



HAL
open science

Multi-temporal recognition of built-up area and land cover changes using machine learning approach in the Metropolis of Aix-Marseille-Provence in France

Sébastien Gadal, Mounir Oukhattar, Solomon Abody Otobo

► **To cite this version:**

Sébastien Gadal, Mounir Oukhattar, Solomon Abody Otobo. Multi-temporal recognition of built-up area and land cover changes using machine learning approach in the Metropolis of Aix-Marseille-Provence in France. Joint Urban Remote Sensing Event (JURSE) 2023, FORTH; RSLab; IEEE; ISPRS, May 2023, Heraklion, Greece. pp.1-4, 10.1109/JURSE57346.2023.10144184 . hal-04127519

HAL Id: hal-04127519

<https://hal.science/hal-04127519>

Submitted on 13 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Multi-temporal recognition of built-up area and land cover changes using machine learning approach in the Metropolis of Aix-Marseille-Provence in France

Sébastien Gadal

Aix-Marseille University, CNRS,
ESPACE UMR 7300, Univ. Nice
Sophia Antipolis, Avignon University
Aix-en-Provence, France
& North-Eastern Federal University,
Republic of Sakha-Yakutia
sebastien.gadal@univ-amu.fr

Mounir Oukhattar

Aix-Marseille University, CNRS,
ESPACE UMR 7300, Univ. Nice Sophia
Antipolis, Avignon University
Aix-en-Provence, France
mounir.oukhattar@etu.univ-amu.fr

Solomon Abody Otobo

Aix-Marseille University, CNRS, ESPACE
UMR 7300, Univ. Nice Sophia Antipolis,
Avignon University
Aix-en-Provence, France
solomon-abody.otobo@etu.univ-amu.fr

Abstract— Over the last thirty years, the extension of built-up areas has affected all areas of the Aix-Marseille-Provence (AMP) Metropolis. Urban sprawl has been particularly important in generating fragmented urban territories with various forms and spatial patterns. The modelling of the evolution of AMP is fundamental in the context of the implementation of climate plans for the horizon of 2050. It is based on spatial modelling by remote sensing of land cover change (LCC) between 1984 and 2021 using a classification approach that combines spectral transformations and indices applied to Landsat 5 TM and Sentinel 2 MSI.

The geo-simulations of the LCC and built-up areas dynamics by 2050 are modelled on Markov chain cellular automata. The future trends of AMP Metropolis are characterised by the evolution of built-up areas, estimated at 4,9% in 1984, 9,6% in 2021 and 11,2% in 2050 mainly to the loss of agricultural lands. The forests and semi-natural environments tend to be mainly more resilient to urban growth. The modelling of spatial dynamics of urbanisation is correlated with those of the evolution of the territorial distribution of populations. Geo-simulation of spatial dynamic changes is one of the decision-making planning tools for better management of the use of AMP territories.

Index Terms—Urbanisation, LCC, remote sensing, machine learning, geo-simulation, Aix-Marseille-Provence Metropolis.

I. INTRODUCTION

Urban growth and land cover changes (LCC) of metropolis in developing countries have drawn considerable attention from urban geographers, urbanists, and regional planners. Population increase and economic development are considered to be major factors directly inducing changes in LCC and urban dynamics. Population growth and economic development induced the construction of new residential, commercial, utility and transport infrastructures impacting the natural environments and agricultural areas. Remote sensing and geographic information system (GIS) have been widely used in recent decades to measure, model, and analyse the LCC [1]. Many land cover change methods and classifier algorithms have been developed for land cover mapping, change modelling, and urban recognition [2, 3, 4]. The methodology implemented on AMP Metropolis combines principal component analysis (PCA) factorial data transformation and machine learning minimum noise fraction (MNF), spectral indices and machine learning Cellular Automata-Markov Chain (CA-Markov) geo-simulation model. PCA and MNF and spectral indices are used to calculate the built-up areas. Spectral transformations and

indices such as MNF, PCA, built-up area index and bare land index have been implemented to extract the built-up areas [5, 6, 7]. Thus, the Markov chain was used to attempt to predict urban sprawl and LCC and their potential impacts on the environment. Markov chain analysis is a stochastic model to predict the spatial change characteristics of land covers based on historical and current geographic conditions [8, 9].

