Deriving Forest Structure from LiDAR data for a Large Watershed

Mark Joselyn and Duncan Munro
Watershed Management Division, Seattle Public Utilities, City of Seattle
25th Annual ESRI International User Conference
San Diego, CA
Paper #UC2024

Abstract

Implementation of the Cedar River Watershed Habitat Conservation Plan, Seattle, WA, poses an ongoing challenge to land managers of what, if any, management action to take. The over-riding goal is to accelerate the development of late seral or 'old growth' forest conditions. Lack of data on existing conditions has made it difficult to identify and prioritize restoration project areas and to make the decision making process transparent to stakeholders. LiDAR data were used to quantify current forest condition and gap distribution. Specific derived metrics include tree height, stand density, and gap distribution. These classifications were verified based on field sample plots. The methods used and results of these efforts will be presented.

The forested 141 square mile watershed is located on the west slope of the Cascade Mountains, ranging in elevation from 500 to 5000 feet. We believe this analysis of LiDAR data across such a broad landscape are unique.

Introduction

The Cedar River Municipal Watershed (CRMW) in King County, Washington, extends eastward from the Puget Sound lowlands to the Pacific Crest in the Cascade Mountains. The CRMW provides high quality, unfiltered drinking water for over 1.3 million people. The City of Seattle owns over 140 square miles essentially encompassing the entire hydrologic drainage. These lands, many of which have been harvested for timber, were acquired over the past century. Today the land is managed by Seattle Public Utilities under the Cedar River Watershed Habitat Conservation Plan (CRW-HCP) approved by the citizens of Seattle in April of 2001. Commitments in the CRW-HCP include maintaining water quality in addition to protecting and restoring wildlife habitat and ecological features and functions of the landscape.

A Strategy for Forest Restoration

Restoring upland forests with the goals of increasing habitat complexity and accelerating the development of late successional forest conditions in second-growth forests is a key component of the CRW-HCP. Ecological restoration is an imprecise science. A restoration philosophy was adopted to help guide our restoration efforts. It identifies restoration of ecological function at the landscape level as a broad goal for our actions.

Ecosystem restoration and management of the CRW is a strategy that attempts to repair the composition, structure, processes, and/or function of human-disturbed ecosystems. To the extent possible, we seek to maintain them as self-regulating natural systems that are integrated with current ecological landscapes and land use and that eventually require minimal human intervention
(CRW Restoration Philosophy, February, 2005)

There is a range of actions that might be taken to accelerate the development of old-growth forest conditions. An Upland Forest Strategic Plan has been developed to provide a framework to guide our restoration activities. The Strategic Plan establishes a method for project site selection and prioritization. In addition it recognizes that there is a finite set of resources, staff, time and money, for implementing any action.

The Strategic Plan identifies information requirements to satisfy the development of a framework for project. The Plan identifies different spatial scales at which data is required, from site specific to landscape level. It also recognizes the need to capture data both on the *extent* and *condition* of habitat. *Extent* is defined as the location, dimensions, area, shape and boundaries of a set of habitat patches. *Condition* is defined as a measure or series of measures that quantify the properties of habitat. This broad definition of *condition* is used in order to encompass physical, compositional and structural properties of habitat.

Specific *conditions*, identified in the Strategic Plan as necessary for landscape level characterization for restoration project site selection and prioritization include:

Stem Density,
Tree Diameter (Diameter at breast height)
Stand Density Index/ Relative Density
Tree Age
Canopy Closure
Site Class
Slope
Aspect
Elevation

LiDAR Data Acquisition and Analysis to Support Forest Restoration

Characterizing the variation in *conditions* across the landscape requires a combination of data acquisition and analysis approaches. Our approach has combined both ground sampling and exploitation of remotely sensed data. Light Detection and Ranging (LiDAR) data has the potential to map some of the key *conditions* at the landscape level. Specifically, tree height, canopy gaps and canopy gap distribution, and tree density may be obtainable from LiDAR data. Efforts to extract this information from a particular LiDAR acquisition are detailed below.

King County Roads Division contracted with 3-Di to have LiDAR data acquired for areas of King County where similar data had not previously been acquired. The acquisition area included the Cedar River Municipal Watershed. These data were made available to Seattle Public Utilities (SPU) through the King County GIS Center via a standing data sharing agreement. As such the SPU was not involved in the RFP or contract specification. While very cost effective for the SPU, this approach also limited what data products are available to us and our ability to affect the acquisition process.

