Identifying rockfall hazards in the Fraser Canyon, British Columbia: a semi-automated approach to the classification and assessment of topographic information from airborne LiDAR and orthoimagery

By

Richard D. Carter

A thesis submitted to the Department of Geological Sciences and Geological Engineering
In conformity with the requirements for
the degree of Master of Science

Queen’s University
Kingston, Ontario, Canada
(September 2018)

Copyright © Richard D. Carter, 2018
Abstract

Rockfall hazards on railway corridors create risk of derailment which can result in damage to property or the environment and cause injury or loss of life. There is interest in understanding the location and severity of such hazards so that management strategies can be implemented. Ongoing collection of high resolution 3D data from terrestrial and airborne platforms has proven to be an effective medium for achieving a better understanding of the spatial-temporal, geological and geomechanical properties of rockfall hazards. The limiting factor is that such data is time consuming to collect and process and needs to be specifically directed. This study aims to provide new office-level screening tools which can be used to identify areas where more refined data collection may offer value.

Airborne LiDAR and orthoimagery for the CN Thompson Fraser Corridor is used to develop new techniques for the classification and interpretation of topographic data for the study of rockfall hazards in the area. In Chapter 3, spectral reflectance in the visible range, extracted from orthoimagery, is used to classify colourized point cloud data in the interest of identifying areas of exposed rock. In Chapter 4, this technique is paired with existing techniques for the geomorphic classification of 3D information using slope angle and use as the main input to a probabilistic rockfall hazard assessment.

Analysis of colour values, paired with conventional morphometric analysis using slope angle, shows promise as a means of classifying topographic data into geomorphic domains. The study successfully derives all inputs for the probabilistic assessment of rockfall hazard from the above mentioned airborne LiDAR and orthoimagery dataset.
Co-Authorship

This thesis is the product of the formal research of Richard Carter. The research was part of a collaborative project and technical guidance was contributed by Dr. Jean Hutchinson and Dr. Dave Gauthier, however the written work is solely that of the author.
Acknowledgements

This thesis is dedicated to my late Grandma, Annie, who, at the age of 96, had no earthly idea about what I was doing back in school but thought it was just great that I get to see the world while doing it.

This research would not have been possible without the support and dedication of Dr. Jean Hutchinson and Dr. Dave Gauthier. Jean, your tireless dedication to education and research is both inspiring and humbling. Thank you for all the opportunities you’ve put in front of me over the last couple of years and for your patience as I’ve developed my knowledge of our field, to which I’m still new. Dave, I cannot properly articulate the level of patience, creativity and intelligence you’ve displayed throughout this process. You’ve been a mentor, a friend and an advocate for me, as well as my work and I’m extremely grateful for that. Further, the hospitality extended to me by you and Rhonda early on will not soon be forgotten.

BGC Engineering has been an integral supporter of my graduate studies, providing me with support by means of funding, resources, data and a leave of absence for me to do this work. There are too many people in the organization to thank by name, but I’d specifically like to mention Dr. Matt Lato as he is the reason I ended up at Queen’s in the first place.

The research was also largely supported by the Railway Ground Hazards Research Program. I’d specifically like to thank Trevor Evans from CN for his willingness to provide expert knowledge, data, site access and anything else we’ve needed over the last couple of years.
David and Zac; thanks for all the fun times and engaging conversations we’ve had over the last couple of years in so many different places. It really feels like we went through this process together, and I’m thankful for that. Looking forward to continuing to work together in the future.

Thank you to the other RGHRP folks at Queen’s for your ongoing support for my work and our overall contribution to the field; Emily Rowe, Megan van Veen, Dr. Ryan Kromer, Alex Graham. Let’s keep this going!

Dr. Mark Diederichs, though not directly involved in my research, has provided me with a ton of support in ensuring that I received the best education possible in my brief time at Queen’s. Rob Harrap has also been a constant source of wisdom and books during this time.

Niki, your patience and support have been everything to me over the last year and a half—but especially in the last few weeks. Thanks for being there to get me through this.

I wouldn’t be the person I am without the ubiquitous role my family has played in my life. My parents, Cam and Rose, have supported me in any way they can in whatever endeavor I’ve undertaken up to and including becoming a full-time student again at the age of 30. A simple “thank you” doesn’t do enough justice to the gratitude I owe these two individuals. Vic and Scott, you’re okay too.

Finally, thank you to all of those I’ve not mentioned by name but have crossed paths with along the way.
# Table of Contents

Chapter 1: Introduction ........................................................................................................ 1
  1.1 Project Overview ................................................................................................ 1
  1.2 Thesis Scope and Objectives ............................................................................. 1
  1.3 Outline ................................................................................................................ 3
  1.4 Summary of Key Findings .................................................................................. 4
  1.5 References ......................................................................................................... 6

Chapter 2: Background and Literature Review ................................................................ 7
  2.1 Rockfall 101 ........................................................................................................... 7
  2.2 The use of remote sensing in the study of rockfall ............................................... 10
  2.3 Searching and Prioritization ............................................................................. 12
    2.3.1 Slope Morphometry ....................................................................................... 14
    2.3.2 Forest Cover ................................................................................................. 16
  2.4 Characterization ............................................................................................... 17
  2.5 Monitoring ........................................................................................................ 19
  2.6 Context of Current Work .................................................................................. 20
  2.3 References ....................................................................................................... 21

