



Land Cover Classification Using ICESat/GLAS Full Waveform Data

S. Ghosh¹ · S. Nandy¹ · S. Patra¹ · S. P. S. Kushwaha¹ · A. Senthil Kumar¹ · V. K. Dadhwal²

Received: 17 October 2015 / Accepted: 30 May 2016 / Published online: 20 June 2016
© Indian Society of Remote Sensing 2016

Abstract In the present study, parameters derived from Ice, Cloud, and land Elevation Satellite/Geoscience Laser Altimeter System (GLAS) full waveform were used for land cover classification in western part of Doon valley, Uttarakhand, India. Three parameters, viz, *height*, front slope angle (*afslope*) and canopy return ratio (*rCanopy*) were extracted from the returned full waveform signals. k-means (KM), partitioning around medoids (PAM), and fuzzy c-means (FCM) with different cluster sizes were used for classifying the land cover types with the help of GLAS-derived parameters. Among the clustering methods, KM performed the best. The overall accuracy (89.41 %) of all methods were quite significant with cluster size three i.e. with three classes forest, mango orchard and other class including agriculture, barren/fallow land, settlement, dry river bed, etc. The accuracy of the PAM (60 %) and the FCM (68.4 %) decreased drastically at four clusters with the separation of agriculture from barren/fallow land. The accuracy of the PAM and the FCM further decreased with increase in the number of clusters whereas KM showed reliable results for all clusters. KM with five clusters was able to distinguish five different land covers, viz, forest, mango orchard, agriculture and barren/fallow land and other class including settlement, dry river bed, etc. with an overall classification accuracy of 72.93 %. The study presents a method for classifying land cover types using GLAS full waveform data.

Keywords ICESat/GLAS full waveform · Land cover classification · k-means · Partitioning around medoids · Fuzzy c-means

Introduction

Land cover classification is widely used for land management and decision support system (Bartholome and Belward 2005; Nandy and Kushwaha 2011). Land cover is influenced and altered by anthropogenic activities and/or climatic factors. A variety of remotely sensed data sources (including multispectral, hyperspectral, microwave, and light detection and ranging (LiDAR) data) and methods have been employed to derive land-cover information. Both multispectral and hyperspectral data can provide only spectral information of an object and may fail to separate objects that are spectrally similar but structurally different (Buddenbaum et al. 2013), whereas LiDAR can provide structural information of the object (Duong 2010).

The Geoscience Laser Altimeter System (GLAS) on-board Ice, Cloud, and land Elevation Satellite (ICESat) provided full waveform data from 2003 to 2009 with near global coverage between $\pm 86^\circ$ latitude (Rosette et al. 2008). The elliptical footprint diameter of the ICESat/GLAS data varies between 51 and 102 m equivalent to an average circular diameter of approximately 70 m and are spaced 172 m apart along tracks (Abshire et al. 2005). The GLAS systematically samples the energy returned from the surface of the ground (Harding and Carabajal 2005). The returned energy (counts or volts) from the earth surface is recorded against time (time/bins). This energy profile is known as full waveform which gathers vital information of the intercepted surfaces using time delay technique. The full waveform data gives the structural information on the

✉ S. Nandy
subrato.nandy@gmail.com

¹ Indian Institute of Remote Sensing, ISRO, Dehradun 248001, India

² National Remote Sensing Centre, ISRO, Balanagar, Hyderabad 500625, India

vertical distribution of the surface (Duong 2010). However, the shapes of the returned GLAS waveforms depend on terrain characteristics (roughness and slope) of the surface reflectance at 1064 nm wavelength, cloud cover, size, shape and orientation of the footprint, canopy thickness, vegetation height, etc. (Neuenschwander 2008).

ICESat/GLAS data products are mainly used for ice sheet elevation estimation, canopy structure analysis, and study of cloud and atmospheric properties (Zwally et al. 2002; Lefsky et al. 2005; Andersen et al. 2015). The full waveform data of GLAS provides new opportunities for land cover classification. Few studies on land cover classification have been carried out in the past decade. These studies used decision tree approach (Duong et al. 2006), curve matching method (Zhou et al. 2015) and machine learning techniques (Liu et al. 2015) to classify land cover types with an overall accuracy of around 73–87 %. Liu et al. (2015) also used multi-sensor (ICESat/GLAS and Landsat Thematic Mapper (TM)/Enhanced TM Plus (ETM+)) integration approach for land cover classification and achieved 91 % overall accuracy.

The present study aims to classify different land cover types using ICESat/GLAS full waveform data. Different land cover types were identified based on three waveform derived parameters, viz, *height*, front slope angle (*afslope*) and canopy return ratio (*rCanopy*) using k-means (KM), partitioning around medoids (PAM), and fuzzy c-means (FCM).