In the Aix-Marseille Provence metropolis, as early as 2000, the law on solidarity and urban renewal (SRU) introduced urban sprawl into the local urban plans: the territorial coherence scheme and local urban plan. The SRU calls for the economical and balanced use of natural, urban, peri-urban, and rural spaces [10]. It is designed for limiting the expansion of urban areas a general principle of sustainable urbanisation with the land possibility constraints of new diffuse build-ups and by restricting the use of minimum sizes for building plots [10]. The mapping of the land cover changes of the Aix-Marseille Provence (AMP) Metropolis is a requirement for planning and managing the dynamics of urbanisation, territorial, potential evolutions, and consequences. It is based on the recognition of artificial areas between 1984 and 2021 with Landsat 5 TM and Sentinel 2 MSI images. Remote sensing data were used based on existing geographic land use databases such as Corine Land Cover (CLC) and the land use pattern (MOS) of 2009 and 2017. The effectiveness of different classification methods has been tested for the analysis and decision-making supports as well as for the CA-Markov geo-simulation of Land covers and built-up dynamics to 2050.

II. MATERIALS AND METHODS

A. Study Area

The Aix-Marseille-Provence (AMP) Metropolis is the largest of the French metropolises territory, with a surface area of 3,173 km². It represents 10% of the surface area of the region SUD “Fig. 1”. It is twice as large as the metropolis of London and three times as large as Berlin or New York [10]. This territory is characterised by the juxtapositions of densely built-up urban areas and forest ridges, agricultural plains, valleys, and plateaux: 61% (1910 km²) of the surface area is occupied by natural areas, 15% (480 km²) is covered by urban areas, and 24% (760 km²) by agricultural areas. The territory of the AMP Metropolis largely overlaps the agglomerations of Marseille and Aix-en-Provence. The 92 districts of AMP Metropolis hosted 1603840 inhabitants in 1982, (density of 509,3 inhabitants/km²) and 1889666 inhabitants, (density of 600

inhabitants/km²) in 2019 “Fig. 2” [11]. AMP Metropolis concentrate 93% of the population of the Boûches-du-Rhône, and 37% of the region SUD. According to census projections, the population should exceed 2 million inhabitants in 2038 [10].

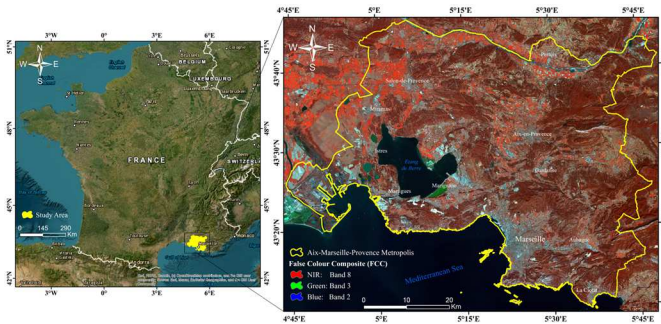


Fig. 1. Location of the study area (source data: USGS EROS Sentinel-2 MSI, Level-2A resampled to 10m; date: 30/03/2021).