Chapter 3: Classifying terrain using the colour values found within colourized 3D pointclouds derived from LiDAR and orthoimagery ....................................................... 29
  3.1 .............................................................................................................................. 29
  3.2 Introduction .......................................................................................................... 29
    3.2.1 K-Means Classification .................................................................................. 32
    3.2.2 Random Forest Classification ........................................................................ 33
    3.2.3 L*a*b Colour Space ....................................................................................... 33
  3.3 Study Area ........................................................................................................... 34
  3.4 Methods and Data .............................................................................................. 37
    3.4.1 Creation of manually classified sample dataset ............................................. 39
    3.4.2 Automated classification of sample dataset................................................... 40
    3.4.3 Measurement of classification success .......................................................... 42
3.4.4 Application of automatic classification to dataset for entire study area ........ 43
3.5 Results .................................................................................................................. 44
3.5.1 Sample dataset .................................................................................................. 44
3.5.2 Application to Study Area Dataset .................................................................. 46
3.6 Discussion ............................................................................................................. 51
3.7 Conclusions ......................................................................................................... 61
3.8 References ........................................................................................................... 64

Chapter 4: Assessing Rockfall Hazard using Semi-Automated Processing Techniques
with LiDAR and orthoimagery: Case Study from CN Railway's Yale Subdivision .......... 67
4.1 Introduction ........................................................................................................... 67
4.2 Engineering Geology of the Fraser Canyon ....................................................... 75
4.3 Methodology ....................................................................................................... 77
4.3.1 Terrain Classification ..................................................................................... 79
4.4 Results and Discussion ....................................................................................... 93
4.4.1 Terrain Classification Results ....................................................................... 93
4.4.2 Assessment of Forest Cover Results ............................................................. 97
4.4.3 Slope-Scale Analysis Results ....................................................................... 97
4.4.4 Hazard Prioritization Results ....................................................................... 98
4.5 Discussion .......................................................................................................... 105
4.6 Conclusions .................................................................................................... 110
4.7 References .................................................................................................... 112

Chapter 5: Summary and Future Work ................................................................. 116
5.1 Goals and objectives revisited ....................................................................... 116
5.2 Colour analysis for the segmentation of 3D Data ............................................ 117
5.3 Combination of Colour Analysis Results with Slope Angle Distribution Method. 120
5.4 Delineation of maximum rockfall runout boundary for each source identified through SAD/Colour analysis ................................................................. 123
5.5 Slope-scale assessment for the runout zone of each source ............................. 123
5.6 Quantification of rockfall hazard at track level by means of parameters generated through this process ................................................................. 124
5.7 Future Work .................................................................................................... 125
5.8 Conclusions ........................................................................................................... 126  
4.8 References ........................................................................................................... 128  
Appendix A: Colour Classification Testing ............................................................... 129  
Appendix B: Code for Calculating Protective Capacity of Forest ....................... 133  
Appendix C: GIS Dataset Overview ...................................................................... 136
List of Figures

Figure 2-1 Framework for the identification, characterization and ongoing monitoring of rockfall using data collected through various remote sensing platforms. ................................................................. 13

Figure 3-1: Depiction of L*A*B colour space, adapted from Bora et al (2015) 34

Figure 3-2: Plan map of study area and extents of airborne LiDAR and orthoimagery data used for this study (highlighted in pink). Background image is ESRI base map satellite data. 36

Figure 3-3: True-colour 3D point cloud of the study area. Inset, a photo of study area taken from helicopter inspection in June 2018 (photo courtesy of CN Rail). ............... 37

Figure 3-4: a diagram highlighting the main components of methodology employed in this study. The inputs are derived from the airborne LiDAR and orthoimagery survey mentioned above ....................................................................................................................................... 39

Figure 3-5: scatter plot of the manually classified validation dataset showing their 2D location along the A and B colour scales along the X and Y axes respectively ........ 47

Figure 3-6: Results of K-Means classification carried out on manually classified dataset. Point colours represent the true colour of the point extracted from the orthoimagery. The coloured areas represent the boundaries delineated through the K-Means classification ........................................................................................................................................ 48

Figure 3-7: A-RGB image of one section of track within the study area. B-results of unsupervised K-Means classification. C-results of Random Forest classification. .... 52

Figure 3-8: Zoomed in images of the boxes highlighted in Figure 8; for each box, an RGB image as shown in addition to the results of both the K-Means and Random Forest classification analyses. .................................................................................................................. 54
Figure 3-9: This plot shows the point cloud for the entire study area in A*B space. In red, are points where Random Forest classified a rock and K-Means classified not rock. In orange are points where K-Means classified a rock and Random Forest classified not rock. The orange points in the top right corner are mostly attributed to lighter coloured vegetation. The cause for the error elsewhere is harder to pinpoint, but the linear nature of these patterns suggest that there is some overfitting happening in the Random Forest classification. .......................... 55

