Journal of Geography and Cartography 2025, 8(1), 10230.
https://doi.org/10.24294/jgc10230

Article

Spatial analysis and classification of land use patterns in Lucknow district, UP, India using GIS and random forest approach

Munizzah Salim[1,*], Shubham Bhattacharjee[1], Neerav Sharma[2], Kavita Sharma[3], Rahul Dev Garg[1]

[1] Department of Civil Engineering, IIT Roorkee, Roorkee 247667, India
[2] Purdue University, IN 47907, United States
[3] Pt. Ravishankar Shukla University, Raipur 492010, India
*** Corresponding author:** Munizzah Salim, munizzah_s@ce.iitr.ac.in

CITATION

Salim M, Bhattacharjee S, Sharma N et al. Spatial analysis and classification of land use patterns in Lucknow district, UP, India using GIS and random forest approach. Journal of Geography and Cartography. 2025; 8(1): 10230. https://doi.org/10.24294/jgc10230

ARTICLE INFO

Received: 19 November 2024
Accepted: 23 December 2024
Available online: 16 January 2025

Abstract: Mapping land use and land cover (LULC) is essential for comprehending changes in the environment and promoting sustainable planning. To achieve accurate and effective LULC mapping, this work investigates the integration of Geographic Information Systems (GIS) with Machine Learning (ML) methodology. Different types of land covers in the Lucknow district were classified using the Random Forest (RF) algorithm and Landsat satellite images. Since the research area consists of a variety of landforms, there are issues with classification accuracy. These challenges are met by combining supplementary data into the GIS framework and adjusting algorithm parameters like selection of cloud free images and homogeneous training samples. The result demonstrates a net increase of 484.59 km^2 in built-up areas. A net decrement of 75.44 km^2 was observed in forest areas. A drastic net decrease of 674.52 km^2 was observed for wetlands. Most of the wastelands have been converted into urban areas and agricultural land based on their suitability with settlements or crops. The classifications achieved an overall accuracy near 90%. This strategy provides a reliable way to track changes in land cover, supporting resource management, urban planning, and environmental preservation. The results highlight how sophisticated computational methods can enhance the accuracy of LULC evaluations.

Keywords: LULC; GIS; ML; spatial analysis; urban planning

1. Introduction

Mapping land use and cover (LULC) is essential to sustainable development, urban planning, and environmental management. Understanding the effects of human activity and natural processes on the environment requires knowledge of the geographical distribution and changes in land use and cover, which is provided by this information [1]. Traditionally, satellite and aerial photo interpretation was done by hand, which took a lot of time and was prone to human mistakes while doing LULC mapping [2]. However, the accuracy, effectiveness, and automation of LULC mapping have been greatly improved by developments in GIS and ML [3].

The integration of various datasets, including topographic maps, socioeconomic data, and remote sensing imagery, is made possible by GIS, which is essential for organizing, interpreting, and visualizing spatial data [4]. Automated pattern recognition and classification in LULC mapping have been possible because of the combination of ML and GIS [5]. The complexity and variability of LULC data have been successfully handled by machine learning, in particular, supervised learning algorithms like Random Forest [6–8], Support Vector Machines [9–12], and Neural Networks [13,14], which provide high classification accuracy (> 85%) even in heterogeneous landscapes [15,16].

The study of high-resolution satellite images, such as data from Landsat, Sentinel, and SPOT, IKONOS satellites, demonstrates the synergy between GIS and machine learning in particular [17–19]. These datasets offer comprehensive data on land cover, but to fully realize their potential, their complexity necessitates the use of sophisticated analytical methods. When combined with GIS, machine learning techniques with its automation, scalability, generalization and adaptiveness allow for important patterns to be extracted from high-dimensional data, resulting in LULC maps that are more precise and comprehensive [20].

Recent research has shown how well GIS and machine learning work together for LULC mapping in a variety of contexts. For instance, Myint et al. [21] mapped Phoenix, Arizona's urban land cover with great precision using object-based image analysis within a GIS framework. Similarly, Akar and Gungor [22] demonstrated the value of machine learning in managing complex landscapes by classifying land cover in Turkey using Support Vector Machines in a GIS setting. Maxwell et al. [23] showed the algorithm's resilience in handling varied land cover types in a different study where they mapped the forest cover in the Brazilian Amazon using Random Forests in conjunction with GIS.