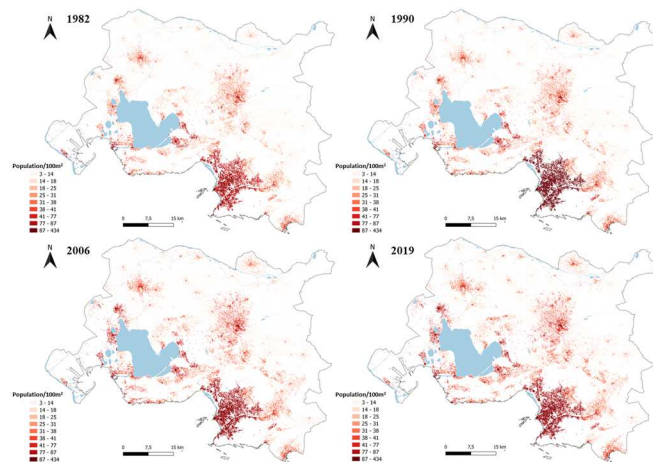


Fig. 2. Spatial distribution of the population between 1982 and 2019 using a 100m² grid that covers all buildings in the AMP Metropolis (Census: <https://www.insee.fr/fr/statistiques>).

B. Data Pre-Processing

Satellite images (Landsat 5 TM and Sentinel 2 MSI) were the main data used for the land cover change evolution and built-up. The choice of Sentinel 2 MSI was due to the high spatial resolution (10 meters resampled). The resampled spatial resolution of Landsat 8 OLI is 15 meters. The temporal dataset covering 37 years between 1984 and 2021, with a strict requirement for data consistency (similar acquisition date), optimal vegetation cover, and minimal cloud and snow cover, led to the selection of satellite images between March and April. The intensity of light received at the Earth's surface is seasonally at its highest during these months of the year. The Landsat 5 TM satellite images were taken on 9 April 1984, 10 April 1990 and 16 March 2010. Similarly, the Sentinel 2 MSI satellite images were taken on 10 April 2017 and 30 March 2021. Atmospheric (Radiometric Calibration) and radiometric (Dark Subtraction) corrections were applied to Landsat 5 TM and Sentinel 2 MSI data. For the Sentinel 2 MSI sensor, the RGB and NIR bands were stacked and then clipped to cover the study area. The Gram-Schmidt (GS) pan-sharpening algorithm was applied to resample the Sentinel 2 MSI images to 10 metres resolutions. Landsat 5 TM data (30 meters) were not resampled to 15 metres as they do not contain the panchromatic band.

C. Image processing

The image processing methodology developed combines image transformations based on PCA and MNF; spectral indices, ISODATA classification, and CA-Markov geo-simulation model. Indeed, the combination of the results of the Minimum Noise Fraction (MNF), Principal Component Analysis (PCA) by arithmetic fusion first; the Normalized Difference Vegetation Index (NDVI), Salinity Index (SI), Normalized Difference Built-up Index (NDBI), Normalized Difference Tillage Index (NDTI), and Built-up Index (BUI) fusion by fuzzy logic in the second, produce good discrimination of geographical objects as well as built-up areas. The results of these fusions are particularly good for urban and land use object recognition.

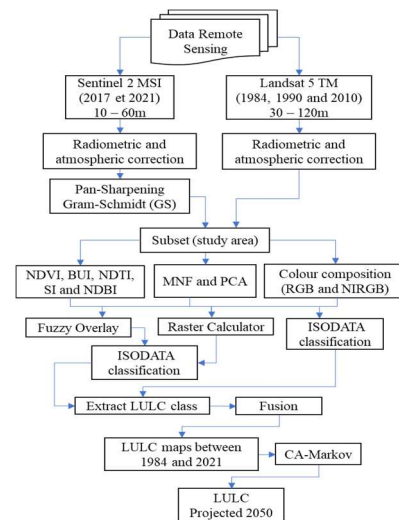


Fig. 3. Flow chart of the working methodology.