Figure 3-10: a-Colourized LiDAR point cloud of study site, with line denoting plan-view of section A-A’ (shown in Figure 10b. b-Section A-A’ consisting of terrain classification and non bare earth LiDAR returns. Points 1 and 2 are highlighted in both A and B and demonstrate how colour classification paired with information regarding slope angle can be used to interpret geomorphic setting for specific sites. The inset in the top left corner of Figure 9B zooms in on area 1 to further highlight the impact of slope angle on superimposing 2D colour into 3D space. The horizontal line at the top of 3-9b is the 2D representation of the colour classification. .......................... 56

Figure 3-11: a) orthoimagery of an area of slope in the Yale subdivision. b) slope angle. c) the vertical distance between bare earth and non-bare earth returns. d) the results of the colour analysis carried out here. e) instances of negative vertical differences resulting from surficial roughness, mostly occurring on areas of high gradient. The purple areas shown in each image are to provide reference for comparison. ............... 60

Figure 4-1: A process developed for the identification, characterization and monitoring of rockfall hazard through the use of remote sensing. The information derived within each step of the process informs the best decision for the next. Broad suggestions for spatial and temporal resolution required to achieve various outcomes are also noted. .......................... 72
Figure 4-2: The risk equation as defined by Porter and Morgenstern (2013). In the middle of the
diagram, for each component of the risk equation, there is a single question related to the
evaluation. On the right, an explicit link between the parameters of the risk equation and this
study is made. This study works to inform the risk equation up to the point of a rockfall reaching
the track, impact and consequence are not considered here. .................................................. 74

Figure 4-3: Plan map of study area—Mile 20–25 of the CN Yale Subdivision, British Columbia,
Canada ........................................................................................................................................ 76

Figure 4-4: High-level overview of the workflow developed in this study. The analysis starts at
the regional scale and uses the individual runout zones to bound analysis at the slope scale for
each potential rockfall source........................................................................................................ 79

Figure 4-5: Gigapixel image of a slope at Mile 22.3 of the Yale subdivision, with potential rockfall
zone of influence identified in purple............................................................................................. 81

Figure 4-6: Visual depiction of the process of combining 3D information from ALS and 2D
information from orthoimagery in order to create a colourized 3D point cloud which can be used
as the input to the unsupervised colour analysis. ....................................................................... 84

Figure 4-7: bare earth LiDAR DEM with annotations which define the features identified through
the Slope Angle Distribution method............................................................................................ 85

Figure 4-8: Top: Bare earth DEMs of two different slopes within the study area. Bottom: Non-
bare earth Returns on top of bare earth returns for same two slopes. Both sites are shown at the
same scale and range from roughly 10 to 700 m a.s.l................................................................. 91

Figure 4-9: Results of the slope angle distribution decomposition. The intersection points shown
represent the cutoff angle for each morphological unit (Flats/Benches:0 -17°; Fans:17 -30°;
Talus Slopes: 30 -50° and Rockfall Sources:50 -90°). ................................................................. 94

Figure 4-10: Results of morphometric classification using the decomposition of slope angles... 95
Figure 4-11 Histogram showing the distribution of slope angle values within the manually classified dataset of 45696 points, discussed in Chapter 3. The black dashed lines and labels represent the results of the Loye et al (2009) method. ................................................................. 96

Figure 4-12: Top-raw orthoimagery for subset of study area around Mile 30 on the CN Yale Subdivision. Bottom-results of classification methodology combining the Slope Families concept with the use of colour for identifying areas of exposed rock. This figure demonstrates how colour can be used in tandem with the Slope Angle Distribution Method to further refine the terrain classification; this allows for the identification of exposed outcrop and exposed talus deposits rather than just areas which meet the slope angle criteria for a given classification. .................. 100

Figure 4-13: a-Output of FiNT (Dorren, 2018). Each point represents a tree, for which FiNT outputs a height and stem diameter. b-cross section B-B’, black dots represent points placed by FiNT, green points represent non-bare earth LiDAR returns. ................................................................. 102

Figure 4-14: Results of the runout zone delineation process for 4 cliffs identified through the terrain classification process. Runout ID is the unique ID given to each source at the end of the terrain classification process (Top image is over a hillshade DEM derived from ALS; bottom is over orthoimagery). ........................................................................................................ 103

Figure 4-15: On the left, the results of the colour classification, right, a graph showing the probability of hazard reaching a track given forest is shown in red while the reduction factor calculated using the Berger and Dorren (2007) method. ................................................................. 104

Figure 4-16: Illustration of how this system can be used in a backwards analysis in the event of a rockfall landing on track. A-the green lines are the extents of all runout cones which intersect the location of the fallen block. These are used in order to find the potential sources which could have released the block, shown in B. C shows the extents of a cone extending +/-20 degrees from the fallen block to all sources which could have been responsible for the block. D shows
these cliffs, coloured by lateral divergence from the fallen block; this is to demonstrate how further research could use the information.

Figure 5-1: A-Bare earth LiDAR model– this provides the 3D geometry for all terrain classification carried out. B-Othoimagery– this is where all colour information is derived from. C-the results of the colour analysis providing a fully classified 3D pointcloud

Figure 5-2: Results of colour analysis from Chapter 3, applied to an area in the Yale Subdivision. Red areas are exposed rock, while green areas are not.

