Technical Details on the Riparian Classification from LiDAR (RCL) Tool

Authors: Sky Jones[1]; Racha El Kadiri[2], PhD; Henrique Momm[3], PhD

[1]Middle Tennessee State University, rj3h@mtmail.mtsu.edu
[2]Assistant Professor, Middle Tennessee State University, racha.elkadiri@mtsu.edu
[3]Asssociate Professor, Middle Tennessee State University, henrique.momm@mtsu.edu

Abstract

A vast catalogue of LiDAR point cloud data is available online in government and private databases, and the amount of data available is constantly increasing as new missions are contracted and completed. Aerially-collected LiDAR point clouds are useful for many analytical and visualization purposes, including the delineation and characterization of vegetation. Under contract with the NRCS, the MTSU Department of Geosciences has produced a model and associated ArcGIS tool to classify riparian landcover called the Riparian Classification from LiDAR (RCL) model. This document details the design decisions and workflow used to generate this model, the scope of the model's applicability and future plans for its modification.

Goals

The main goal of this project is to create a reproducible workflow for classifying landcover in the riparian buffer using only a LiDAR point cloud as input. We have termed such a workflow an RCL (Riparian Classification from LiDAR) model; a specific RCL model has been packaged for ease of use as an ArcGIS tool called the RCL tool.

While both aerial imagery and LiDAR have been used for years to train landcover classification models, the specifics on how to properly create training data, implement a machine learning model and then evaluate the quality of the output are typically unclear to those unfamiliar with landcover classification. Thus, we have endeavored to make the RCL model as robust as possible to variations in study area physiography and LiDAR collection methods so that the final model is general; that is, the model is agnostic of how or where the input data is collected, and so can be depended on to produce quality output without exhaustive pre-input preparation and analysis.

An Introduction to LiDAR

Light Detection and Ranging, or LiDAR, uses laser pulses to measure the distance between a sensor and a target. Upon striking the target, the laser pulse is reflected and the sensor records the time between the pulse leaving the LiDAR apparatus and returning to the sensor. This information can be used to determine the distance between the sensor and object. In coordination with a GPS system and when mounted on aircraft, LiDAR systems can be used to generate detailed representations of the ground below the flightpath. These data, called *point clouds*, typically contain millions of points each representing a *return* (reflected laser pulse). Each point is associated with geographic information (northing, easting, elevation) as well as ancillary information, such as the *intensity* of the return (a measure of how reflective the target is) and the *return number*. Individual laser pulses from the LiDAR apparatus spread as they travel, resulting in an ever-widening *footprint*. If the footprint strikes an object with multiple elevation levels (such as a building edge or a tree), then the originating pulse will return in multiple pulses rather than one discrete pulse; the order that the split pulses return determines the return number for each returned pulse. Both return splitting and intensity are sensitive to the type of LiDAR apparatus used and flight characteristics. LiDAR vendors will occasionally classify landcover using proprietary models, but the classification methods vary in quality and detail.

Traditionally three types of elevation models are generated from aerially-collected LiDAR point clouds. A Digital Surface Model (DSM) is created by generating an interpolated surface using only the first returns in the point cloud, and roughly represents the elevation of the ground or, if present, above-ground structures such as trees and power lines. A Digital Elevation Model (DEM) is created by interpolating only the last returns in a point cloud, and roughly represents the elevation of the ground without any above-ground structures. Occasionally points classified as buildings will be excluded when generating a DEM, but vendor classifications cannot always be relied on for this. The third elevation product, the Digital Height Model, is the difference between the DSM and the DEM. This roughly

represents the height of above-ground structures. Because the DHM is often used as a proxy for canopy height it is sometimes called the Canopy Height Model (CHM).

These data products, as well as those derived from them such as slope models and Haralick textures can be used to train machine learning algorithms to identify the unique morphometric signatures associated with different landcover types. Creating a reproducible workflow to do this is a key goal of this project.