After testing and comparing other algorithms (Support Vector Machine, Maximum Likelihood and Multi-Layer Perceptron), the ISODATA classifier gave the best result for the classification of LCC at the scale AMP Metropolis. We performed classification based on the fusion image of colour composition (RGB and NIRGB), spectral indices (NDVI, NDTI, SI, NDBI and BUI), PCA and MNF. Similarly, we applied ISODATA on each colour composition, transformation and spectral index and then extracted the best mapping results from LCC. This processing step is necessary to extract and merge the best results to obtain an optimised image with a high classification accuracy “Fig. 3”. Indeed, NDVI is used to identify burnt areas, water surfaces, arable fields, and vegetation density. NDTI and SI are used to distinguish wetlands from built-up areas and arable land used for agriculture. NDBI and BUI are used mainly to recognise artificial territories, knowing that NDBI is an index that highlights urban areas where reflectance is higher in the shortwave-infrared (SWIR) spectrum than in the near-infrared (NIR) spectrum. In addition, to reduce data redundancy, the MNF and PCA allow the highlighting of urban morphology (road network, urban fabric, spatial distribution of urban areas) and geomorphological shapes of subsurface structures. The more specific recognition of each of these indices and transformations allows the ISODATA algorithm to easily separate the pixel values of each land cover unit. The accuracy of the land cover maps was evaluated and

validated using "natural" RGB colour composition images from the WORLDVIEW-3 satellite with a resolution of 30 cm by geo-visualisation on the Google Earth platform and by using the 2009 and 2017 MOS maps of the AMP Metropolis (crige-paca.org).

D. Prediction of future LCC dynamics

The Markov Cellular Automaton (CA-Markov) model was used to predict the LCC evolution of the AMP metropolis in 2050. CA-Markov is a probabilistic Change/Time Environment/Simulation model based on the LCC maps generated by the ISODATA classifier from 1984 to 2021. The LCC maps classified for 1984, 1990, 2010, 2017 and 2021 were used to acquire the dynamic change of LCC classes between 1984 and 2021, generating intermediate maps of potential LCC transitions and a matrix of transition probabilities from 1984 to 2021 and from 2021 to 2050. Thus, transition probabilities and

suitability zones are created by running the probabilistic Markov matrix before running the CA-Markov geo-simulation processing [12].

III. RESULTS AND DISCUSSIONS

The land cover maps of 1984, 1990, 2010, 2017 and 2021, obtained by the ISODATA unsupervised classifiers are presented in Fig. 4, together with their projection to 2050 obtained by the CA-Markov geo-simulation model. The maps show current and future territorial dynamics in five land cover classes according to those proposed by the Regional Centre for Geographic Information of Provence-Alpes-Côte d'Azur (CRIGE-PACA). The cartography of LCC classes includes artificial territories (red colour), agricultural territories (beige colour), water surfaces (electric blue colour), wetlands (light blue colour), and forests with semi-natural environments (green colour).