Figure 5-3: Results of Loye et al (2009) Slope Angle Distribution method

Figure 5-4: The results of the combined approach of using slope angle and colour to classify terrain

Figure A-1: Results of K-Means Classification on a randomly generated sample dataset of 2500 points

Figure A-2: Scatter plot of Random Forest classification on randomly generated sample dataset of 2500 points, 95% total accuracy

Figure A-3 Results of Random Forest model built using incrementally larger sample data sets, applied to the entire point cloud of 7 million points

Figure A-4: Results of K-Means Classification on the full dataset for the study area. Cluster centres for both the validation dataset and the full dataset are shown.
List of Tables

Table 2-1: taken from Higgins and Andrew (2012) a general overview of rockfall failure mechanisms and their common geological settings ................................................................. 8

Table 2-2: Overview of most common rockfall triggers and influence on rockfall from Higgins and Andrew (2012) ........................................................................ 9

Table 3-1: Contingency table from testing K-Means clustering with different proportions of manually classified dataset 49

Table 3-2: Contingency table from Random Forest classification using different proportions of the manually classified dataset to train the model. Each model is applied to all 44690 points. Here, ‘n’ refers to the number of points used to train the model. .... 50

Table 4-1: Parameters required for the calculation of risk reduction offered by forest cover as suggest by Berger and Dorren (2007). 86

Table 4-2: Main functions of Flow-R model ................................................................... 89

Table 4-3: Parameters used in Flow-R model for this study in order to delineate maximum runout boundaries for each potential source identified through the terrain classification process. ........................................................... 89

Table 4-4: summary information derived through downslope analysis for four rockfall runout zones shown in Figure 11. ............................................................ 101

Table 5-1: Classification results of both approaches to classification of points 120

Table 5-2 Summary of results of slope angle distribution analysis ......................... 121
List of Abbreviations

ALS     Airborne Laser Scanning
CN      Canadian National (Rail)
CP      Canadian Pacific (Rail)
DBH     (Tree) Diameter at Breast Height
DEM     Digital Elevation Model
FiNT    Find Individual Trees
LiDAR   Light Detection and Ranging
OHP     Oblique Helicopter Photogrammetry
RGHRP   Railway Ground Hazard Research Program
RHRA    Rockfall Hazard and Risk Assessment (CN)
RHRS    Rockfall Hazard Rating System (Oregon)
SAD     Slope Angle Distribution
SfM     Structure From Motion
TLS     Terrestrial Laser Scanning
TRB     Transportation Research Board
Chapter 1: Introduction

1.1 Project Overview

The rugged, mountainous terrain of Western Canada is conducive to rockfall formation and release. Rockfalls are the detachment of a block from a rock mass which then propagates down slope (Cruden and Varnes, 1996; Higgins and Andrew, 2012). Along railways, rockfalls can result in derailments which can pose risk to human life, infrastructure as well as the natural environment. Moreover, rockfalls make up a significant portion of economic losses caused by slope hazards along railway corridors.

The Railway Ground Hazard Research Program (RGHRP), established in 2003, is a collaborative research initiative involving CP Railway, CN Railway, Transport Canada, The University of Alberta and Queen’s University. The program’s objective is to advance our understanding of ground hazard processes as they relate to railway operation, and develop new techniques for identifying, characterizing and monitoring them. The Queen’s University section of the RGHRP group specifically looks at rockfall hazards and since 2007 has been focused on exploring the use of remote sensing technology in rockfall hazard management. The work carried out for this thesis deals specifically with regional-scale rockfall hazard assessment using airborne LiDAR and orthoimagery data and is part of the Queen’s component of the RGHRP.

1.2 Thesis Scope and Objectives

There are two main components of this research. First, we examine the utility of spectral reflectance in the visible range (colour) in orthoimagery as a means of identifying
topographic features and classify terrain for the purposes of rockfall hazard identification and assessment. The accuracy and uncertainty associated with deploying automated classification techniques using colour, for these purposes, is assessed. Two different statistical classification approaches are examined to determine how to best go about using colour data in this manner.

The second component is an application of this technique to a rockfall hazard identification study on the CN Rail Yale subdivision. The findings from the first part of the research, which are reported in Chapter 3, are applied to the creation of a fully classified spatial dataset for the study area containing geomorphic classification on the basis of both slope angle and colour. This is used, in Chapter 4, as the main input to a study aimed at using surficial characteristics to identify evidence of rockfall activity. The use of forest cover as a mechanism for rockfall protection is also explored.

The overarching goal of this project was to improve our understanding of what information can be derived from a single LiDAR/orthoimagery survey for regional rockfall hazard assessment for natural slopes (i.e., rather than anthropogenically cut slopes). Conventional assessment has involved rock mass characterization and hazard identification in the field or hazard scale data collection using terrestrial laser scanning or photogrammetry; both of these approaches have proven to be effective. Recent advancements in LiDAR equipment, as well as data storage and computational capability have made it possible to collect and interpret vast amounts of high resolution spatial data at an elevated level of detail. Moreover, these advancements have led to increased accessibility of advanced statistical algorithms which were previously only employed by computer scientists and statisticians. These capabilities were leveraged
throughout this research. The full resolution of the data collected is maintained and analysis is carried out in 3D space wherever possible.

