Geocomputation for Urban Planning in the era of big data: Developing automated tools for analyzing lidar data to inform urban planning and policy in Portland, OR

J. L. Voelkel[1], V. Shandas[2], M. Rao[3], A. M. Thompson[4]

[1]Portland State University
506 SW Mill St., Suite 750, Portland, OR 97201
Telephone: 1 (971) 404-1843
Email: jvoelkel@pdx.edu

[2]Portland State University
506 SW Mill St., URBN 350, Portland, OR 97201
Telephone: 1 (503) 725-5222
Email: vshandasl@pdx.edu

[3]Portland State University
506 SW Mill St., Suite 750, Portland, OR 97201
Telephone: 1 (503) 725-4043
Email: mraol@pdx.edu

[4]Portland State University
506 SW Mill St., Suite 750, Portland, OR 97201
Telephone: 1 (503) 725-4043
Email: amt7@pdx.edu

1. Introduction

Currently 54.0% of the world's population lives in urban environments and by 2050 that number will increase to 66.4% (United Nations 2014). However, there are many unintended consequences of rapid urbanization such as urban air pollution and equitable distribution of services. Therefore, there is an urgent need to better characterize the urban environment in order to minimize these unintended consequences. The urban forest is one of the critical but under-characterized components of the urban environment that shows great potential in mitigating some of these unintended consequences. It has been shown that urban green spaces contribute to human health (Donovan 2013), well-being (Dallimer 2012), ecosystem services (Escobedo 2011), and societal health (Donovan 2010). However, the urban forest, like urban areas in general, shows a high degree of spatial heterogeneity at a fine spatial scale, and the lack of fine spatial characterization presents a challenge in optimizing the benefits of the urban forest. Lidar data can be used to characterize the urban forest at this fine spatial scale, but requires time, computational power, and specialized knowledge, making it inaccessible to many researchers.

In this study, we present two geocomputational tools to automate and speed up the processing of lidar data. We apply these tools to develop metrics characterizing the urban forest in Portland, OR at a 1-meter resolution. The metrics include presence, height, and density of canopy. We address the challenges presented in characterizing the urban forest with lidar data in order to make lidar accessible and practical to environmental scientists, demographers, and urban planners.

2. Methods

2.1 Data

The lidar data for the Portland Metropolitan Area was obtained as unclassified .las files through the Oregon Lidar Consortium as 0.75-minute USGS topographic quad tiles. The data was collected during leaf-on conditions in the summer of 2014 and contains approximately 8 points/m^2 (OLC 2014). Data on current infrastructure (buildings, roads, water bodies) was used for visualization and verification (Metro 2015).

2.2 Processing

In order to create an urban canopy dataset, the .las files were processed in software to determine what features exist (buildings, trees, and ground). A raster dataset representing canopy presence is created from this new pointcloud. Supplemental raster datasets (fig. 1) are then made from the feature-classified pointcloud which are used in the creation of canopy height and density datasets.