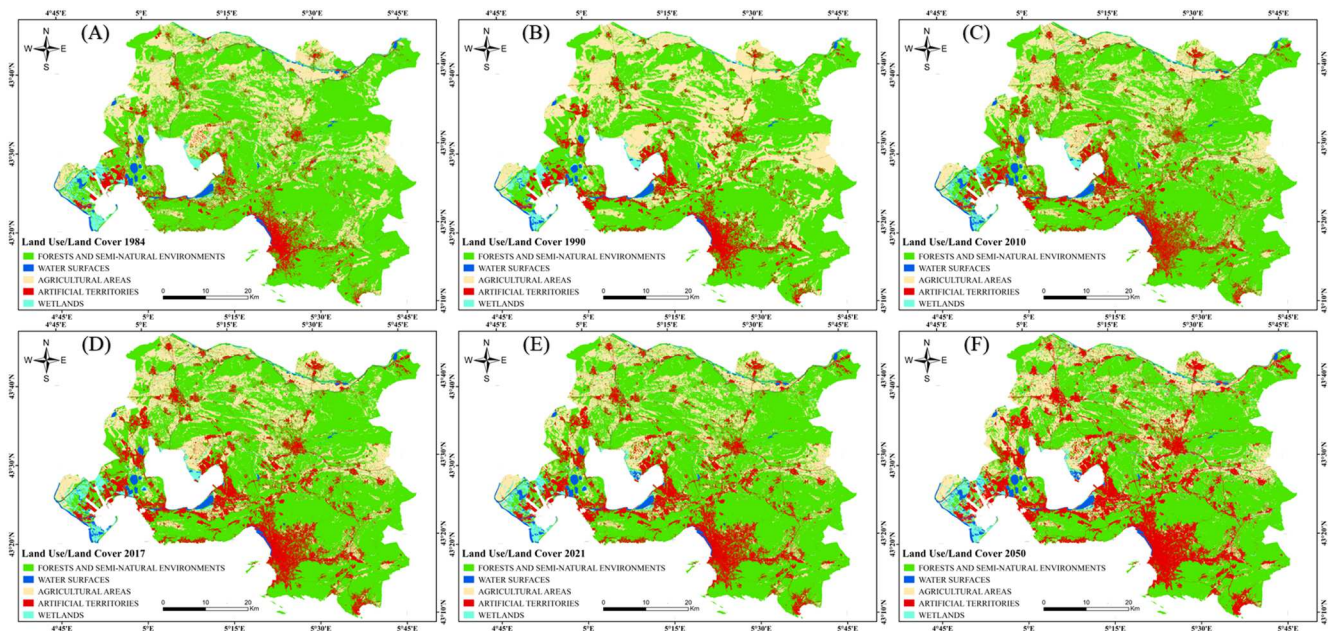


Fig. 4. Observed LCC maps for (A) 1984, (B) 1990, (C) 2000, (D) 2017, (E) 2021 and (F) 2050 predicted.

The multi-temporal statistical analysis of the LCC change classes of AMP Metropolis between 1984 and 2050 (Table I) illustrates in Fig.4. The statistical analysis results highlight the territorial dynamic driven by the urbanisation processes. These spatial and territorial dynamics will increase in the coming years according to the CA-Markov mode of geo-simulation. The spatial statistics generated with Landsat 5 TM and Sentinel 2 MSI show significant development of urban areas (4,9% to

11,2% of the AMP Metropolis area in 2050). The main changes are observed in transport networks (road and rail networks) and urban structures (spatial organisation, and densification of the urban fabric). This increase in built-up areas is accompanied by an increase in covered surfaces of forests and semi-natural environments and; a decrease in agricultural land (13,9% to 9,7%).

TABLE I
LCC TYPE AREA OF EACH TYPE OF LCC IN 1984, 1990, 2010, 2017, 2021 AND 2050

LCC Type	Land cover class surfaces											
	1984		1990		2010		2017		2021		2050	
	Km ²	%	Km ²	%	Km ²	%	Km ²	%	Km ²	%	Km ²	%
Artificial areas	252,5	4,9	298,4	5,8	363,2	7,0	486,2	9,4	498,3	9,6	581,7	11,2
Agricultural areas	713,2	13,9	869,8	16,9	660,7	12,8	609,5	11,8	554,7	10,7	503,1	9,7
Forests and semi-natural areas	2041,1	39,7	1878,6	36,5	1998,0	38,7	1929,8	37,3	1973,1	38,1	1934,7	37,2
Wetlands	93,5	1,8	63,9	1,2	76,2	1,5	81,7	1,6	80,0	1,5	80,5	1,5
Water surfaces	53,7	1,0	43,7	0,9	57,2	1,1	52,9	1,0	58,0	1,1	56,7	1,1

Referring to the statistics of the land cover unit "artificial areas" (Table I and Fig. 5), they have evolved from 4,9% (252,5 km²)

in 1984 to 9,6% (498,3 km²) in 2021 with a positive rate of variation of 197,3%. According to the projection for 2050, the

built-up areas of the AMP Metropolis would be 11,2% (581,7 km²) with a positive rate of variation of 116,7% between 2021 and 2050. The rapid urbanisation of the metropolis is linked to the demographic dynamics: external population, natural growth

of the population and the socioeconomic development of the metropolis. The urban space of the AMP Metropolis is currently expanding at a rate of 200 ha/year, or 2 km²/year [13].