Notwithstanding these developments, there are still difficulties in using machine learning and GIS for LULC mapping. The quality of the input data (data clarity, high resolution, good temporal extent) and the choice of suitable ML methods have an impact on the accuracy of LULC maps [24]. Furthermore, complex data fusion techniques are needed to maximize the information retrieved from each source when merging multi-source data into GIS, such as topographic, radar, and optical imaging [25]. More research is required to obtain more scalability, efficiency and further improve the potential of LULC mapping using GIS and machine learning.

With an emphasis on approaches, case examples, and difficulties, this study examines the state of LULC mapping using GIS and the Random Forest model now. Recent breakthroughs in remote sensing technologies, particularly with satellite imagery and Geographic Information Systems (GIS), have revolutionized the ability to monitor and analyze LULC changes at global, regional, and local scales. These innovations have enabled more precise mapping of land cover types, identification of land use patterns, and detection of environmental changes such as deforestation and urban expansion [26].

2. Study area

Lucknow Metropolis (**Figure 1**) lies between the coordinates of 26°30′ N to 27°10′ N latitudes and 80°30′ E to 81°13′ E longitudes. It is the capital city of India's most populous state Uttar Pradesh. Lucknow is situated in the middle of the Gangetic Plain and spreads on the banks of the river Gomati, a left-bank tributary of river Ganga. The height of Lucknow city above mean sea level is 123 m. The total land area of Lucknow city is 310 Sq. km. Lucknow has an extensive network of roads and railways and has grown all around in a radius of 25 km. Half of the rainfall occurs from June to October when the city gets an average rainfall of 896.2 mm (35.28 in) from the southwest monsoon winds, and occasionally frontal rainfall from the northeast monsoon will occur in January. Lucknow district is a densely populated district of Uttar Pradesh that witnessed remarkable expansion, growth, and development activities such as significant building construction, construction of highways, etc. Such a rapid increase in land consumption and modifications on land use and land cover changes need to be addressed through spatiotemporal dynamics of various LULC classes.

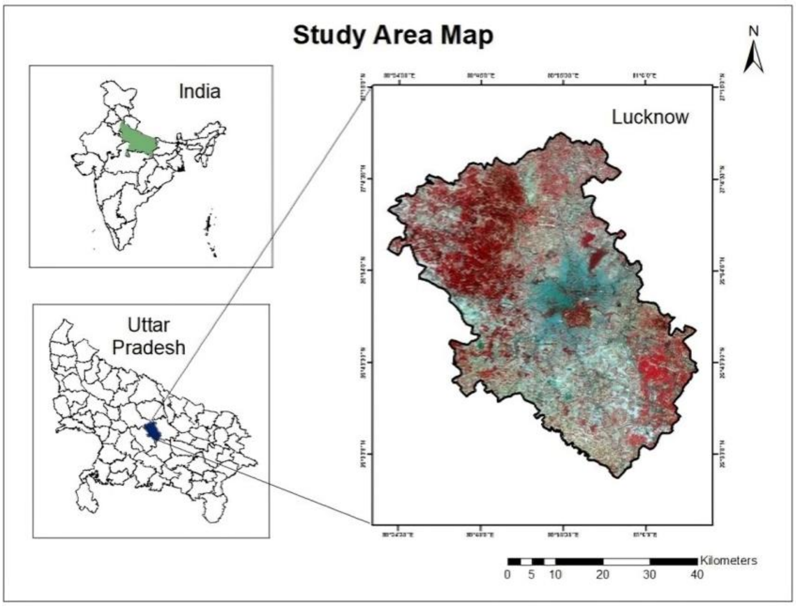

Figure 1. Location map showing the study area.

3. Data used and methodology opted

Cloud-free Landsat series (**Figure 2**) of datasets (resolution ~ 30 m) for the years 2004 (TM sensor) and 2024 (OLI sensor) have been deployed for the present study. Image classification was performed on the Google Earth Engine (GEE) platform. ArcMap was used for data visualization and preparation of maps.