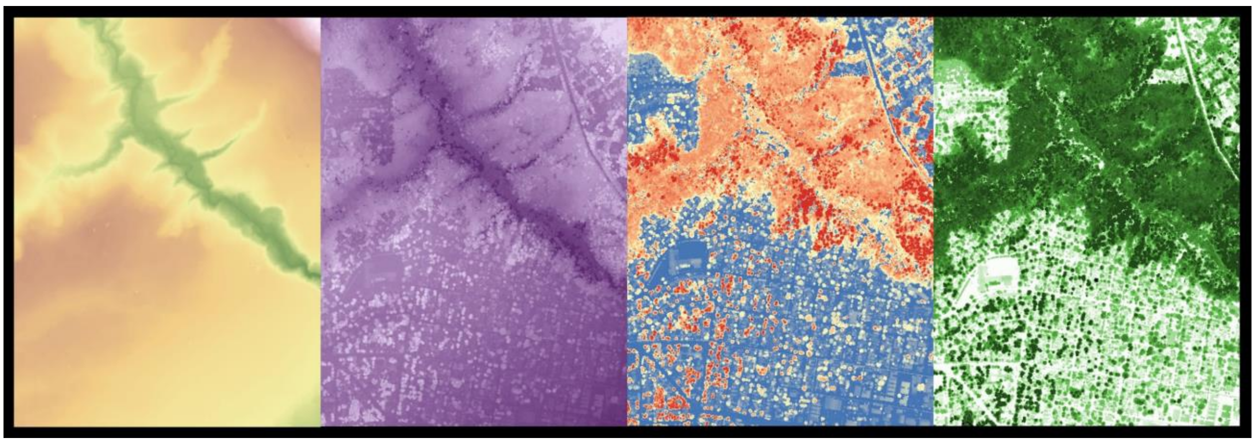

Figure 1. *From Left to Right:* 1.0m resolution examples of a Digital Elevation Model, Digital Surface Model, feature height above ground raster, and canopy height above ground raster. All created directly from lidar data with our developed tools.

2.3 Automation of Workflow

In order to maximize the efficiency of our workflow, we developed automation tools to take on the bulk of the processing. The first tool iterates through lidar files and automatically processes them, outputting a classified pointcloud identifying canopy, buildings, and ground. The second tool creates a larger composite pointcloud from smaller processed tiles by merging them together. Next, this tool creates all desired raster outputs with no need for additional processing.

3. Results and Discussion

3.1 The Urban Forest

As seen in fig. 2, our tools are able to process data covering large areas while possessing incredibly high resolution.