Materials and Methods

Study Area

The area selected for the present study lies in the western part of Doon valley (29.7–30.7°N and 77.4–78.2°E), Uttarakhand, India (Fig. 1.). The forest of the area can be broadly classified as Tropical Moist Deciduous Forest (Champion and Seth 1968) dominated by *Shorea robusta* and its associates. Major forest tree species are *S. robusta*, *Mallotus philippensis*, *Terminalia tomentosa*, *Ehretia laevis*, *Lagerstroemia parviflora* and *Tectona grandis* plantations. The area has a predominantly sub-tropical monsoonal climate with temperature ranging from 2 to 40 °C. The average annual rainfall is about 2000 mm. The abundance of different land covers, accessibility, and data availability acted as the guiding factors for the choice of the study area.

Data and Methodology

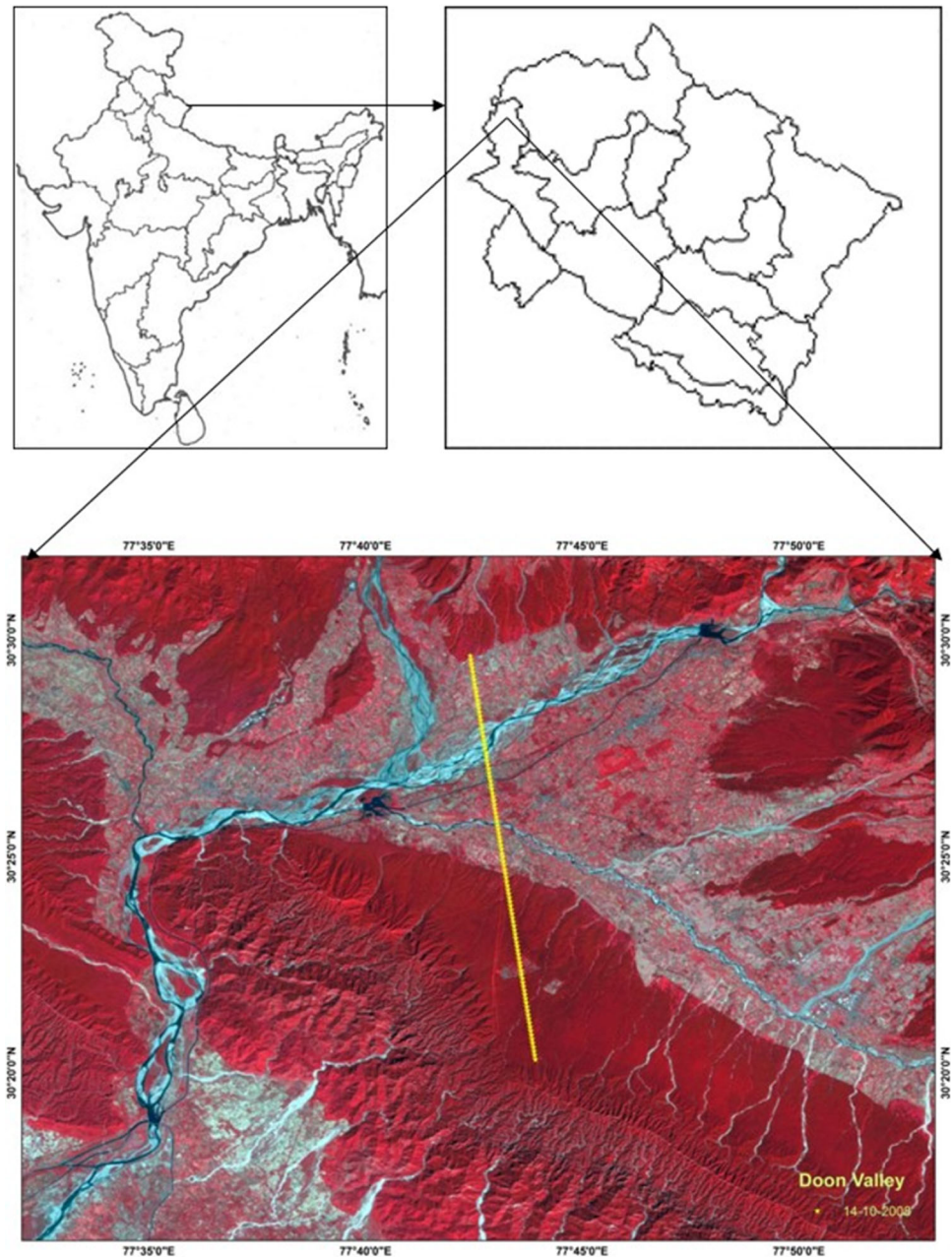
ICESat/GLAS full waveform datasets were used in this study. The GLA01 and GLA14 of version 33 datasets were procured from National Snow and Ice Data Centre

(NSIDC, <http://nsidc.org/data/icesat>) (Zwally et al. 2011). GLA01 is global altimetry data containing transmitted and received waveforms and corresponding sensor gains, whereas GLA14 contains precise geo-location of the footprint centre. GLA01 and GLA14 products are linked by Coordinated Universal Time and shot number. In the present study, eighty-five footprints covering different land cover types were selected. The footprints which have less than ten degree ground slope were considered, as the ground slope is a crucial factor for GLAS data processing. Shuttle Radar Topography Mission (SRTM) digital elevation model was used for determining the ground slope.