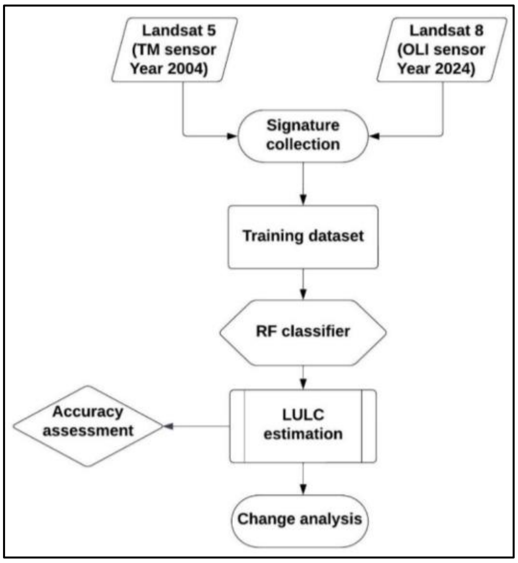

Figure 2. Methodological flowchart for the present study.

The study area boundary and respective cloud free and Top of Atmosphere (TOA) corrected satellite images were imported into the GEE console. Designated image bands (NIR-Red-Green) were assigned to the R-G-B color code for obtaining the fine-scale False color composite (FCC). Signatures corresponding to different land cover classes viz. built-up areas, forests, agriculture, water bodies, and wasteland were collected carefully to prepare the training samples for further use. Training samples on the pure pixels distributed throughout the imagery were selected to maintain class homogeneity. To obtain a singleton set of classes, the different identifiers of the class samples were merged into one. The signatures were then trained respectively for classifying the distinct classes through the satellite imagery.

An ML-based Random forest (RF) classifier was used to classify the imagery into different classes mentioned above based on the collected training samples. Being a bagging algorithm, RF depicts low prediction error for better accuracy. RF classifier works with multiple decision trees (DT). It reduces the variance of the individual DT through random selection. The prediction of a target variable (classes) was made with the usability of the maximum vote provided by each decision tree for every image pixel. To depict the correctness of the classification results, an accuracy assessment through confusion matrix carrying users and producers accuracy was made using Google Earth images. Areas for different classes were computed and transitions in the LULC classes were obtained between the years 2004 and 2024.

4. Results and discussions

The training samples collected were based on the land distribution of the study area, and visual comparison of the natural and false color composite images. There is a considerable amount of change between LULC classes for the years 2004 and 2024. Spatiotemporal changes are depicted in **Figure 3** and **Table 1**.

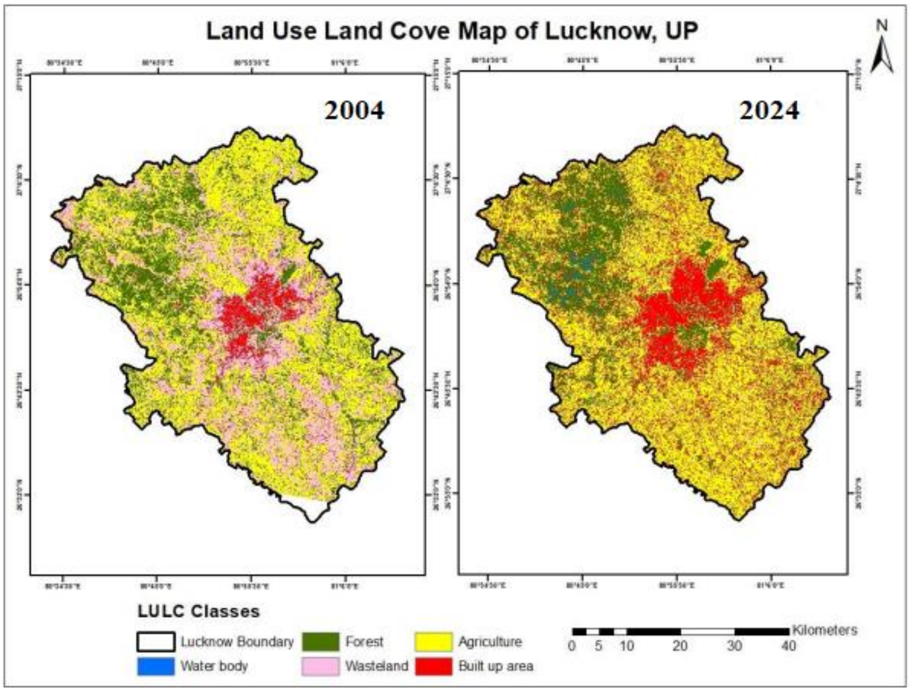

Figure 3. Spatiotemporal LULC dynamics in lucknow.