Seven data products were available to SPU:

Data Product	Description
Digital Ground Model. (DGM)	LiDAR last returns aggregated to 6 foot postings (pixels) representing mean ground elevation within each pixel
Digital Ground Model (DGM) Last Point Returns	ASCII files of all last return points with x,y,z values for each point.
Digital Ground Model (DGM) hillshade	Hillshade image (values range from 0-255) based on the DGM
5' Elevation Contours	Shapefile of contour lines with a 5 foot spacing derived by contractor
Digital Surface Model (DSM)	LiDAR first returns aggregated to 6 foot postings (pixels). Elevations represent the elevation of the top of the vegetation canopy elevation for each 6 foot pixel where canopy is present.
Digital Surface Model (DSM) hillshade	Hillshade image (values range from 0-255) based on the DSM.

From these delivered products attempts were made to derive maps of three *conditions*:

Tree Height
Canopy Gaps
Tree Density

In the process a number of other intermediate products were created and other data sources were utilized to perform these analyses. In particular, a classified image of broad land cover classes (coniferous, deciduous, water, non-forested) was used a various stages n the processing.

Plot level data gathered in the field was leveraged to assess the accuracy of LiDAR derived information products. At the time of this writing data has only been processed for one township (6 x 6 miles) in the lower watershed. This is due primarily to challenges in acquiring all the required data products from King County. Delivery of all 'source' data sets is anticipated by July, 2005. This area has less topographic relief than the upper watershed near the Cascade crest. This has limited our ability to evaluate the effect of more rugged terrain on the results.

Developing Maps of Tree Height with LiDAR

Inputs: DSM, DGM (both 6 foot by 6 foot cells)

A raster surface representing the top of the tree canopy (height of vegetation above the ground) can be derived from the original deliverables by subtracting the DGM from the DSM. The advantage of this approach is that it is quick and straight forward, It does not require dealing with any of the surface points data sets. The resulting values are an estimate of tree or vegetation height above the ground surface.

Comparison to ground truth data indicate a high degree of agreement with the LiDAR derived estimates of tree height (Figure 1).

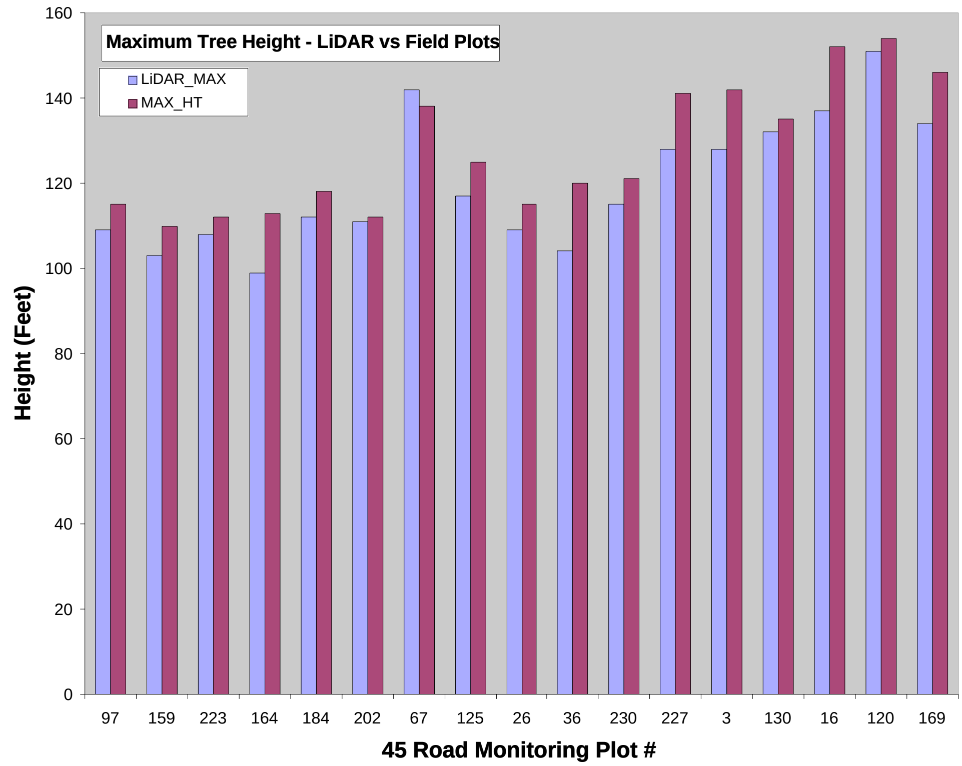

Figure 1. Comparison of tree heights estimated from LiDAR and tree height measured at fixed radius field plots.