Methods

A watershed in western Tennessee (HUC 080102040304) was the primary focus for this study. Data for a 2012 LiDAR mission entirely covering the watershed was obtained and used to create multiple data products, including but not limited to a DEM, DSM, DHM, and derived slope models. Approximately 6.4% of the watershed's landcover was manually classified into one of 19 categories **(Table 1)**, which was then used along with the LiDAR-derived data to train a decision tree to classify landcover. This process was repeated for six additional watersheds across the continental US **(Table 2)** in order to evaluate the effects of physiography and LiDAR vendor on model validity; training coverage accounted for approximately 10.0% of all land in the study areas.

A TN-specific model trained only on data derived from the west TN watershed (HUC 080102040304) as well as a general model trained on all watersheds were created. The general model was allowed access to allow derived datasets, but the TN model was restricted to datasets that can be easily reproduced using ArcGIS software and ArcPy. The full explanation of data types used in each model can be found in **Table 4.** Additionally, though 18 landcover types were manually classified, model output was restricted to 2 classes (binary models: trees and all other) or 3 classes (ternary models: trees, herbaceous vegetation and all other) due to insufficient differences in LiDAR signature between most classes.

Though this model is primarily intended to classify riparian landcover, landcover throughout the entirety of the watersheds was classified. Doing so increases the amount of training data available to the model and diversifies the LiDAR-signatures encountered during training.

Results

The decision trees for a ternary and binary, western Tennessee-specific (HUC 080102040304) ArcGIS-friendly models are shown in **Figures 1 and 2**. More general models **(Figures 3 and 4)** are provided as well. Quality metrics for each model are available in **Table 3**. "ArcGIS-friendly" means the model uses data products that are easily generated using native ArcGIS tools. A sample classification is provided in **Figure 5**.

Discussion

Both the binary general model and the binary TN model exhibit high F1-scores, indicating strong predictive power, and each do so using a decision tree with a single branch. This is not particularly surprising given that the binary models need only to differentiate tree and non-tree cover, but such binary classification schemes are still useful for many remote-sensing and watershed investigations. The significant contributor to model misclassification in these schemes are building edges and powerlines, both of which produce return splitting that is similar to the splitting produced by trees. Though return numbers are not directly used in the model, this splitting causes artifacts in the DHM that are similar to the signature of trees. Misclassifications of this sort are difficult to eliminate due to the relative infrequency of building edges and powerlines compared to other landcover types.

The ternary general and ternary TN models both exhibit acceptable predictive power in general, but are relatively limited in their ability to identify vegetative cover. In general herbaceous produces a faint, inconsistent LiDAR signature and often accounts for a small portion of total landcover during the leaf-off season when most LiDAR missions are flown,

including those whose data is included in this study; this reduces the available training samples and likely causes training sample inconsistency. The TN model's predictive abilities are reduced further due to its data restrictions; both roughness and Haralick textures were excluded from this model due to the difficulty of generating them in ArcGIS, but the DHM roughness and cluster shade are relatively diagnostic of the presence or absence of vegetation. Like the binary models, some model inaccuracy can be attributed to misclassification of building edges and powerlines, though significant inaccuracy is due to rough tilled soil, which produces a signature that is similar to the rough textures of herbaceous vegetation.

The binary and ternary TN models were packaged into an ArcGIS tool; these are pretrained models that can be used without supplying a training classification file. The binary model is expected to be relatively generalizable outside of urban areas because it classifies exclusively based on DHM signature. Because a signature on the DHM occurs due to return splitting, and because non-urban areas do not contain significant areas of return-splitting buildings or powerlines, it can be assumed that any DHM signature is due to a tree.

The ternary TN model may generalize within TN, but it is unknown how the model may be affected by differing LiDAR collection methods (such as sensor type of flight altitude). The general model appears robust to both changes in LiDAR collection methods and physiography, but this model could not be packaged within the RCL tool because it requires roughness and Haralick texture data products, both of which are difficult to generate natively in ArcGIS.