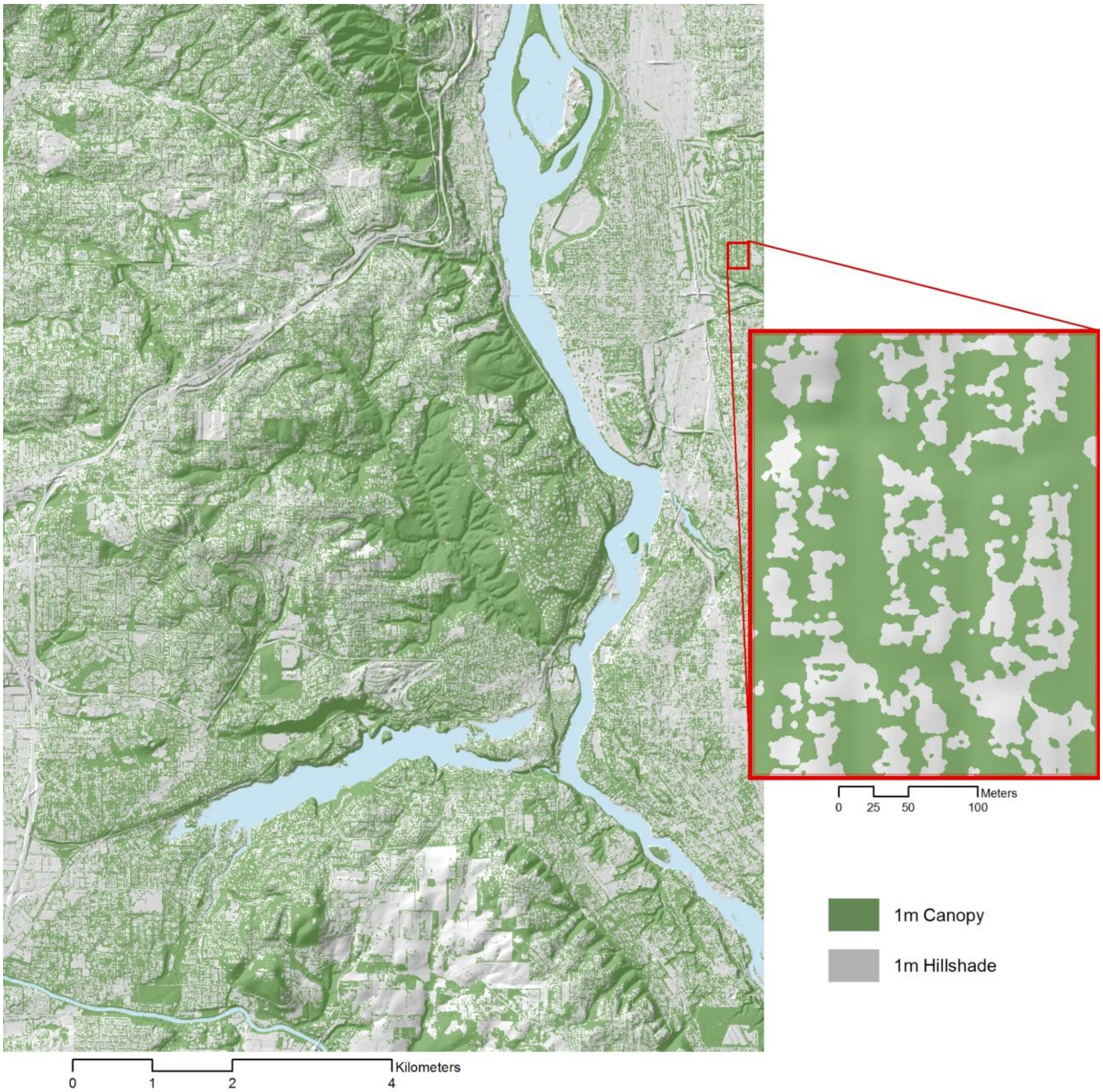

Figure 2. A 7.5-minute topographic quad (and detail) within the Portland Metro Area displaying a canopy presence dataset created with our automation tools.

3.2 Automation Tools

The tools created were developed to maximize efficiency. The raw lidar processing tool can be run simultaneously on multiple machines to expedite processing. By classifying data automatically, the amount of human hours spent working on the raw data drop drastically. The amount of additional quality assurance for a USGS 7.5-minute topographic quad is minimal to create accurate 1m raster datasets from the automatically classified lidar data. To make the lidar data even more accessible to researchers, the tools created for raster extraction and canopy metrics have been turned into a Python Toolbox for easy integration into the industry-standard ESRI software suite.

In addition to raster creation tools, we have also developed an interpretation tool which is accessed through a Python Toolbox. This tool, an automatic sampler, can: create multiple buffers from a random set of points at specified distances for use in statistical data collection; work around incorrect statistics when sampling rasters with overlapping polygon masks; calculate statistics for both binary and non-binary rasters; merge and join all tables together with unique naming convention to make further study of the data easier seamless. This tool has been used to examine and error-check our lidar-derived datasets, however it is not limited to this alone - it can be used to greatly simplify and automate any task where large amounts of spatial statistics must be collected over many raster variables.

3.3 Creating a Scheduler

Many commonly used pre-existing geoprocessing tools currently can only run a single task on a single core. This software shortfall greatly limits the speed of processing on our highly-resolved lidar-derived raster datasets. For computation-heavy analyses we have created a time-saving scheduling tool to optimize both speed and computation of large datasets.

The scheduler splits larger rasters into smaller tiles. The size of these tiles was heuristically determined and allows for geoprocessing tasks to be run which would take exponentially longer on larger rasters. Next, the scheduler adds these tiles to a table within a File Geodatabase which keeps track of the directory path of each tile. When multiple instances of the process are open, this tile list keeps track of what has not been worked on, what is currently processing, and what is completed. The number of processes run is dependent on hardware limitations - in our study we are able to speed up the process by a factor of 13.44 (tab. 1). Though simple in concept and construction, this tool allows us to take full advantage of hardware by bypassing single-core software limitations.

# of Processes Run	Estimated Minutes	Estimated Days	Factor vs. 1 Process
1	16082.62	11.17	1
4	4071.93	2.83	3.95
8	2122.80	1.47	7.58
12	1646.57	1.14	9.77
16	1303.56	0.91	12.34
20	1196.46	0.83	13.44

Table 1. Increased processing time based on an estimation of 18 7.5-minute USGS topographic quads worth of lidar data needed to cover the Portland Metro Area.

3.4 Integration with Planning Practices

As seen in fig. 3, the data created from the lidar processing is far more resolute than commonly available data. We combine these highly resolved metrics with built environment data, American Community Survey data, and observational data to better understand the link between urban tree canopy and the role it plays in mitigating urban air pollution (Rao 2014) and the urban heat island effect. Our automatic sampling tool has already been successfully used by researchers to efficiently compare observational point data to a multitude of raster datasets.