In general the LiDAR generated vegetation height under estimated the maximum tree height based on the field data. This is not surprising as the individual LiDAR point returns will not necessarily capture tree apexes and the 6 foot pixels are an average of numerous data points. The agreement provides a high level of confidence in the delivered raster products representing tree height.

Characterizing Canopy Gaps and Mapping Canopy Gap Distribution with LiDAR

Inputs: Tree Height (6 foot cells), DGM (6 foot cells), DSM points, conceptual model of what constitutes a canopy gap, land cover layer to mask non-forested areas.

Canopy gaps are openings in the forest canopy. They indicate structural diversity and are of interest to ecologists for differentiating forest stands and potentially prioritizing restoration actions. Canopy gaps have both a vertical and horizontal dimension. The vertical dimension is what creates the opening in the canopy, the horizontal dimension equates to the area or size of the opening.

To identify canopy gaps it was first necessary to develop a quantitative definition of what constitutes a gap. The first attempt was simply to find areas where tree height was essentially zero. These are areas where no vegetation was detected. These data can then be pulled through a mask of ‘known forest’ and subsequently filtered to apply a ‘minimum gap size’ criteria.

A more sophisticated definition of a canopy gap accounts for the fact that in a forest where maximum tree heights range from 100’ to over 200 feet, a canopy gap may have significant vegetation, including young trees, growing in it. The initial quantification of gaps did not capture these areas. A definition was developed that defines a gap as any opening, over a certain size, where the understory is less than one third of the average tree height (Figure 2).

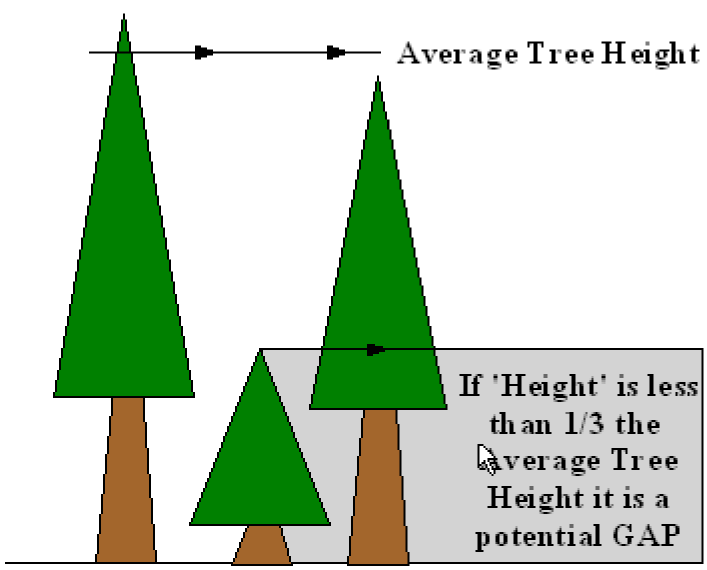

Figure 2. Schematic representation of a canopy gap.

This presented a challenge of defining, and subsequently quantifying, average tree height. By definition an average requires more than one value. In this case, average tree height implied a spatial unit of analysis larger than the pixel size of the source data (6 x 6 feet). A decision was taken to aggregate data to tiles of 125 feet by 125 feet (~ 1/3 acre) (Figure 3). Several factors influenced this decision. The delivery format for the point return data was 7500 foot blocks, each containing approximately 5 million either ground surface or first return (surface) points. This is divisible by 125 so the need to mosaic different point data sets is avoided. Also, a tile size of ~ 1/3 acre is comparable to the plot sizes used for gathering field measurements. The source rasters of surface and ground elevations contain 6 foot pixels. Each 125 foot tile contains approximately 441 of these 'source pixels'.