All models display improved classification power when the study area is restricted to the riparian corridor, which typically lacks the anthropogenic structures and tilled soils that contribute to a significant portion of model misclassifications.

It is important to note that all models were trained using leaf-off (fall-winter) LiDAR data, which comprises the bulk of publicly-funded LiDAR missions. Thus, these models may not generalize to LiDAR collected during other seasons even if they generalize along other parameters such as physiography or LiDAR collection method.

Future Plans

Further research into generalizing the model is planned. In particular, the effect that different LiDAR collection methods (e.g., sensor type, flight attributes) have on model predictions will be investigated in greater detail. Fortifying the model against these changes is essential for creating a tool that minimizes user input because it eliminates the need to create custom training classes for each LiDAR dataset. Proving true generality will require testing models against entirely unseen study areas and evaluating performance, which will involve generating training data for new study areas, particularly in Tennessee.

Since more research is needed to generalize the RCL model, the RCL tool will be updated to accept a landcover classification training file in the meantime. This would allow cause the tool to generate and use a custom-trained classification model rather than the packaged RCL model if desired while still abstracting away the step-by-step process of training a machine learning classification model, though some understanding of training sample selection and bias will still be required. Such custom models also have the advantage of being able to use non-standardized data such as return numbers and intensity which can lead to classification improvements.

Acknowledgements

Many thanks to the National Resources Conservation Service for providing funding for this project.

Tables and Figures

Table 1. Landcover classes used to train the model. The more specific categories (left column) were grouped together (right column) in order to improve model quality before training.

Category	**Reclassification**
Forest	Trees
Linear Trees	Trees
Individual Trees/Small Clusters	Trees
Building Tops	Other
Building Edges	Other
Dirt/Bare Field	Other
Crops	Herbaceous Vegetation
Rough Vegetation	Herbaceous Vegetation
Other Impervious Surfaces	Other
Water	Other
Snow	Other
Bare Rock	Other
Sand	Other
Wetlands	Herbaceous Vegetation
Power Lines	Other
Charred Trees and Vegetation	*(excluded)*
Utility Easement	Herbaceous Vegetation
Large-Scale Urban	*(excluded)*
Canyon	Other

Table 2. Locations used to generate the general model. The western TN HUC (080102040304) was of particular focus for this study, and a special TN-specific model was generated using it.

HUC12	180500020905	070801050901	130202090102	080102040304	010500021301	030902040303	140801040103	**TOTAL**
Name	Lobos Creek-Frontal San Francisco Bay Estuaries	Walnut Creek	Ox Spring Canyon	North Fork Forked Deer River Middle	Branch Lake	Middle Fakahatchee Strand State Preserve	Mineral Creek	
Location	San Francisco, CA	Kelley, IA	Near Alamo, NM	Western TN	Penobscot, ME	Naples, FL	Central CO	
Type	Urban	Cropland	Mountains, Desert	Cropland	Rugged Coastal	Flat coastal	Mountainous	
Size (sq km)	25.3	50.7	66.1	161.1	80.0	74.7	136.8	**594.7**
Trained Area (sq km)	1.3	8.8	7.7	10.3	10.4	8.1	13.2	**59.6**
Training Fraction	5.1%	17.3%	11.6%	6.4%	13.0%	10.8%	9.6%	**10.0%**
F1-Score	78.5%	98.4%	88.0%	86.1%	93.3%	96.3%	89.8%	**90.1%**

Table 3. Quality metrics for each different models The TN model was training using only data from HUC 080102040304, while the general mode was training on all 7 watersheds. The TN model (both ternary and binary) are what is packaged in the RCL tool.