The GLAS full waveform (from GLA01) is assumed to be a sum of Gaussian components. Gaussian decomposition method was used to process the data as it assumes that both transmitted and received waveforms are Gaussian in nature and can be fitted reasonably well using Gaussian peaks (Brenner et al. 2003). The transmitted waveform $W_x(t)$ is assumed to have a bell shape and modelled as a Gaussian function (Eq. 1):

$$W_x(t) = A_x e^{-\frac{(t-x)^2}{2\sigma_x^2}} \quad (1)$$

where, A_x is the amplitude of transmitted pulse; x is the mean value representing the peak location and σ_x represents width of transmitted pulse at half power. The received waveform is modelled as a sum of Gaussian components (Duong 2010). The returned waveform energy was normalized and decomposed into different Gaussian components. Next, the parameters extracted from the waveforms by Gaussian decomposition were used to categorise land cover types. Three parameters, viz, *height*, front slope angle (*afslope*) and canopy return ratio (*rCanopy*) (Table 1) were extracted from the waveforms for clustering of eighty-five ICESat/GLAS footprints over the study area. The *height*, *afslope* and *rCanopy* parameters were chosen because these three parameters have enough potentiality to distinguish the characteristics of land cover types. Height of major land cover types in the study area, viz. forest, mango orchard, agricultural crop and barren/fallow land may vary. But when waveform returns from heterogeneous or rough terrain, vegetation information gets mixed with the ground return which may lead to erroneous height estimation. So, height parameter may fail to distinguish the land cover types accurately. Hence, *afslope* and *rCanopy* were also considered along with *height* in land cover classification. The *afslope* is angle from vertical to vector from waveform begin to peak of the canopy return energy. This provides the information about vertical variability of the upper canopy and canopy density (Boudreau et al. 2008). Whereas *rCanopy* is the canopy return ratio meaning canopy return energy to total return waveform energy i.e. nadir-projected vegetation cover area

Fig. 1 Location of study area

versus total area. So, $rCanopy$ represents total canopy cover in the GLAS footprints (Harding and Carabajal 2005). The canopy height, canopy cover, and upper canopy variability with structural information of the canopy can be derived using these three parameters. Combining these three parameters, different land cover types were identified for eighty-five footprints using KM, PAM and FCM clustering methods based on Euclidean distance. Cluster sizes of three, four and five were considered for each method. For each cluster the major land cover types were assigned based on field knowledge (Table 2). Clustering, a primitive unsupervised classification technique, is basically partitioning of

dataset into different groups, where the data in same cluster are considered as similar type.

The KM is one of the simplest unsupervised algorithms which is known to solve well-known clustering problems easily (MacQueen 1967). The procedure follows easy and simple way to classify a given set of data through a fixed number of clusters (k). The k centroids are defined for each cluster. These centroids should be placed in a cunning way because different location causes different results. Better choice, therefore, is to place them as far away as possible from each other. The objective of KM algorithm is to minimize the squared error objective function (Eq. 2).

Table 1 GLAS waveform parameters (Duong 2010)

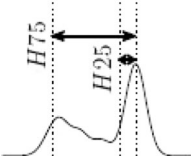
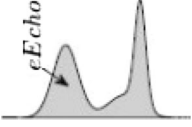

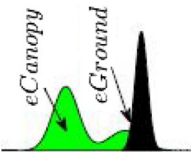
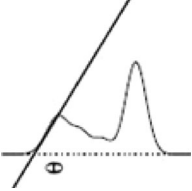
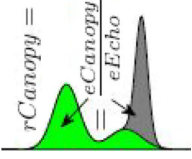
| Sl. No | Parameters | Definition | Visualization |
|--------|----------------|---|---|
| 1. | <i>Height</i> | Height at which 100 % of the return energy occurs |  |
| 2. | <i>eEcho</i> | It is the received energy that means the area below the waveform between waveform begin and waveform end |  |
| 3. | <i>eGround</i> | eGround is the total intensity of the last Gaussian mode i.e. the ground return energy |  |
| 4. | <i>eCanopy</i> | eCanopy is the difference between return waveform energy (eEcho) and ground return energy (eGround) eCanopy = eEcho – eGround |  |
| 5. | <i>afslope</i> | Angle from vertical to vector from waveform begin to peak of the canopy return energy |  |
| 6. | <i>rCanopy</i> | It is the canopy return ratio means the canopy return energy (eCanopy) divided by return waveform energy (eEcho) |  |

Table 2 Land cover types for different cluster sizes

| Sl. No | Land cover types in cluster size | | |
|--------|----------------------------------|-------------|--------------------|
| | 3 | 4 | 5 |
| 1. | Forest | Forest | Forest |
| 2. | Orchard | Orchard | Orchard |
| 3. | Others | Agriculture | Agriculture |
| 4. | – | Others | Barren/fallow land |
| 5. | – | – | Others |

$$J = \sum_{j=1}^k \sum_{i=1}^n |x_i^{(j)} - c_j|^2 \tag{2}$$

where, J is objective function, k is the number of clusters, n is number of cases, x_i is for each case i of data set x, c_j is

the centroid for cluster j and $|x_i^{(j)} - c_j|$ is the distance function.

