# Comparing Curve Matching Approaches for Land Cover Classification Using Waveform Data

Yuhong Zhou[1], Fang Qiu[2]

[1]University of Texas at Dallas. 800 W Campbell Rd. GR31, Richardson, TX 75080-3021, USA
Email: yxz102020@utallas.edu

[2]University of Texas at Dallas. 800 W Campbell Rd. GR31, Richardson, TX 75080-3021, USAcontact postal address
Email: ffqiu@utallas.edu

**Abstract**

LiDAR waveform data have been increasingly available to perform land cover classification. Numerous studies have been focused on either the discretization of waveforms to obtain denser 3D point clouds (Mallet and Bretar 2009) or the extraction of metrics from waveforms to characterize their shapes (Mallet et al., 2011). In contrast, the direct use of the full waveform curve itself, which contains more comprehensive and accurate vertical structural information of ground features within the footprint, has been scarcely investigated (Zhou and Qiu 2015). The objectives of this study were to utilize the complete waveform curve directly to differentiate objects having distinct vertical structures using different curve matching approaches. We adapted four curve matching approaches: waveform matching root sum squared differential area (WMRSSDA), waveform angle mapper (WAM), Kolmogorov-Smirnov distance (KSD), and Kullback-Leibler (KL) divergence, which were extended from the existing curve matching methods that were used for pixel-based hyperspectral image classification or object-based high spatial image classification. These approaches are designed to assess the similarity between an unknown waveform and individual reference waveforms. We also developed two new curve matching approaches: cumulative matching curve root sum squared differential area (CCMRSSDA) and cumulative curve angle mapper (CCAM) to assess the similarity of cumulative distribution function (CDF) of waveform curves. Among them, WMRSSDA, WAM and KL are based on the original waveform curves, while CCMRSSDA, CCAM, and KSD are based on the cumulative waveform curves.

**Keywords:** waveform, KL, KS, SAM, ICESat/GLAS.

## 1. Introduction

Light Detection And Ranging (LiDAR) is an active remote sensing technology which uses visible, near-infrared, or short-wave infrared laser beams to measure ground elevation, while also providing information on the vertical structure of geographical objects (Zhou and Troy, 2008). There are two types of LiDAR data: discrete-return LiDAR and full-waveform LiDAR (Ussyshkin and Theriault, 2011). Currently, discrete-return LiDAR data are the mainstream commercial product and a great number of off-the-shelf tools are available for its processing (Ussyshkin and Theriault, 2011). Up to six returns for each transmitted laser pulse are typically recorded, with x, y, elevation, and intensity measured for each. These measurements have been extensively used for land cover classification (Sasaki et al., 2012).

Full waveform, a relatively new LiDAR product, records the quasi-continuous time-varying strength of the return signal from the illuminated area (called the waveform footprint) using small time intervals (e.g., 1nanosecond). It has become increasingly popular in the last decade. LiDAR waveforms are able to provide thousands of measurements for each transmitted laser pulse (Mallet et al., 2011; Ussyshkin and Theriault, 2011). Due to this finer vertical resolution, the waveform offers an enhanced capability to reflect the vertical structures of geographical objects (Wagner et al., 2008).

Numerous studies have been focused on either the discretization of waveforms to obtain denser 3D point clouds (Mallet and Bretar 2009) or the extraction of metrics from waveforms to characterize their shapes (Mallet et al., 2011). The complete curve of a full-waveform contains substantially more information than the discretized returns or summary metrics extracted from the waveform. However, the complete waveform curves have rarely been utilized to perform land cover classification because of the absence of appropriate approaches for dealing with their complexity. In an earlier paper (Zhou et al., 2015), we proposed a curve matching approach to directly compare waveform curves using Kolmogorov-Smirnov (KS) distance as an alternative to the methods based on the simplification of waveforms. Preliminary results showed that this approach outperformed a widely adopted rule-based method using waveform-derived metrics. This result suggested that a curve matching approach to waveform-based land cover classification is worth further exploration.

Besides the KS based curve matching approach we developed in Zhou et al. (2015) and the KL based approach in Zhou and Qiu (2015), we are unaware of other studies using curve matching approaches to quantify the similarity between waveform curves for land cover classification. In the current study, four additional curve matching approaches for waveform classification are proposed. The first is curve root sum squared differential area (CRSSDA). The second is curve angle mapper (CAM). Both CRSSDA and CAM were originally designed for pixel-based hyperspectral data and were extended to object-based hyperspatial image classification. Here, we extend them to LiDAR full waveforms. We also develop two new curve matching approaches based on CDF curves, cumulative curve root sum squared differential area (CCRSSDA) and cumulative curve angle mapper (CCAM), to compare the similarity of cumulative distribution functions (CDFs) of waveform curves. These approaches are all compared with each other and with our previous KS and KL based approaches.

## 2. Methodology

In this study, large-footprint waveform data from the Geoscience Laser Altimeter System (GLAS) onboard the Ice, Cloud, and land Elevation Satellite (ICESat) were used to explore the potential of our curve matching approaches because of its availability for free, regular repetition, and global coverage. Detailed information on ICESat waveforms can be found in Zhou et al. (2015).