TN, Ternary			
	Precision	**Recall**	**F1-Score**
Trees	97.3%	91.4%	94.3%
Herb. Veg.	49.5%	61.7%	55.0%
Other	91.2%	88.7%	89.9%
Accuracy	85.8%		
Macro Avg.	79.4%	80.6%	79.7%
Weighted Avg.	87.3%	85.8%	86.4%
TN, Binary			
	Precision	**Recall**	**F1-Score**
Trees	92.1%	97.3%	94.6%
Other	98.8%	96.2%	97.4%
Accuracy	96.5%		
Macro Avg.	95.4%	96.8%	96.0%
Weighted Avg.	96.7%	96.5%	96.6%
General, Ternary			
	Precision	**Recall**	**F1-Score**
Trees	96.4%	93.8%	95.1%
Herb. Veg.	17.5%	60.0%	27.1%
Other	93.7%	77.6%	84.9%
Accuracy	86.9%		
Macro Avg.	69.2%	77.1%	69.0%
Weighted Avg.	92.9%	86.9%	89.2%
General, Binary			
	Precision	**Recall**	**F1-Score**
Trees	95.4%	96.8%	96.1%
Other	94.8%	92.7%	93.7%
Accuracy	95.2%		
Macro Avg.	95.1%	94.7%	94.9%
Weighted Avg.	95.2%	95.2%	95.1%

Table 4. Explanation of LiDAR-derived data used in creating the RCL models. The "Model Restriction" column shows which data were Some data that was generated was not used to train the models at all ("Not analyzed") or was used to train the general model but not TN model ("Yes").

LiDAR-Derived Data Name	Data Explanation	Model Restriction	Restriction Explanation
Digital Surface Model (DSM)	The elevation of last return	*Not analyzed*	Raw elevation values are not normalized and thus cannot be compared between different study areas
Digital Elevation Model (DEM)	The elevation of the first return	*Not analyzed*	
Digital Height Model (DHM)	The difference between the DEM and DSM	No	
Return Number	The average number of returns per pulse	*Not analyzed*	The values of these data are partially a function of sensor type and flight characteristics rather than purely of the sensed landform
Return Intensity	The average intensity of the returns	*Not analyzed*	
Slope of DSM	The slope of the raster in degrees	No	
Slope of DEM		No	
Slope of DHM		No	
Roughness of DSM	The absolute value of the maximum difference between a pixel and its adjacent neighbors	Yes	Focal statistics in ArcGIS are computationally expensive; this data was excluded to reduce RCL tool run time
Roughness of DEM		Yes	
Roughness of DHM		Yes	
DHM Energy	A measure of textural uniformity	Yes	Haralick textures have no native implementation in ArcGIS
DHM Entropy	A measure of randomness	Yes	
DHM Correlation	A measure of neighborhood correlation	Yes	
DHM Inverse Difference Moment	A measure of texture homogeneity	Yes	
DHM Inertia	A measure of contrast	Yes	
DHM Cluster Shade		Yes	
DHM Cluster Prominence		Yes	
DHM Haralick Correlation		Yes	

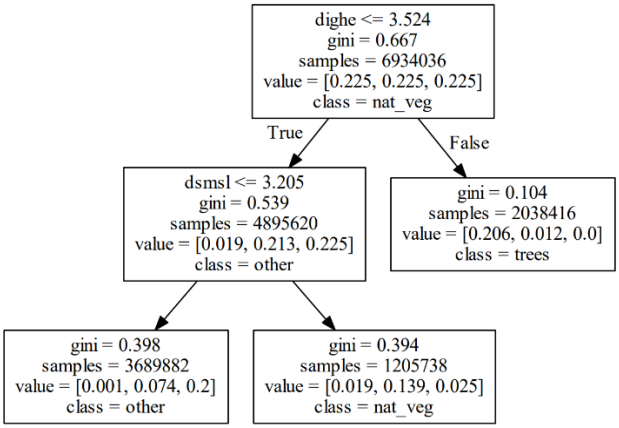

Figure 1. The decision tree for the ternary, western Tennessee specific, ArcGIS friendly model. This model and its binary counterpart is packaged in the RCL tool. *dighe = DHM, dsmsl = slope of DSM*