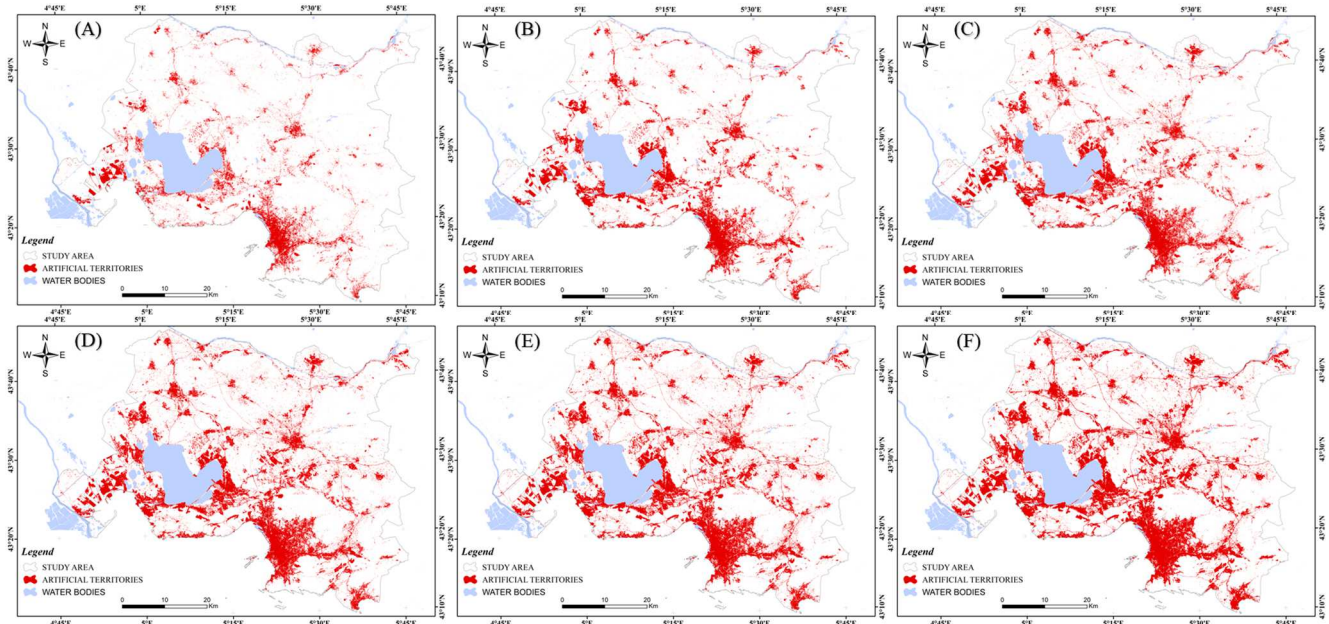


Fig. 5. Evolution of built-up areas between 1984 and 2050.