Kauffman and Rousseeuw (1990) proposed a clustering algorithm PAM which maps a distance matrix into a specified number of clusters. The medoids are generally a robust representations of the cluster centres, which is particularly important in the common context that many elements do not belong well to any cluster. The PAM is the algorithm to find a local minimum for the k-medoids problem which may not be the optimum, but it is faster than exhaustive search. Instead of taking the mean value of the objects in a cluster as a reference point, a medoid, which is the most centrally located object in a cluster can be used. Thus the partitioning method can still be performed based on the principle of minimizing the sum of the dissimilarities between each object and its corresponding

reference point. It minimizes the sum of pair-wise dissimilarities instead of sum of squared euclidean distances. This forms the basis of the k-medoids method (Kauffman and Rousseeuw 1990).

The FCM method allows one piece of data to belong to two or more groups and is frequently used in pattern recognition (Bezdek et al. 1981). In the FCM, the following objective function (Eq. 3) is minimized:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m |x_i - c_j|^2; \quad 1 \leq m < \infty \tag{3}$$

where, m is any real number >1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the i th of d-dimensional measured data, c_j is the d-dimension centre of the cluster. An iterative optimization of the above mentioned objective function is carried out in fuzzy partitioning method, with the update of membership u_{ij} (Eq. 4) and the cluster centres c_j (Eq. 5):

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \tag{4}$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \times x_i}{\sum_{i=1}^N u_{ij}^m} \tag{5}$$

This iteration will stop when, $\max_{ij} \left\{ \left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right| \right\} < \varepsilon$ where, ε is a termination criterion between 0 and 1; and k is the number of iteration steps. This procedure converges to a local minimum or a saddle point of J_m .

After performing different clustering algorithms with different cluster sizes, the datasets were verified with field observations and GoogleEarth imagery. The overall classification accuracy was evaluated for each method and their corresponding clusters. Also, clustering validation techniques were applied to evaluate the goodness of clustering results. There are two types of clustering validation technique: external validation and internal validation. External validation is based on prior knowledge about database whereas internal validation technique relies on the intrinsic information of the data alone. In the present study, external information about the land cover classes is available. Therefore, in the current situation external validation measures are chosen for comparing the clustering methods. Entropy, and coefficient of variation and F-Measure were considered to validate the classification techniques for different cluster sizes (Rendón et al. 2011).

Entropy measures the purity of the clusters with respect to a given cluster label (Wu 2012). The value of entropy ranges from zero to any real number based on the variety of the objects in the clusters. So, the class distribution of the objects in each cluster is needed to understand for

computing the entropy of a dataset. The entropy of the dataset is computed as follows (Eqs. 6, 7):

$$E_j = - \sum p_{ij} \log(p_{ij}) \tag{6}$$

$$E = \sum_{j=1}^m \frac{n_j}{n} E_j \tag{7}$$

where, n_j = size of cluster j, m = number of cluster, n = total number of objects and p_{ij} is the probability of assigning an object of class i to cluster j.

Also, the coefficient of variance of the objects of each cluster was calculated to examine the dispersion of the resulting dataset. This is a dimensionless quantity (ratio of standard deviation to mean) for comparing the variance of the population. A higher value of the coefficient of variance implies larger variation in the dataset. The F-Measure is about clustering quality. It ranges from zero to one, the value near to one indicates higher clustering quality. The F-Measure of cluster j and class i is calculated as follows (Eq. 8):

$$F(i, j) = \frac{2 \frac{n_{ij}}{n_i} \frac{n_{ij}}{n_j}}{\frac{n_{ij}}{n_i} + \frac{n_{ij}}{n_j}} \tag{8}$$

where, n_{ij} is the number of objects of class i that are in cluster j, n_j is the number of objects in cluster j, and n_i is the number of objects in class i.

Results and Discussion

The KM, PAM, and FCM clustering methods were applied with different clusters to extract land cover information from GLAS full waveform (Figs. 2, 3). Proper selection of features used for classifier training is very important as it increases classification accuracy. Initially forest and mango orchard were considered as major classes and agriculture, barren/fallow land, settlement, dry river bed, etc. were grouped as other class. The KM, PAM, and FCM with three clusters were examined considering these three classes. By comparing classified footprints with actual field scenario, around 89 % accuracy was achieved for all clustering methods (Table 3). This enforced to do the same exercise with more land cover classes. So cluster size was increased to four: forest, mango orchard and agriculture were taken as key classes and barren/fallow land, settlement, dry river bed etc. were grouped into other class. When clustering size became four the accuracy for all clustering methods decreased drastically (Table 3). Out of three clustering methods, the KM showed considerably better accuracy. The land cover classes were further classified into five classes to see how the clustering methods behave with increase in cluster size. Forest, mango orchard,