Table 1. Area changes in different LULC classes.

Classes	Area in 2004 (in km^2)	Area in 2024 (in km^2)
Built up area	157.48	642.07
Forest	611.16	535.72
Agriculture	1035.67	1313.60
Water bodies	23.58	22.02
Wasteland	698.84	24.32

Built-up area (**Figure 4**) increases from 157.48 km^2 to 642.07 km^2 (net increase ~ 484.59 km^2). This is because of rapid urbanization i.e. transmission of rural to urban in Lucknow. Many wastelands were transformed into settlements. Transforming wastelands into state and municipal ownership partially addresses the issue of limited space for high-rise buildings in urban areas, particularly through infill development. The growth was prominently seen over the eastward and southward sides of Lucknow. Growth around the Central Business district is also visible.

Figure 4. Google earth snapshot showing built-up areas of lucknow.

Forest cover (**Figure 5**) decreases from 611.16 km^2 to 535.72 km^2 (net decrease ~ 75.44 km^2). This is because many forested areas are converted into agricultural areas as well and deforestation is becoming a leading problem in the country [27]. So many forests were deforested and converted into built-up areas for settlement zones as well. Clearing forests results in the destruction of habitats for numerous plant and animal species, many of which are at risk of extinction. Additionally, deforestation disrupts the carbon cycle, as trees are essential in absorbing carbon dioxide, a key greenhouse gas. In the absence of trees, carbon is released into the atmosphere, further driving climate change. For instance the forested patch near Barkhurdarpur was degraded significantly throughout the study period.

Figure 5. Google earth snapshot showing forested areas of lucknow.

Agricultural area (**Figure 6**) increases from 1035.67 km^2 to 1313.60 km^2 (net increase ~ 277.93 km^2). This is because many forested areas were converted into agricultural areas as well as many fallow lands have been converted into cropped and matured cropped areas. Many wastelands based on their land suitability for agriculture have also been converted into agricultural areas. In the Misripur area forested patch was converted into agricultural area.

Figure 6. Google earth snapshot showing agricultural areas of lucknow.

The water body is not much affected throughout these 20 years. Forests and settlements cover most of the periphery areas of rivers and water bodies. Gomati River (**Figure 7**) crosses Lucknow. There are some other water bodies in Lucknow such as Kathauta Jheel.

Figure 7. Google earth snapshot showing the gomati river crossing lucknow.

Wasteland shows a major variation in Lucknow as they decrease from 698.84 km^2 to 24.32 km^2 (net decrease ~ 674.52 km^2). Most of the wastelands have been converted into urban areas (**Figure 8**) and agricultural land based on their suitability with settlements or crops. In 2004 generally, wastelands surrounded urban areas whereas in the year 2024, most of them have been converted into built-up areas.

Figure 8. Google earth snapshot showing wasteland converting to built-up areas.

Accuracy assessment was performed for the classified images for both years and confusion matrices were generated using the classified and reference points (collected from high-resolution Google Earth imagery). The confusion matrices with several training samples generated for both years depict classification accuracies of 89 % for the year 2004 (**Table 2**) and 90 % for the year 2024 (**Table 3**).

Table 2. Confusion matrix for the year 2004 (G.T signifies ground truth).

Classes	**Ref 1**	**Ref 2**	**Ref 3**	**Ref 4**	**Ref 5**	**G.T**
Built-up	80	0	0	3	0	80
Forest	0	45	0	7	0	54
Agriculture	0	6	32	0	0	38
Waterbody	0	0	4	51	0	53
Wasteland	9	0	0	2	51	65
Total	89	51	36	63	51	290

Table 3. Confusion matrix for the year 2024.