IV. CONCLUSIONS

The spatial modelling of the evolutions and changes of the artificial territories in the Aix-Marseille-Provence metropolis over the last 37 years is fundamental regarding the issue of space consumption due to the rapid processes of peri-urbanisation, urban fragmentation, with the increase of the population and the economic development. To analyse, map and model the dynamics and possible territorial transformations, a machine learning approach combined with spectral indexes and ISODATA classifier to identify and map the land covers was applied to Landsat 5 TM and Sentinel 2 MSI multispectral data. Land cover change cartographies were used to map and simulate LCC and the evolution of the built-up areas of the metropolis in 2050 at the regional scale. The analysis of LCC showed a clear continuous increase of the artificial territories from 252,5 km² in 1984, 298,4 km² in 1990, 363,2 km² in 2010, 486,2 km² in 2017, 498,3 km² in 2021 and 581,7 km² in 2050. This increase is partially resulting from the development of the motorway networks: more than half of the population and 2/3 of the positions in the AMP Metropolis are located outside the urban centres [10]. Moreover, the urban spatial expansion could be also explained by the economic development and by the growth of the population which went from 509,3 inhabitants/km² in 1982 to 602,9 inhabitants/km² in 2019 “Fig. 2”. According to the forecasts of the statistics of the urban planning agency of the Marseille agglomeration (AGAM) the population should exceed 2 million inhabitants in 2038 [11]. The analysis of land use dynamics is an essential prerequisite for urban studies, planning, and territorial schema of development, particularly in the context of urban growth and implementation of environmental planning standards relative to the climate master plans.

REFERENCES

- [1] A.H. Chughtai, H. Abbasi, I.R. Karas” “A review on change detection method and accuracy assessment for land use land cover”, *Remote Sensing Applications: Society and Environment*, Vol. (22), 2021, pp. 2352-9385.
- [2] J. Zhang, X. Zhang, X. Tan, X. Yuan “Extraction of Urban Built-Up Area Based on Deep Learning and Multi-Sources Data Fusion—The Application of an Emerging Technology in Urban Planning” *Land.*, Vol. 11, 2022, pp.1212.
- [3] S. Puttinaovarat, P. Horkaew “Urban areas extraction from multi-sensor data based on machine learning and data fusion” *Pattern Recognit. Image Anal.*, Vol. (27), 2017 pp.326–337.
- [4] M. Rahimzad, S. Homayouni, A. Alizadeh Naeini, S Nadi " An Efficient Multi-Sensor Remote Sensing Image Clustering in Urban Areas via Boosted Convolutional Autoencoder (BCAE)” *Remote Sensing.*, vol. 13(13), 2021, pp.2501.
- [5] U. A. M. E. Ali, M. A. Hossain, M. R. Islam, "Analysis of PCA Based Feature Extraction Methods for Classification of Hyperspectral Image” *2nd International Conference on Innovation in Engineering and Technology (ICIET)*, 2019, pp. 1-6.
- [6] Y. Zha, J. Gao, S. Ni “Use of normalized difference built-up index in automatically mapping urban areas from TM imagery” *International Journal of Remote Sensing.*, Vol. 24(3), 2003, pp.583–594.
- [7] Y. Zhou, G. Yang, S. Wang, L. Wang, F. Wang, X. Liu “A new index for mapping built-up and bare land areas from Landsat-8 OLI data” *Remote Sensing Letters.*, Vol. 5(10), 2014, pp.862–871.
- [8] M. Sharjeel, A. Zahir, H. Mateeul, M.G. Badar “Monitoring and Predicting Land use/Landcover Change Using an Integrated Markov Chain & Multilayer Perceptron Models: A Case Study of Sahiwal Tehsil” *Journal of GeoSpace Science.*, Vol. 1(2), 2016, pp. 43-59.
- [9] K. Hazhi, J. Javad, K. Jabbar, A. Parisa “Monitoring and prediction of land use/land cover changes using CA-Markov model: a case study of Ravansar County in Iran” *Arabian Journal of Geosciences.*, Vol 11, 2018, pp. 592.
- [10] Urban planning agency of the Marseilles agglomeration (AGAM), “space consumption Best trajectory “, *Regards Environnements.*, December 2018, N°80.
- [11] INSEE (2019) “France, portrait social “, *INSEE publication*, 292 p.
- [12] E. Gidey, O. Dikinya, R. Sebebo “Cellular automata and Markov Chain (CA-Markov) model-based predictions of future land use and land cover scenarios (2015–2033) in Raya, northern Ethiopia”, *Model. Earth Syst. Environ* Vol. 3, 2017, pp.1245– 1262.