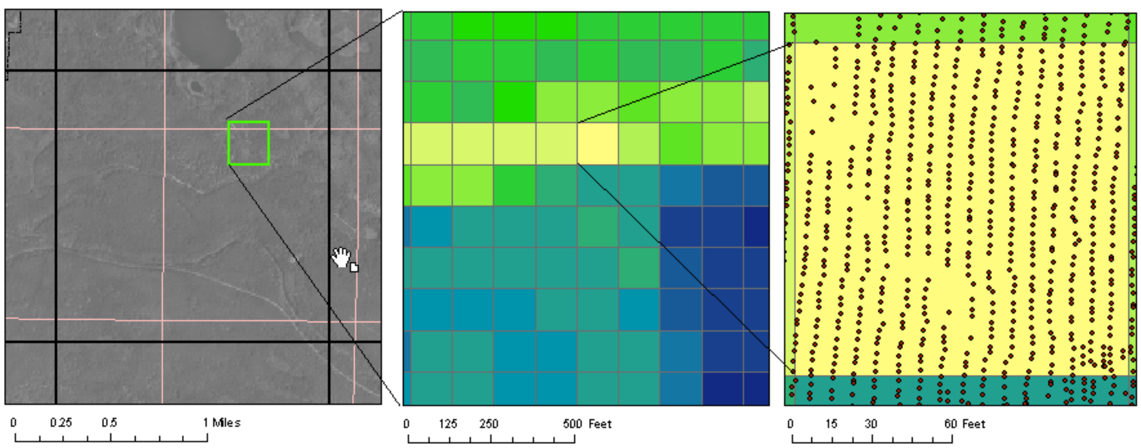

Figure 3. Partitioning the landscape into spatial units. Left: Black lines indicate the 7,500 foot 'blocks' in which point data were delivered. Middle: 125 foot tiles created for data analysis. Shading indicates the number of 'first return' points per tile. Right: Actual 'first return' points for a single tile.

Tree height values for each of the 6 foot pixels within each tile (125 x 125 feet) were 'binned' into 2 percentile classes. The class where 85 percent of the pixels were accounted for was taken as the point defining average tree height. This class value was then multiplied by the maximum tree height for that tile. For example, if a tile had a maximum tree height of 200 feet and 85 percent of all pixels fell below the 80th percentile the average tree height would be:
0.80 X 200 = 160 feet. If 85 percent of the points fell below the 90th percentile the average tree height would be 0.90 x 200 = 180 feet.

Once the average tree height was calculated for each 125 foot tile a routine was run to flag each pixel with a tree height value less than one third of this derived average tree height. Subsequently, the output grid was converted to a polygon coverage and polygons smaller in area than 576 sq ft or ~1.3 acres (equal in area to 4 by 4 source pixels) were eliminated.

The final step is to mask out non-forested areas such as roads and waterbodies. There are known issues along edges created by cutting history. For example, stand boundaries where tree heights differ significantly generate spurious average tree heights and result in 'gaps' being identified that are fictitious.

Because this analysis relied on the tree height data there is a high degree of confidence in the output. This has been confirmed by comparison with a variety of aerial photographs

Mapping Stem Density with LiDAR

Stem density, the number of trees per acre, varies across the landscape. It is an indicator, when combined with other information, about the status of a particular site.

Historically forest inventories have identified stand boundaries (polygons) and then attributed the stands. Significant limitations arise from this approach. Delineation of stand boundaries is often based on photo interpretation making it hard to repeat. Stands are in part delineated on the basis of land ownership rather than solely upon the *conditions* within the forest. The approach assumes that within stand variation is less than between stand variation. It is assumed that the stand is homogeneous and that a single value for each *condition* for a stand (e.g., stem density expressed as trees per acre) is representative of the entire stand. . To provide a more spatially explicit representation of stem density a method was developed to derive estimated stem density based on the LiDAR products described previously. The method derives stem density for each pixel in a raster and creates a basis for identification of hetereogeneity at different scales. The method does not preclude creating suites of polygons by association of adjacent pixels for which like values of any single or suite of *conditions* are estimated.

The Inverse Weighted Distance grid function within ArcGIS was used to generate a surface from the first return LiDAR points (Figure 4). This function has the property of not extrapolating above or below actual data values contained in the input. This was necessary given the very close spacing of the input data points and the fact that other potential methods (e.g., polynomial functions or kriging) generated surface values well beyond the range of reasonable values for tree heights. Initially, each point dataset was converted to a point coverage. The 'z' value associated with each point in the delivered data indicates its elevation above sea level, not a tree height.

The LATTICESPOT command was used to generate a 'height above the ground' value ('z' value relative to the ground) for each point. The ground surface used as an input to the LATTICESPOT command was the 6 foot pixel DGM. This decision was taken after investigating using the raw DGM points to create a TIN of the ground surface. The variation in the ground points created sinks and peaks that led to height values for the surface points that were unreasonable. This variation is largely removed by the aggregation that occurs in creation of the 6 foot pixel ground surface.