1.3 Outline

A review of literature regarding rockfall hazards and the use of remote sensing in geoscience and geological engineering is provided in Chapter 2. Derivatives of topographic data from LiDAR, namely slope angle and forest metrics, are explored as along with their utility in the study of rockfall hazards. This chapter was put in the context of a framework for the systematic collection of 3D topographic data based on the goals of the study; this framework guides the spatial and temporal resolution requirements for data based on the desired outcomes of the work being undertaken.

Chapter 3 and 4 contain original research. In Chapter 3, a methodology for the analysis of colour as a means of classifying topographic data was presented and tested. In Chapter 4, this methodology was combined with work by Loye et al (2009) which uses slope angle to classify topographic data into geomorphic units. Flow-R (Horton et al, 2013) was used to delineate rockfall runout boundaries for rockfall sources identified in the above-mentioned terrain classification exercise. In addition, individual trees were identified using FiNT (Dorren, 2017). These outputs were used as inputs into an assessment at the slope scale, which is defined by the extents of the runout zone for each source. Using these spatial extents, all terrain down slope of a given source was assessed for evidence of rockfall activity and the methodology from Berger and Dorren (2007) was applied to quantify the impact of forest cover on the probability of rockfalls from a given location reaching the track. All of this information was combined in a probabilistic hazard assessment exercise aimed at informing the hazard components of
the risk equation developed by Porter and Morgenstern (2012); the end result is a partial risk rating for each source zone.

A summary of the findings and comments on the relative strengths and weaknesses of the methods utilized in this work is provided in Chapter 5. Based on this, future research opportunities which can build on the work carried out here will be discussed.

1.4 Summary of Key Findings

In Chapter 3, colour was determined to be an effective means for classifying topographic data and a methodology for doing so is presented. After testing K-Means clustering (MacQueen, 1967) and Random Forest classification (Breiman, 1996) it was determined that the Random Forest Classification results in more accurate classification and generates errors which are more predictable and less abundant than those generated through K-Means clustering. Therefore, it was concluded that the output of the Random Forest Classification would be a suitable input for the geomorphic classification of terrain in the interest of identifying and characterizing rockfall hazards at the regional scale.

In Chapter 4 a methodology was developed for rockfall hazard assessment using only inputs derived from a single airborne LiDAR and orthoimagery dataset. Simple rockfall runout modelling is utilized in order to define the extents of the slope scale for each rockfall source and a preliminary demonstration of how the role of forest cover in rockfall hazard susceptibility can be examined in a semi-automated fashion at the regional scale. The outcomes of the analysis presented in Chapter 4 are listed below and are included as Appendix C within this document:
• A raster dataset containing the results of the terrain classification using slope angle and colour analysis.

• A GIS dataset containing individual trees and their dimensions for the entire study area.

• A GIS dataset containing the runout zone polygon for each source identified through terrain classification.

• The results of the slope scale analysis, a GIS database containing attributes related to the terrain, slope geometry and forest cover for each runout zone.

• The tabular results of the probabilistic assessment of rockfall hazard for the study area using the above-mentioned GIS database.
1.5 References


Chapter 2: Background and Literature Review

2.1 Rockfall 101

The rugged, mountainous terrain of Western Canada is conducive to rockfall, which can pose risk to human life, infrastructure as well as the natural environment. Rockfalls are the detachment of a block from a rock mass which then propagates down slope (Cruden and Varnes, 1996; Higgins and Andrew, 2012). For railroads, rockfalls are a constant problem (Martin, 1988). Transportation corridors which traverse mountainous terrain are particularly exposed to rockfall hazards (Piteau and Peckover, 1978).

To be capable of sourcing rockfalls, a slope must be sufficiently steep in order to expose the discontinuities or erodible material which bound rock masses (Higgins and Andrew, 2012). Furthermore, the configuration of these features guide what types of rockfall mechanisms are possible for a given slope. For example, planar slides are structurally controlled and require that the slope face and discontinuity plane be dipping in the same direction. Wedge failures, on the other hand, require that multiple discontinuity planes intersect along a line which plunges out of the slope at an angle steep enough to sustain movement; failure occurs along that intersection. Table 2-1, taken from Higgins and Andrew (2012), provides a general description of common rockfall failure modes. Regardless of failure mechanism, for rockfall to occur, the slope below the rock mass must be steep enough to sustain movement despite the presence of vegetation and other features which may absorb the energy of the initial failure (Higgins and Andrew, 2012).
Table 2-1: From Higgins and Andrew (2012) a general overview of rockfall failure mechanisms and their common geological settings