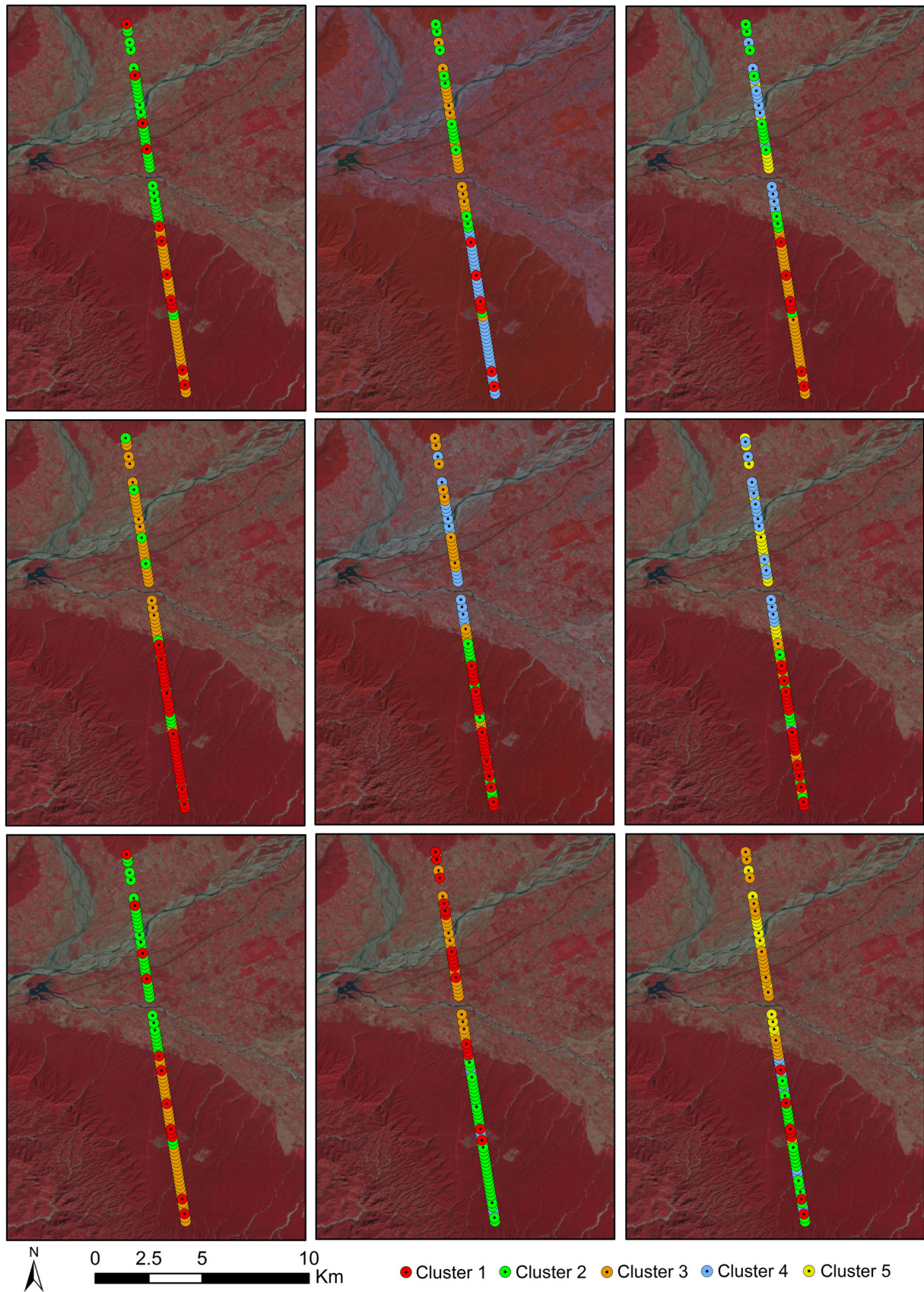


Fig. 2 Classified GLAS footprint over Landsat 7 (ETM) imagery of 14 October 2008

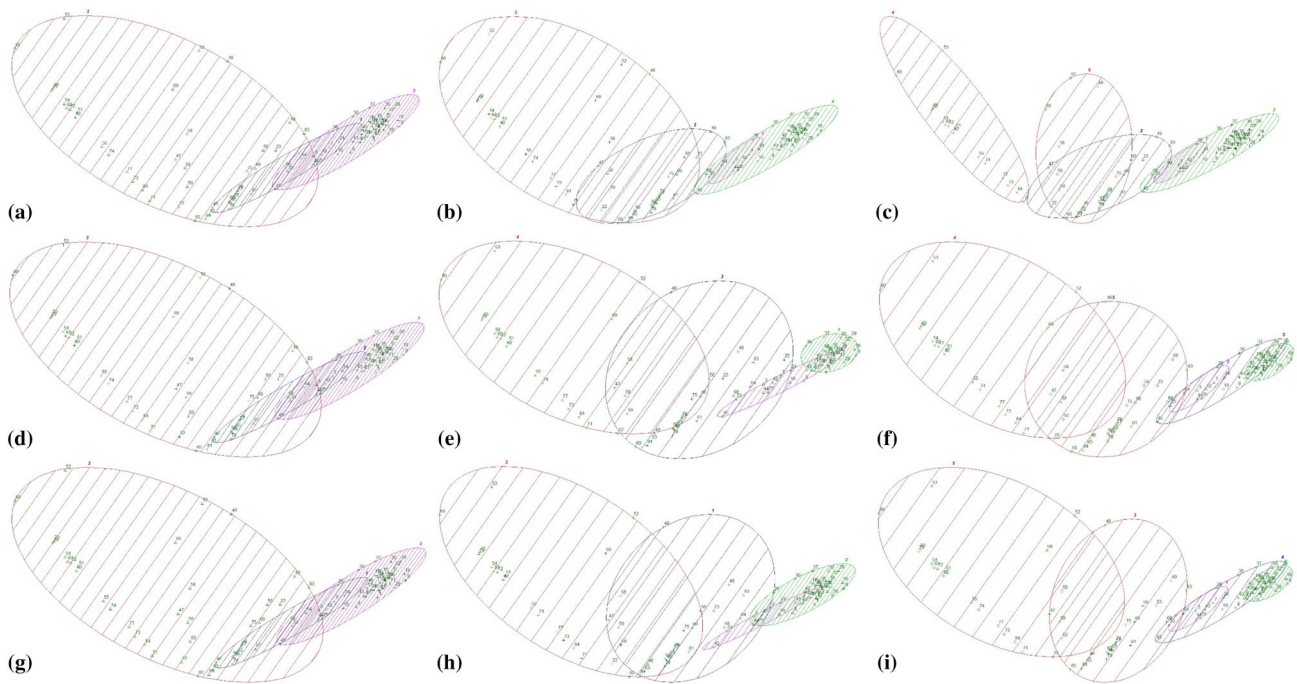


Fig. 3 Clustering results. KM with **a** 3 clusters **b** 4 clusters **c** 5 clusters; PAM with **d** 3 clusters **e** 4 clusters **f** 5 clusters and FCM with **g** 3 clusters **h** 4 clusters **i** 5 clusters

Table 3 Overall land cover classification accuracy of different clustering methods

| Methods | Accuracy (%) in cluster size | | |
|---------|------------------------------|-------|-------|
| | 3 | 4 | 5 |
| KM | 89.41 | 72.94 | 72.94 |
| PAM | 89.41 | 60.00 | 42.35 |
| FCM | 89.41 | 68.24 | 60.00 |

agriculture and barren/fallow land were selected as four distinct classes, while remaining classes were clumped together as other class (Table 2). Accuracy of PAM decreased rapidly to 42.35 % (Table 3). The FCM also showed a decreasing trend in accuracy. Interestingly, accuracy of KM did not fall further. Accuracy of each class corresponding to each method with respective cluster size was also computed (Table 3). PAM was unable to separate agricultural field from settlement and dry river bed. FCM identified agricultural class considerably well after excluding barren/fallow land class. For all cases, the KM gave significantly good result.

Accuracy of KM, PAM and FCM classification varies with cluster sizes. To investigate the reason, external clustering validation techniques were adopted. The entropy and coefficient of variation of resulting clustering sizes of different clustering techniques were computed. Then difference between coefficient of variation (DCV) of true and