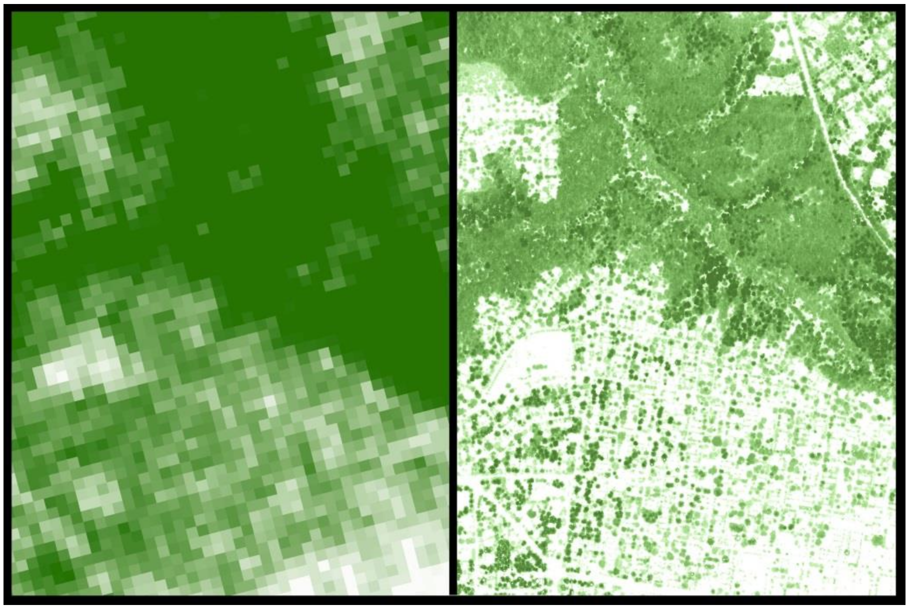

Figure 3: Two examples of data for the same rough geographic area. *Left*: 30m Canopy Percent (Xian 2011), *Right*: Our 1.0m Lidar Canopy Density Metric.

4. Acknowledgements

We wish to acknowledge the Institute for Sustainable Solutions for their support in our study of the urban environment, and the Oregon Lidar Consortium for their efforts in organizing the funding for, collection, and dissemination of high-quality lidar data to the public.

5. References

Dallimer M, Irvine KN, Skinner AMJ, Davies ZG, Rouquette JR, Maltby LL, Warren PH, Armsworth PR and Gaston KJ, 2012, Biodiversity and the Feel-Good Factor: Understanding Associations between Self-Reported Human Well-being and Species Richness. *BioScience*, 62: 47-55.

Donovan G H, Butry D T, Michael Y L, Prestemon JP, Liebhold AM, Gatziolis D, and Mao MY, 2013, The relationship between trees and human health: evidence from the spread of the emerald ash borer.*American Journal of Preventive Medicine*, 44(2): 139–45.

Donovan GH and Prestemon JP, 2010, The Effect of Trees on Crime in Portland, Oregon. *Environment and Behavior*, 44(1): 3–30.

Escobedo JF, Kroefer T and Wagner JE, 2011, Urabn forests and pollution mitigation: Analyzing ecosystem services and disservices. *Environmental Pollution*,159: 2078-2087.

Metro, 2015, http://rlisdiscovery.oregonmetro.gov/

NLCD, 2011, http://www.mrlc.gov/nlcd2011.php

OLC, 2014, http://www.oregongeology.org/sub/projects/olc/

Rao M, George LA, Rosenstiel TN, Shandas V and Dinno A, 2014, Assessing the relationship among urban trees, nitrogen dioxide, and respiratory health. *Environmental Pollution*, 194: 96-104.

United Nations, 2014, *World Urbanization Prospects*.

Xian G, Homer C, Dewitz J, Fry J, Hossain N and Wickham J, 2011, The change of impervious surface area between 2001 and 2006 in the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 77(8): 758-762.