<table>
<thead>
<tr>
<th>Failure Mode</th>
<th>Failure Description</th>
<th>Susceptible Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple structurally controlled planar sliding</td>
<td>Sliding of one or multiple rock blocks on a single discontinuity surface</td>
<td>Rocks with discontinuities daylighting in and striking nearly parallel to the slope face. All rock types may slide on joint sets; sedimentary and volcanic rocks may slide on weak beds or bedding planes; metamorphic rocks may slide along foliation.</td>
</tr>
<tr>
<td>Wedge sliding</td>
<td>Sliding of one or multiple blocks along the line of intersection of two non-parallel planes</td>
<td>Rock masses with well-defined orthogonal discontinuity sets, foliation bedding planes or weak beds in which the intersection of two discontinuities plunges out of (daylights) a slope face. May occur in any igneous, sedimentary and metamorphic rocks.</td>
</tr>
<tr>
<td>Toppling</td>
<td>Rotation of columns or block(s) of rock about a fixed base and out of the slope</td>
<td>Hard rocks with regular, parallel discontinuities dipping steeply into the slope, with or without cross joints. Common in foliated metamorphic rocks and steeply dipping layered sedimentary rocks, block-jointed granites, steep faces of columnar basalts, and thinly bedded slate.</td>
</tr>
<tr>
<td>Circular sliding</td>
<td>Failure occurs along a surface that approaches a circular shape and is not a zone of particular weakness, merely the line of least resistance</td>
<td>Highly fractured or low intact-strength material; highly fractured rocks with so many discontinuities that it is effectively homogenous; weak sedimentary rocks with bedding that does not dip out of slope; clay-rich rocks softened by loosening and wetting; smectite-rich clay shales and altered tugs, rocks adjacent to major fault zones, and coal-bearing formations.</td>
</tr>
<tr>
<td>Complex structurally controlled buckling and kick-band slumping</td>
<td>Compressional failure of columns or slabs parallel to the rock slope face.</td>
<td>Thiny bedded, weak sedimentary rocks inclined steeply and parallel to the slope surface; shale-sandstone and shale-chert sequences, coal measures and foliated metamorphic rocks.</td>
</tr>
<tr>
<td>Block Torsion</td>
<td>Failure block spins about a hinge where sliding has been impeded</td>
<td>Blocky rock that contains a series of discontinuities and obstructions.</td>
</tr>
</tbody>
</table>
Rock slump Backward rotation of single or multiple blocks as they fail in series, resembling a classic soil slump Materials with vertical and horizontal zones of weakness, allowing for sliding on the basal rupture surface and between blocks or along bedding at the same time. Commonly, sedimentary units or stratified units.

Rockfall triggers are the external factors which influence the instability of rock mass and the initiation of movement (Pantandelis, 2009). Climatic factors such as freeze thaw and high levels of precipitation have been found to have the most substantial impact on rockfall frequency (Piteau, 1977; Bunce et al, 2006; Manciotta, 2015;). Table 2-1, from Higgins and Andrew (2012) summarizes the main triggering processes relevant to rockfall. This table lists each triggering factor and provides a brief explanation of their influence over rockfall frequency.

Table 2-2: Overview of most common rockfall triggers and influence on rockfall from Higgins and Andrew (2012)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>● Has the potential to cause erosion of finer grained material, leaving larger blocks unstable; and&lt;br&gt; ● Causes recharge of groundwater leading to increased pore pressures within the slope.</td>
</tr>
<tr>
<td>Snowmelt</td>
<td>Same influences as precipitation, although more likely to occur when the rock slope surface is frozen and therefore the water backs up.</td>
</tr>
<tr>
<td>Freeze-Thaw</td>
<td>● As water infiltrates into discontinuities and freezes, it undergoes a volume increase, leading to increased pressures; &lt;br&gt; ● Can cause opening of cracks and joints, pushing rocks out from the slope face; and&lt;br&gt; ● Most critical times are late fall/early winter, and late winter/early spring when temperatures are hovering near zero.</td>
</tr>
<tr>
<td>Wind</td>
<td>● Can cause erosion of small particles, leaving larger blocks unsupported; &lt;br&gt; ● Can cause loosening of tree roots, triggering rockfalls</td>
</tr>
</tbody>
</table>
Particles eroded more easily on south-facing slopes due to dry conditions.

Human/Animal Activity
- Damage to rock face can be caused by poor blasting practices (especially on slopes that were blasted before modern techniques were developed);
- Vibrations induced by blasting can cause instabilities;
- Vibrations can also be caused by train movement and the use of heavy construction equipment; and
- Rockfalls can also be initiated by the movement of animals walking on a slope.

Earthquakes
- Long-duration vibrations and high accelerations from earthquakes have the potential to trigger rockfalls.

2.2 The use of remote sensing in the study of rockfall

In this study, regional-scale airborne LiDAR and orthoimagery datasets were used for the identification and characterization of rockfall prone areas. In addition, one of the broader goals of this work was to guide the study of rockfall processes using remote sensing data collection; this can be through terrestrial or airborne platforms. Therefore, it is important to establish working knowledge of available remote sensing platforms and their capabilities; that is done here.

Remote sensing and GIS have been widely utilized in the assessment of rockfall susceptibility. The Parachute Project (Cloutier et al, 2015) highlights the use of remote sensing technology for the purposes of hazard assessment. Parachute makes use of terrestrial and airborne LiDAR throughout the project to identify and characterize hazards through terrain analysis as well as the assessment of structural geology and rockfall susceptibility rating. This study made use of slope angle to identify potential rockfall sources. The aim of the research presented here was to determine how other
information, such as colour, could be used to further refine this terrain classification process.