resulting clustering sizes were calculated. DCV and entropy of different resulting clustering sizes were plotted (Fig. 4a). It is observed that entropy value of the PAM for cluster size 4 is smallest minimum (0.1271), but, DCV is largest (0.4042) for cluster size 4. Entropy values of cluster size 4 of KM (0.1631) and FCM (0.1649) are more or less similar, but DCV is less for KM (0.0088). Clustering result of KM is much closer to true cluster distribution. In cluster size 4, it was found that KM performed better than PAM and FCM (Table 3). So, it is observed that DCV plays a vital role in clustering quality than entropy. This is due to biased effect of entropy, especially when data have highly imbalanced true clusters. It is observed that if the entropy measure is only used to validate the clustering methods, the validation results could be misleading (Wu 2012). For cluster size 5 (Fig. 4a), similar kind of situation was arisen for PAM as it was for cluster size 4. Entropy and DCV values of KM (0.3031, 0.2064) and FCM (0.3214, 0.1455) are almost similar. However, classification accuracy (Tables 3, 4) of KM for cluster size 5 is higher than FCM in the same cluster size. So, entropy and DCV were unable to justify the higher classification accuracy of KM than FCM for cluster size 5. In this context, F-Measure test was performed. F-Measure combines precision and recall concepts from information retrieval (Rendón et al. 2011). Precision can be a measure of a classifiers exactness whereas recall can be a measure of a classifiers completeness. Combining these F-Measure gives the quality of the clustering. It was found that (Fig. 4b, c),