Classes	**Ref 1**	**Ref 2**	**Ref 3**	**Ref 4**	**Ref 5**	**G.T**
Built-up	72	3	0	0	0	71
Forest	0	60	6	2	0	69
Agriculture	0	6	38	0	0	44
Waterbody	3	0	1	56	0	60
Wasteland	6	9	0	0	61	78
Total	81	78	44	58	61	322

To reduce deforestation, effective strategies include fostering reforestation efforts, enforcing stricter land-use regulations, and advocating for sustainable forestry practices. Managing urban sprawl requires the implementation of smart growth initiatives, encouraging denser development, and enhancing public transportation systems to minimize the need for widespread urban expansion. In the realm of sustainable agriculture, promoting agroecology, diversifying crops, and adopting organic farming practices can help maintain soil health and biodiversity while decreasing reliance on harmful chemicals. By combining these approaches, we can achieve a more harmonious balance between environmental preservation and urban development [28].

The Random Forest classifier is a highly effective and commonly used machine learning algorithm, but it does have certain limitations. One significant issue is its potential to become computationally demanding and slow, especially with large datasets or when a high number of trees are included in the forest. Furthermore, due to the complexity, interpreting Random Forest models can be difficult. There is also a risk of overfitting if the model is not properly tuned, particularly when the number of trees or the depth of the trees is excessively large. These issues can be addressed through hybrid ML models and improved algorithms [29,30] which will fulfil the large computation time gap and the problems of overfitting.

5. Conclusions

This study aimed to identify and analyze general trends in LULC Changes that have taken place in Lucknow district over 20 years using Landsat satellite imagery and ML-based image classification in the GEE platform. The key findings of this study revealed that the major LULC classes of Lucknow district identified include agriculture, forest, wasteland, built-up areas, and water. Significant building construction and deforestation were major drives of LULC dynamics in the Lucknow district over the study period from 2004 to 2024. This study showed a continuous decrease in wasteland areas in the district. Consequently, the urban built-up area and the agriculture area increased. Some future insights could be the integration of IoT for smart technology based embedded infrastructure, waste management etc and urban growth prediction modelling for decision and policymaking. Some other recommendations could be the relationship between urban growth and urban heat islands etc. Such study is required to support environmental policy; physical planning purposes, sustainable land use, and land development.

Author contributions: Conceptualization, MS, SB, and NS; methodology, MS, SB and NS; software, MS and SB; validation, MS and SB; formal analysis, MS, SB and NS; investigation, MS and NS; resources, SB; data curation, MS, SB and NS; writing—original draft preparation, MS, SB and NS; writing—review and editing, MS, SB, NS, KS and RDG; visualization, MS; supervision, KS and RDG; project administration, KS and RDG. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Lambin EF, Turner BL, Geist HJ, et al. The causes of land-use and land-cover change: Moving beyond the myths. Global Environmental Change. 2001; 11(4): 261–269.
2. Foody GM. Status of land cover classification accuracy assessment. Remote Sensing of Environment. 2002; 80(1): 185–201.
3. Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in remote sensing: an applied review. International Journal of Remote Sensing. 2018; 39(9): 2784–2817. doi: 10.1080/01431161.2018.1433343