Parameters used in the IDW command included:

Barriers: None
Power: 2.9
SAMPLE search method
Num_points: 8
Max_radius: 16
Cell Size: 0.5 feet

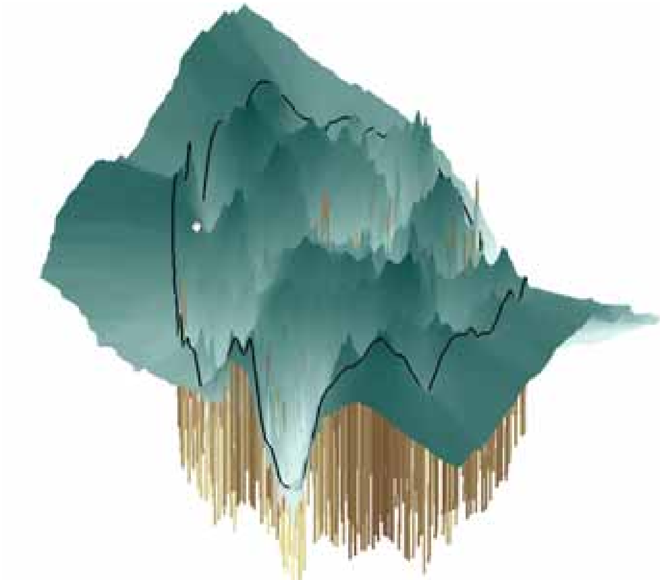

Figure 4. IDW surface created from LiDAR first returns.

Various approaches were investigated to derive tree density from this surface. The method employed involves the following steps for each 125 for by 125 foot tile:

- Create a small radius focal mean surface of the original (IDW) surface of canopy.
- Create a second focal mean surface of the same canopy surface with a larger radius.
- Difference these two surfaces.
- Threshold the resultant surface to identify the upper one sixth of the distribution
- Count the number of polygons that result as surrogates for trees
- Scale result to represent trees per acre (multiply number of trees by 2.78784)

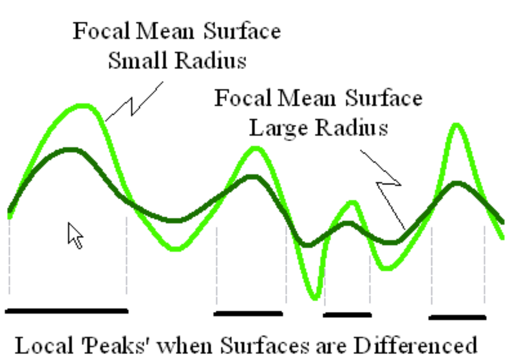

Figure 5. Schematic representation of identifying peaks by comparing two 'focal mean' surfaces.

Varying radial distance for the focal mean steps was investigated. The results were compared to plot level field data. It was determined that three differing levels of focal mean pairs best calibrated the results to the plot level data. Specifically,

For tiles:
-greater than 60% non-forested – ignore
-with deciduous composition > 25 % use radial distances of 2 feet and 8 feet
-average tree height < 100 feet use radial distance of 3 and 10 feet
-average tree height > 100 feet use radial distances of 4 and 12 feet

The resultant map of tree density showed some patterning. This was attributed in large part, to the range of LiDAR returns within a given tile. The values range from 89 returns per tile (in a tile dominated by water which absorbs the laser pulses) to 2853 returns per tile within the 36 square mile study area (Township 22 N, Range 7 E).

The empirical development and calibration of the method is similar to published studies on the derivation of stem density from LiDAR data. In this case, the variable nature of the source data and the scope of the landscape over which it is to be applied suggested that an empirical approach to model development would be best.

Delivery of additional data is anticipated by July at which time these methods will be employed to derive stem density estimates for the entire watershed. At present field measurements are being taken to assess the accuracy of the estimated stem densities.

Conclusion

The current study illustrates new potential for mapping forest habitat *conditions* at landscape scales. Success in such mapping tasks will enable us to manage the habitat resources we are responsible for in a holistic fashion. As with many mapping methodologies that leverage remotely sensed data there are some areas of uncertainty as to the degree to which we can extend the empirical approach to mapping forest habitat *conditions* across the landscape. The method of de-limiting tree crowns by thresholding the upper 6th quantile from the differenced 'focal mean' surfaces has been developed by experimentation. Further work is needed to evaluate the methodology.

At the current time we have not tested the empirical approach in forest habitat with very high stem density (> 500 stems per acre). In these cases we predict that the lack of significant height variation in the canopy surface will limit use of this method because the focal mean filtering approach will fail to identify any polygons representing tree apexes.

Similarly, we have not applied the method in areas of steep terrain in the CRMW. Earlier work has indicated that the error in location of LiDAR pulse returns increases in more rugged terrain as the footprint of the instrument is influenced by the viewing geometry of the sensor and the slope and aspect of the terrain.