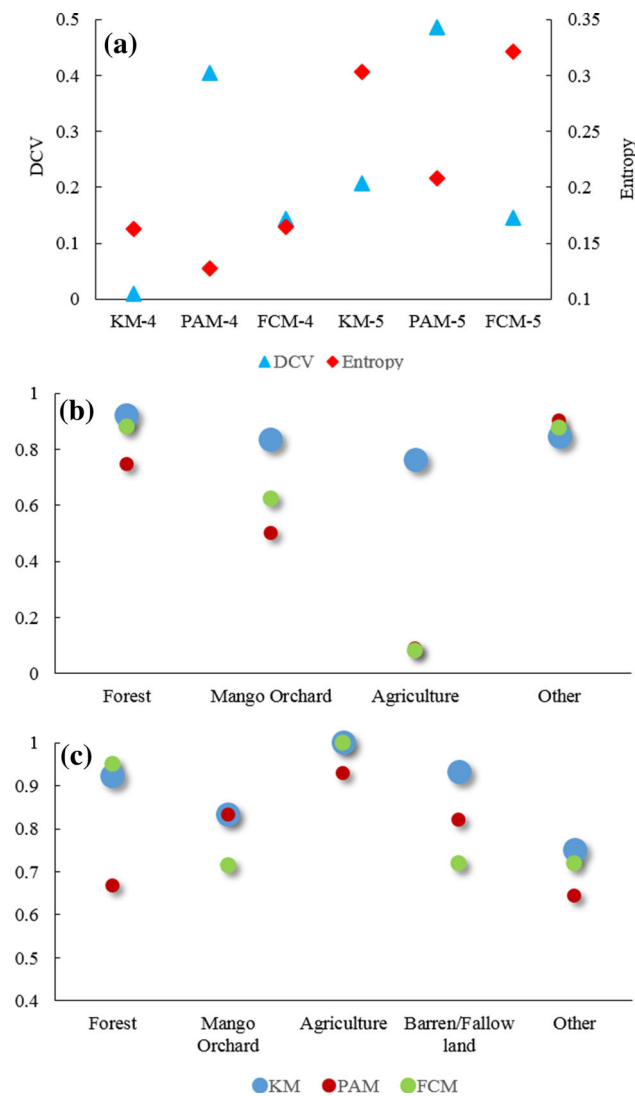


Fig. 4 Clustering validation results: **a** DCV and entropy values with cluster size 4 and 5 **b** F-Measure value for cluster size 4 and, **c** F-Measure value for cluster size 5

KM performed better than PAM and FCM for both the cluster sizes 4 and 5 by F-Measure values. The entropy and DCV could not justify the accuracy because of large class imbalance in the data set. When large class imbalance exists in the input dataset clustering method can predict the value of the majority class for all predictions and achieve a high classification accuracy (Wu 2012). So, the entropy fails to determine accuracy correctly in the present study. The KM gave best result because the parameters *height*, *afslope* and *rCanopy* are well separated from each other. The complexity of the KM, PAM and FCM are $O(nk)$, $O(k(n-k)^2)$, and $O(nk^2)$ respectively. Due to this the overhead needed for computing and managing the proximity vector explains why the PAM and FCM are quite slower than KM. Hence, KM clustering was found comparatively fast, robust and efficient than other two methods in the present study (Table 4).

Conclusions

In the present study, three partition-based clustering algorithms, viz, KM, PAM and FCM were applied to ICESat/GLAS derived parameters, viz, *height*, *afslope* and *rCanopy* for classifying land cover types. The results showed that among the three clustering algorithms, the KM clustering algorithm performed the best (with cluster size 4 and 5) with 72.94 % accuracy. The KM algorithm was able to distinguish the different land cover types more efficiently than the other two methods. However, the classification method could be improved to get better discrimination between the agriculture and barren land. This research suggests a new and promising way to derive land cover information from ICESat/GLAS full waveform data. GLAS data is freely available with near global coverage and can potentially be used for large-scale land cover

Table 4 Accuracy of land cover classification using different clustering methods

| Methods | Accuracy (%) | | | | | | |
|---------|--------------|--------|---------------|-------------|--------------------|--------|----------|
| | Cluster size | Forest | Mango orchard | Agriculture | Barren/fallow land | Others | Over all |
| KM | 3 | 85.71 | 80.00 | – | – | 94.74 | 89.41 |
| | 4 | 85.71 | 60.00 | 7.69 | – | 88.00 | 72.94 |
| | 5 | 85.71 | 60.00 | 84.62 | 43.75 | 55.56 | 72.94 |
| PAM | 3 | 85.71 | 80.00 | – | – | 94.74 | 89.41 |
| | 4 | 59.52 | 80.00 | 0.00 | – | 88.00 | 60.00 |
| | 5 | 50.00 | 60.00 | 0.00 | 43.75 | 55.56 | 42.35 |
| FCM | 3 | 85.71 | 80.00 | – | – | 94.74 | 89.41 |
| | 4 | 78.57 | 60.00 | 0.00 | – | 88.00 | 68.24 |
| | 5 | 88.10 | 60.00 | 23.08 | 25.00 | 44.44 | 60.00 |

classifications. Moreover, the results derived from the GLAS waveform analysis is also useful for comparison, validation or updation of land cover classification data obtained by other methods.