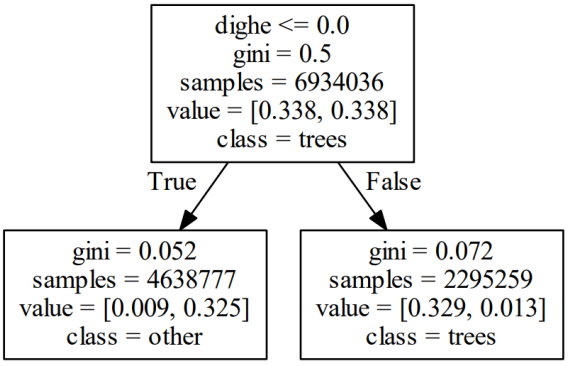

Figure 2. The decision tree for the binary, western Tennessee specific, ArcGIS friendly model. This model and its ternary counterpart is packaged in the RCL tool. *dighe = DHM*

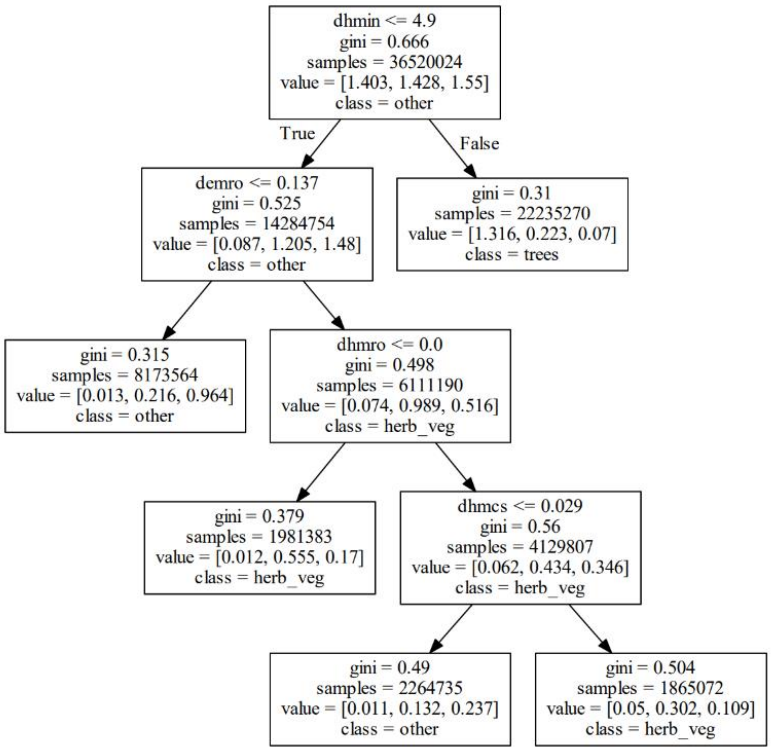

Figure 3. The decision tree for the ternary general model. *dhmin = inverse difference moment of DHM, demro = roughness of DEM, dhmro = roughness of DHM, dhmcs = cluster shade of DHM*

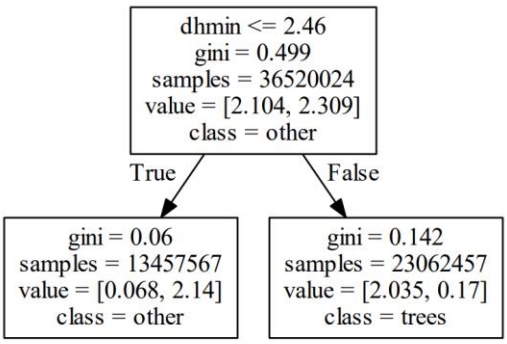

Figure 4. The decision tree for the binary general model. *dhmin = inverse difference moment of cluster shade*