LiDAR is commonplace in geoscience and geological engineering (McCaffrey et al 2005). Specifically, it has proven to be a useful tool for the identification, characterization and monitoring of slope hazards (Lim et al. 2005; Rosser et al. 2005; Kromer et al. 2015; Kromer et al. 2017). Lato et al (2015) demonstrate that Airborne Laser Scanning (ALS), Terrestrial Laser Scanning (TLS), and Oblique Helicopter Photogrammetry (OHP) are all viable technologies for identifying anthropogenic and naturally occurring change in rock masses along transportation corridors and that deciding which of these technologies to employ is a matter of cost, scale and level of detail necessary. Moreover, Lague et al. (2013) and Kromer et al. (2015a) demonstrate the ability to identify the displacement vector of observed change in high resolution topographic data. In response, high-temporal TLS has been largely utilized to characterize rock slope hazards in detail (van Veen et al, 2016) identify structural controls and pre-failure deformation trends (Kromer et al, 2017; Rowe et al, 2017) and quantify movement of blocks on talus slopes (Bonneau et al, 2018).

Figure 2-1 is a framework for rock slope hazard management using remote sensing. The framework highlights how spatial and temporal scale guides what is achievable through analysis of point cloud data and provides a platform for decision making based on the needs of a specific project or availability of data. The framework is intentionally progressive in nature to allow for more refined hazard evaluation at each step. The implicit output at each point in the framework is the location of sites to be used in the next step; that is, the prioritization phase guides which areas need to be
carried through to the characterization phase, while the results at the characterization phase will guide what is useful, possible and attainable at the monitoring stage, which is inherently multi-temporal. The temporal frequency necessary is also guided by the results of the characterization phase. The following text will provide a detailed overview of the use of point cloud data in rockfall hazard assessment using this framework as a reference point to show how different tools and applications fit together in order to achieve the goal of advanced understanding of the spatial and temporal components hazards stemming from rock slope instability.

2.3 Searching and Prioritization

There is no turnkey solution for the study of regional rockfall activity to be applied broadly within a given study area. Rather, the approaches employed by studies aiming to assess regional rockfall hazards are generally determined based on the availability of relevant data, budget and time constraints, as well as the level of detail required. Most studies include the combination of geomorphological and/or geological data, event inventories and climate information (e.g. Baillifard et al, 2003; Frattini et al 2007; Copons and Vilaplana, 2008).
Gauthier et al (2017) state that there are two questions that need to be answered in the preliminary stages of any framework aimed at assessing rock slope hazards using
remote sensing: where are rockfalls possible, and where are rockfalls likely? The Searching and Prioritization section of the framework outlined in Figure 2-1 aims to answer these questions at the regional scale to highlight sites which should be reassessed at a more refined scale. In this study, surficial conditions were assessed in order to identify geomorphic features which are indicative of rock fall, i.e. rockfall sources and rockfall deposits.

2.3.1 Slope Morphometry

The analysis of topographic characteristics to classify terrain is broadly defined as morphometry. Several morphometric approaches are relevant to rockfall assessment. Slope angle has been used extensively as a means of identifying rockfall sources and deposits (e.g. Loye et al, 2009; Michoud et al, 2012; Ansari et al, 2016; Fanos et al, 2018). Loye et al (2009) analyzed the distribution of slope angles within a DEM to define morphological units such as talus slopes, steep slopes and rockfall sources. They found that the regional-scale susceptibility trends determined through this method are consistent with field observations made regarding rockfall activity in the canton of Vaud, Switzerland. The principles of this method are founded on work by Strahler (1950) which suggests that the slope angles of a given morphological unit in a given area will consistently be distributed around a prominent mean slope angle value for that unit. Therefore, the method decomposes the distribution of slope angles for a study area into as many normal (Gaussian) distributions as there are geomorphological units. Michoud et al (2012) combined Loye’s work with runout modelling and found that, together, the two concepts provide a viable, low-cost solution for conventional rockfall susceptibility assessment without using field investigation.
Beyond slope angle, other topographic data can be generated which help to define the nature of rockfall hazard in a given area. For example, Copons and Vilaplana (2008) classify slopes into units according to the properties indicative of instability such as orientation of discontinuities, joint characteristics, weathering and seepage. Moreover, Jasiewicz & Stepinski (2013) propose a DEM-based approach to terrain classification using simple pattern recognition surrounding a point. The patterns they identify are called *Geomorphons* (abbreviation of geomorphologic phonotypes). Using this method, the authors were able to automate the process of delineating flats, peaks, ridges, shoulders, spurs, slopes, pits, valleys, foot slopes and hollows. Similarly, Iwahashi and Pike (2007) use a combination of slope angle, surface texture and convexity to classify terrain into the same groups. Understanding the location of these types of features is important in the study of rockfall, though the link is not explicitly made in either of the works cited.

Morphometric assessment is of value in the Searching and Prioritization phase of the framework outlined in Figure 2-1 in that the output will implicitly answer the question of where rockfalls are possible and provide some of the necessary context for assessing whether or not they are probable. These two questions are central to the work being presented here; techniques for semi-automated interpretation of colourized LiDAR data to classify terrain were presented. The output of this process was subsequently used as the primary input into a hazard assessment workflow for the study which explicitly linked terrain features to rockfall activity.
2.3.2 Forest Cover

In the context of natural slopes, forest cover is an important element to consider as it can have a substantial impact on the effect of rockfall on infrastructure. Forest cover can potentially interfere with the momentum of a rockfall and reduce energy or stop the propagation downslope altogether. In the alpine regions of Western Europe, this concept has been embraced for several centuries and protection forests (Turner and Duffy, 2012) have been planted and managed.