The same dataset employed by Zhou et al. (2015), located in Dallas County, Texas, USA, was used in this study. For consistency with our previous study, the same three land use categories of open space, buildings, and trees were employed.

Curve matching approaches measure the curve similarity between two samples, one usually with known class (training sample) and the other with its class to be predicted (testing). For this purpose, after a set of preprocessing procedures, the waveform dataset was divided into training and testing subsets. For this reason, the reference waveforms were simply randomly selected from the training dataset to avoid biasing any of the six curve matching approaches. The random selection process was iterated 15 times, resulting in 15 sets of reference waveforms to determine the average performances of the six curve matching approaches. In order to maintain consistency with our earlier study, each set contained the same number of reference waveforms as selected by the principle component analysis (PCA) used previously. The significance of the performance differences between any two curve matching approaches was assessed with the pairwise McNemar's chi-square test

Generally, the performance of one curve matching approach is highly determined by its ability to capture the differences between within-class similarity and between-class similarities also termed between-class separability. Therefore, the within-class and between-class curve similarities for the methods based on original waveform curve and those based on cumulative waveform curves are investigated to give a better insight into the performances of these methods using one of the randomly selected dataset. Since different curve matching approaches employed measures in different scales or units, the resultant measures cannot be compared directly. These measures are normalized using relative spectral discriminatory probability (RSDPB) that was originally employed to compare performance of different hyperspectral classification methods to allow comparison of their performance (Ghiyamat et al., 2013). The within-class similarities are obtained by selecting the testing waveforms from the same category as that of the reference waveforms. For example, the curve similarities within the building category (Building-Building) were obtained by drawing both the testing waveforms and the reference waveforms from the building category. The between-class similarities are obtained by selecting the testing waveforms from a category that is different from that of the reference waveforms. For example, the between-class similarity (Building-Tree) was obtained by drawing the testing waveforms from the building category and the references waveforms from the tree category.

## 3. Results

A performance comparison of the ability of the six curve matching approaches to identify three land use classes (building, tree and open space) based on 15 sets of randomly selected references was conducted. KL provided the highest average classification accuracy, closely followed by CCRSSDA and CCAM, but there were no significant differences between any two of these methods. However, they all significantly outperformed KS, CRSSDA, and CAM. Generally, approaches based on the CDF-curve of waveforms are preferred over those based on the original waveform curves when the same algorithm is used for classification. However, the CDF-curve based approaches are not always superior to those based on original waveform curves when different algorithms are used.

Just as KL, KS, CAM, and CRSSDA were extended from algorithms originally developed for hyperspectral and high spatial image classification, CCAM and CCRSSDA, which were newly developed for waveform classification, could also be applied to pixel-based hyperspectral image classification, object-based high spatial image classification, and the fusion of imagery and LiDAR waveform data. Future research is needed to assess if these cumulative frequency based approaches would achieve superior results in these application areas as they generally did for waveforms.

## 4. Acknowledgements

The authors would like to thank the National Snow and Ice Data Center by distributing ICESat data.

## 5. References

Ghiyamat, A., Shafri, H.Z.M., Mahdiraji, G.A., Shariff, A.R.M., and Mansor, S., 2013, Hyperspectral discrimination of tree species with different classifications using single- and multiple-endmember. International Journal of Applied Earth Observation and Geoinformation 23 (1): 177-191.

Ussyshkin, V. and Theriault, L.. 2011, Airborne Lidar: Advances in Discrete Return Technology for 3D Vegetation Mapping. Remote Sensing 3 (3): 416-434.

Mallet, C., and Bretar, F., 2009, Full-waveform topographic lidar: State-of-the-art. ISPRS Journal of Photogrammetry and Remote Sensing, 64 (1): 1-16.

Mallet, C., Bretar, F., Roux, M., Soergel, U., and Heipke, C., 2011, Relevance assessment of full-waveform lidar data for urban area classification. ISPRS Journal of Photogrammetry and Remote Sensing 66 (6 SUPPL.), S71-S84.

Sasaki, T., Imanishi, J., Ioki, K.., Morimoto, Y., and Kitada, K.. 2012, Object-Based Classification of Land Cover and Tree Species by Integrating Airborne LiDAR and High Spatial Resolution Imagery Data. Landscape and Ecological Engineering 8 (2): 157-171.

Wagner, W., Hollaus, M., Briese, C., and Ducic, V., 2008, 3D Vegetation Mapping using Small-Footprint Full-Waveform Airborne Laser Scanners. International Journal of Remote Sensing 29 (5): 1433-1452.

Zhou, W. and Troy, A.. 2008, An Object-Oriented Approach for Analysing and Characterizing Urban Landscape at the Parcel Level. International Journal of Remote Sensing 29 (11): 3119-3135.

Zhou, Y., and Qiu, F., 2015, Fusion of high spatial resolution WorldView-2 imagery and LiDAR pseudo-waveform for object-based image analysis, ISPRS Journal of Photogrammetry and Remote Sensing, 101: 221-232.

Zhou, Y., Qiu, F., Al-Dosari, A. A., and Alfarhan, M. S.. 2015, ICESat Waveform-Based Land-Cover Classification using a Curve Matching Approach. International Journal of Remote Sensing 36 (1): 36-60.