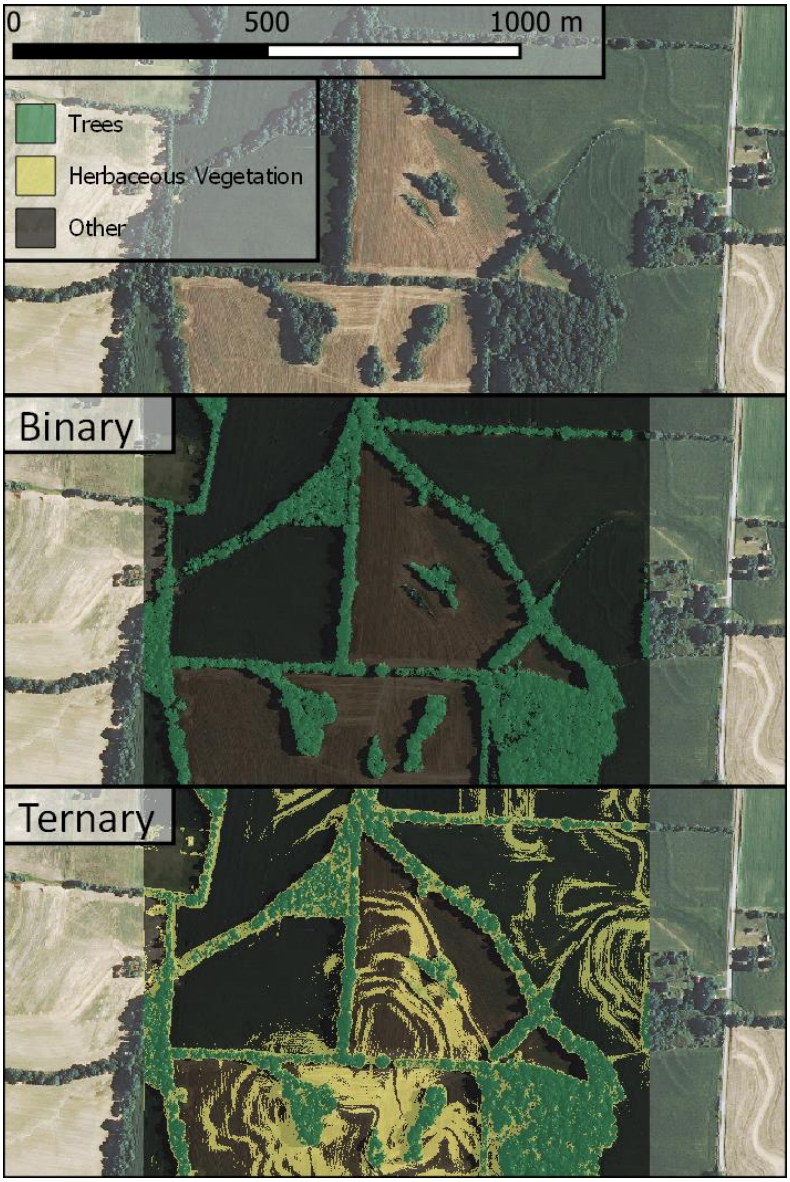

Figure 5. Example output for the TN models. Note that the tilled agricultural fields create significant misclassifications in the ternary model; for this reason it is recommended that the model is restricted to classification in the riparian buffer, where misclassification-inducing tilled fields, buildings and power lines are uncommon. The binary model is less prone to such errors, though powerlines and building edges may still be misclassified.

References

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6), 610–621. doi: 10.1109/tsmc.1973.4309314

Kashani, A., Olsen, M., Parrish, C., & Wilson, N. (2015). A Review of LIDAR Radiometric Processing: From Ad Hoc Intensity Correction to Rigorous Radiometric Calibration. *Sensors*, 15(11), 28099–28128. doi: 10.3390/s151128099

Momm, H., Easson, G., & Kuszmaul, J. (2009). Evaluation of the use of spectral and textural information by an evolutionary algorithm for multi-spectral imagery classification. *Computers, Environment and Urban Systems*, 33(6), 463–471. doi: 10.1016/j.compenvurbsys.2009.07.007