Acknowledgments The present work was carried out as part of ISRO's Earth Observation and Application Mission (EOAM) project. The authors are thankful to NSIDC for providing the ICESat/GLAS data. The authors wish to acknowledge Divisional Forest Officer and staff of Kalsi Soil and Water Conservation Division, Government of Uttarakhand for field support.

References

- Abshire, J. B., Sun, X., Riris, H., Sirota, J. M., McGarry, J. F., Palm, S., & Liiva, P. (2005). Geoscience laser altimeter system (GLAS) on the ICESat mission: on-orbit measurement performance. *Geophysical Research Letters*, 32(21), L21S02. doi:10.1029/2005GL024028.
- Andersen, M. L., Stenseng, L., Skourup, H., Colgan, W., Khan, S. A., Kristensen, S. S., Andersen, S. B., Boxa, J. E., Ahlstrøma, A. P., Fettweis, X., & Forsberg, R. (2015). Basin-scale partitioning of Greenland ice sheet mass balance components (2007–2011). *Earth and Planetary Science Letters*, 409, 89–95.
- Bartholome, E., & Belward, A. S. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, 26(9), 1959–1977.
- Bezdek, J. C., Coray, C., Gunderson, R., & Watson, J. (1981). Detection and characterization of cluster substructure I. Linear structure: fuzzy c-lines. *SIAM Journal on Applied Mathematics*, 40(2), 339–357.
- Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., & Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112(10), 3876–3890.
- Brenner, A., Zwally, H., Bentley, C., Csathó, B., Harding, D., Hofton, M., Minster, B., Roberts, L., Saba, J., Thomas, R., & Yi, D. (2003). Derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights. Algorithm Theoretical Basis Document. https://nsidc.org/data/icesat/pdf/Atbd_20031224.pdf. Accessed 20 June 2015.
- Buddenbaum, H., Seeling, S., & Hill, J. (2013). Fusion of full-waveform LiDAR and imaging spectroscopy remote sensing data for the characterization of forest stands. *International Journal of Remote Sensing*, 34(13), 4511–4524.
- Champion, H. G., & Seth, S. K. (1968). *A revised survey of the forest types of India*. New Delhi: Manager of Publications, Government of India.
- Duong, V. H. (2010). *Processing & application of ICESat large footprint full waveform laser range data*. TU Delft: Delft University of Technology.
- Duong, H., Pfeifer, N., & Lindenbergh, R. (2006). Full waveform analysis: ICESat laser data for land cover classification. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36, 30–35.
- Harding, D. J., & Carabajal, C. C. (2005). ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophysical Research Letters*, 32(21), L21S10. doi:10.1029/2005GL023471.
- Kauffman, L., & Rousseeuw, P. J. (1990). *Finding groups in data. An introduction to cluster analysis*. New York: Wiley.
- Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Del Bom Espirito-Santo, F., & de Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, 32(22), L22S02. doi:10.1029/2005GL023971.
- Liu, C., Huang, H., Gong, P., Wang, X., Wang, J., Li, W., Li, C., & Li, Z. (2015). Joint use of ICESat/GLAS and landsat data in land cover classification: A case study in Henan province, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(2), 511–522.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 281–297.
- Nandy, S., & Kushwaha, S. P. S. (2011). Study on the utility of IRS 1D LISS-III data and the classification techniques for mapping of Sunderban mangroves. *Journal of Coastal Conservation*, 15(1), 123–137.
- Neuenschwander, A. L. (2008). Evaluation of waveform deconvolution and decomposition retrieval algorithms for ICESat/GLAS data. *Canadian Journal of Remote Sensing*, 34(sup2), S240–S246.
- Rendón, E., Abundez, I., Arizmendi, A., & Quiroz, E. (2011). Internal versus external cluster validation indexes. *International Journal of computers and communications*, 5(1), 27–34.
- Rosette, J. A. B., North, P. R. J., & Suarez, J. C. (2008). Vegetation height estimates for a mixed temperate forest using satellite laser altimetry. *International Journal of Remote Sensing*, 29(5), 1475–1493.
- Wu, J. (2012). *Advances in K-means clustering: a data mining thinking*. New York: Springer Science & Business Media.
- Zhou, Y., Qiu, F., Al-Dosari, A. A., & Alfarhan, M. S. (2015). ICESat waveform-based land-cover classification using a curve matching approach. *International Journal of Remote Sensing*, 36(1), 36–60.
- Zwally, H. J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., et al. (2002). ICESat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34(3), 405–445.
- Zwally, H., Schutz, R., Bentley, C., Bufton, J., Herring, T., Minster, J., Spinhirne, J., & Thomas, R. (2011). GLAS/ICESat L2 Global Land Surface Altimetry Data. Version 33, 27.