4. Longley PA, Goodchild MF, Maguire DJ, et al. Geographical Information Systems: Principles, Techniques, Applications, and Management. Wiley; 2005.
5. Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing. 2011; 66(3): 247–259. doi: 10.1016/j.isprsjprs.2010.11.001
6. Breiman L. Random forests. Machine Learning. 2001; 45(1): 5–32.
7. Svoboda J, Štych P, Laštovička J, et al. Random Forest Classification of Land Use, Land-Use Change and Forestry (LULUCF) Using Sentinel-2 Data—A Case Study of Czechia. Remote Sensing. 2022; 14(5): 1189. doi: 10.3390/rs14051189
8. Tikuye BG, Rusnak M, Manjunatha BR, et al. Land Use and Land Cover Change Detection Using the Random Forest Approach: The Case of The Upper Blue Nile River Basin, Ethiopia. Global Challenges. 2023; 7(10). doi: 10.1002/gch2.202300155
9. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20(3): 273–297. doi: 10.1007/bf00994018
10. Ul Din S, Mak HWL. Retrieval of Land-Use/Land Cover Change (LUCC) Maps and Urban Expansion Dynamics of Hyderabad, Pakistan via Landsat Datasets and Support Vector Machine Framework. Remote Sensing. 2021; 13(16): 3337. doi: 10.3390/rs13163337
11. Shi D, Yang X. Support Vector Machines for Land Cover Mapping from Remote Sensor Imagery. In: Li J, Yang X. (editors) Monitoring and Modeling of Global Changes: A Geomatics Perspective. Springer Remote Sensing/Photogrammetry. Springer, Dordrecht; 2015.
12. Waske B, Benediktsson JA. Fusion of Support Vector Machines for Classification of Multisensor Data. IEEE Transactions on Geoscience and Remote Sensing. 2007; 45(12): 3858–3866. doi: 10.1109/tgrs.2007.898446
13. Kavzoglu T, Mather PM. The use of backpropagating artificial neural networks in land cover classification. International Journal of Remote Sensing. 2003; 24(23): 4907–4938. doi: 10.1080/0143116031000114851
14. Sisay G, Gessesse B, Fürst C, et al. Modeling of land use/land cover dynamics using artificial neural network and cellular automata Markov chain algorithms in Goang watershed, Ethiopia. Heliyon. 2023; 9(9): e20088. doi: 10.1016/j.heliyon.2023.e20088
15. Pal M, Mather PM. Support vector machines for classification in remote sensing. International Journal of Remote Sensing. 2005; 26(5): 1007–1011. doi: 10.1080/01431160512331314083
16. Shao Y, Lunetta RS. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS Journal of Photogrammetry and Remote Sensing. 2012; 70: 78–87. doi: 10.1016/j.isprsjprs.2012.04.001
17. Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing. 2007; 28(5): 823–870. doi: 10.1080/01431160600746456
18. Muttitanon W, Tripathi NK. Land use/land cover changes in the coastal zone of Ban Don Bay, Thailand using Landsat 5 TM data. International Journal of Remote Sensing. 2008; 26. doi: 10.1080/0143116051233132666
19. Zhou W, Huang G, Troy A, et al. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: A comparison study. Remote Sensing of Environment. 2009; 113(8): 1769–1777. doi: 10.1016/j.rse.2009.04.007
20. Li J, Shao G. Big data in smart forest management. Forest Ecology and Management. 2014; 331: 1–3.7
21. Myint SW, Gober P, Brazel A, et al. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sensing of Environment. 2011; 115(5): 1145–1161. doi: 10.1016/j.rse.2010.12.017
22. Akar Ö, Gungor O. Integrating GIS and Support Vector Machines for Land Use and Land Cover Classification. Journal of Applied Remote Sensing. 2012; 6(1): 063512.
23. Maxwell AE, Strager MP, Warner TA, et al. Comparison of NAIP orthophotography and RapidEye satellite imagery for mapping of forest disturbances in the southeastern United States. Forest Ecology and Management. 2016; 378: 222–236.
24. Friedl MA, McIver DK, Hodges JCF, et al. Global land cover mapping from MODIS: Algorithms and early results. Remote Sensing of Environment. 2002; 83(1–2): 287–302.
25. Gong P, Howarth PJ. Frequency-based contextual classification and grey-level vector reduction for land-use identification. Photogrammetric Engineering and Remote Sensing. 1992; 58(4): 423–437.
26. Turner BL, Meyer WB. Land use and land cover in global environmental change: considerations for study. International Social Science Journal. 1991; 43: 669–679.
27. Forest Survey of India. State of Forest Report, ISFR. Ministry of Environment, Forest and Climate Change. 2021.

28. Weng Q. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. Remote Sensing of Environment. 2012; 117: 34–49. doi: 10.1016/j.rse.2011.02.030
29. Rey SJ, Anselin L. PySAL: A Python library of spatial analytical methods. In: Handbook of Applied Spatial Analysis. Springer Nature; 2010. pp. 175–193.
30. Friedl MA, Sulla-Menashe D, Tan B, et al. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment. 2010; 114(1): 168–182. doi: 10.1016/j.rse.2009.08.016