

[27] U. Falchi, “IT tools for the management of multi-Representation geographical information”, Int. J. Eng. Technol., vol. 7, 2017, pp. 65–69. https://doi.org/10.14419/ijet.v7i1.8810

[28] CopernicusEMS-Service_Overview_Brochure, https://emergency.copernicus.eu/mapping/sites/default/files/files/CopernicusEMS-Service_Overview_Brochure.pdf [Accessed on 23/06/21]

[29] SENTINEL-2 User Handbook, ESA, 2015. https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook [Accessed on 28/05/2021]

[30] G. Anfuso, J. A. M. Del Pozo, "Assessment of coastal vulnerability through the use of GIS tools in South Sicily (Italy)", Environmental Management, vol. 43.3, 2009, pp. 533-545. https://doi.org/10.1007/s00267-008-9238-8

[31] J. W. Rouse, R. H. Haas, J. A. Schell, D. W. Deering, “Monitoring vegetation systems in the Great Plains with ERTS”, NASA special publication, vol. 351, 1974, pp. 309.

[32] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, L. G. Ferreira, “Overview of the radiometric and biophysical performance of the MODIS vegetation indices”, Remote sensing of environment, vol. 83.1-2, 2002, pp. 195-213. https://doi.org/10.1016/S0034-4257(02)00096-2

[33] S. K. McFeeters, "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features", International journal of remote sensing, vol. 17.7, 1996, pp. 1425-1432. https://doi.org/10.1080/01431169608948714

[34] K. Mishra, P. Prasad, "Automatic extraction of water bodies from Landsat imagery using perceptron model", Journal of Computational Environmental Sciences, 2015. https://doi.org/10.1155/2015/903465

[35] J. P. Lacaux, Y. M. Tourre, C. Vignolles, J. A. Ndione, M. Lafaye, “Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal”, Remote Sensing of Environment, vol. 106.1, 2007, pp. 66-74. https://doi.org/10.1016/j.rse.2006.07.012

[36] V. Gond, E. Bartholomé, F. Ouattara, A. Nonguierma, L. Bado, “Surveillance et cartographie des plans d'eau et des zones humides et inondables en régions arides avec l'instrument VEGETATION embarqué sur SPOT-4”, International Journal of Remote Sensing, vol. 25.5, 2004, pp. 987-1004. https://doi.org/10.1080/0143116031000139908

[37] K. Herndon, R. Muench, E. Cherrington, R. Griffin, “An assessment of surface water detection methods for water resource management in the Nigerien Sahel”, Sensors, vol. 20.2, 2020, pp. 431. https://doi.org/10.3390/s20020431

[38] D. Kaur, A. Kaur, S. Gulati, M. Aggarwal, “A clustering algorithm for software fault prediction”, In 2010 International Conference on Computer and Communication Technology (ICCCT), IEEE, September 2010, pp. 603-607. https://doi.org/10.1109/ICCCT.2010.5640474

[39] J. Pérez-Ortega, N. N. Almanza-Ortega, A. Vega-Villalobos, R. Pazos-Rangel, C. Zavala-Díaz, A. Martínez-Rebollar, “The K-Means Algorithm Evolution” In Introduction to Data Science and Machine Learning, IntechOpen, 2019.

[40] L. Macaire, N. Vandenbroucke, J. G. Postaire, "Color image segmentation by analysis of subset connectedness and color homogeneity properties", Computer Vision and Image Understanding, vol. 102.1, 2006, pp. 105-116. https://doi.org/10.1016/j.cviu.2005.12.001

[41] Y. Zhang, X. Li, J. Zhang, D. Song, “A study on coastline extraction and its trend based on remote sensing image data mining”, In Abstract and Applied Analysis, Hindawi, vol. 2013, January 2013. https://doi.org/10.1155/2013/693194

[42] S. V. Stehman, J. D. Wickham, "Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment", Remote Sensing of Environment, vol. 115.12, 2011, pp. 3044-3055. https://doi.org/10.1016/j.rse.2011.06.007

[43] M. B. Giannini, C. Parente, "An object based approach for coastline extraction from Quickbird multispectral images", Int. J. Eng. Technol, vol. 6.6, 2015, pp. 2698-2704.

[44] D. Costantino, M. Pepe, G. Dardanelli, V. Baiocchi, “Using optical Satellite and aerial imagery for automatic coastline mapping”, Geographia Technica, vol. 15.2, 2020, pp. 171-190. https://doi.org/10.21163/GT_2020.152.17