Prior to 2002, there was a lack of empirical data regarding the relationship between forest cover and rockfall propagation (Turner and Duffy, 2012). At this time the ROCKFOR Project group (Berger et al, 2002) carried out a number of field experiments which determined the mechanisms by which trees dissipate rockfall energy. The study found that not only do trees absorb rockfall impacts directly by means of several mechanisms, but the interaction between trees also has an impact; therefore, the distance between trees is an important factor.

Since this time, many authors have explored the protective capacity of forests and developed methods for quantifying it (e.g. Brauner et al, 2005; Berger and Dorren, 2007; Fuhr et al, 2015; Bourrier et al, 2015; Dupire et al, 2016), and tools have been developed with the goal of quantifying the capacity of forest cover to reduce risk created by rockfall (e.g. Brauner et al, 2004; Berger and Dorren, 2007). These studies generally use the size, distribution and density of forest cover along a 2D profile line to calculate the reduction in energy along the line. More recently, Dupire et al (2016) determined that only 3 characteristics are necessary to adequately quantify the protective effect of forest against rockfalls:
1) Total basal area—which is the total amount of tree stem present over a given area.

2) Mean diameter at breast height (DBH)—a common metric used in forestry which defines the diameter of a given tree at about 1.3 m above the ground.

3) Length of forest in maximum slope direction—which measures the amount of forest covered slope in the direction a rock is likely to fall.

The impact of forest cover on the probability of rockfall reaching the railway track within the study area was examined. Within the study area, there is widespread presence of rockfall activity in that it is largely made up of steep natural rock slopes marked by talus deposits at the lower reaches (Piteau, 1977). The ubiquity of evidence of rockfall throughout the study area makes quantifying hazard based on slope geometry and terrain alone a difficult task; most of the study area would likely be rated highly using these parameters. Therefore, it is clear that forest cover is an important factor to consider when attempting to identify where prominent hazards exist which should be addressed by railway operators. This is a channel of information which provides a useful context for the recent rockfall history of a given site.

2.4 Characterization

The characterization phase (in Figure 2-1) assumes that a first pass of data collection with an oblique facing platform (terrestrial LiDAR, terrestrial/helicopter photogrammetry) at the slope scale for the locations flagged through the searching and prioritization phase already exists. This would act as the baseline for ongoing change detection and temporal analysis. The data requirements here are more stringent than
the previous stage; higher resolution data are necessary to identify discontinuities and, in the case of steep slopes, oblique data are required for accurate assessment of structural geology and geomechanical properties. Structural characterization should be undertaken in this stage as identification of the presence (or lack thereof) of discontinuities within the slope. This step allows for further refinement of the results of the Searching and Prioritization phase through manual inspection of data. Areas flagged as being prone to rockfall activity in the Searching and Prioritization phase may be ruled out upon collection and inspection of this initial oblique dataset. Many studies have demonstrated that site-specific rockfall failure mechanisms can be accurately identified through structural characterization of oblique facing topographic data collected using TLS or OHP (e.g. Lato et al, 2012; Rowe et al, 2017 Kromer et al, 2017; Matasci et al, 2018;). Cruden and VanDine (2013) found that small movement can often be observed before block detachment. Kromer et al (2017) demonstrates the ability to observe this pre-failure deformation of individual blocks before they fail through temporal analysis of high-resolution TLS data in order to determine when a block is likely to detach and fall. This stage is crucial in determining an appropriate schedule for further data collection based on the outcome here. For example, seasonal and brittle displacement shows a trend of deformation leading up to failure and would therefore benefit from near real-time monitoring (Crosta and Agliardi 2003). In simple terms; the objective here is to design a schedule for data collection which compliments the processes at hand for a given site.

Along with understanding the mechanisms for a given hazard, knowledge of the potential impact of the hazard is also pertinent. Rockfall runout modelling can be used
to quantify the likelihood of a rockfall of a given volume from a given location reaching an element at risk.

2.5 Monitoring

The information derived within the characterization stage of the framework presented in Figure 2-1 (e.g. failure mechanism, block sizes, density of potential failure masses) dictates whether or not there will be benefit from ongoing data collection for the purposes of change detection and other analyses which require multi-temporal datasets. These factors will also dictate the temporal resolution needed to meet objectives. In the monitoring stage (in Figure 2-1) the assessment of change is used to monitor existing hazards on varying temporal scales. Change detection analysis has been proven capable of detecting change at the centimeter to millimeter scale (Abellán et al, 2010). Abellán (2006) and van Veen (2016) demonstrate how repeated data collection can be used to identify rockfall events. As temporal scale increases, so too does the level of detail, spatial resolution and confidence in change data. The appropriate temporal scale required for a given slope is guided by what processes govern the hazard. Kromer et al (2017) and Rowe et al (2017) demonstrate that pre-failure deformation can be observed for toppling and wedge failures at temporal resolutions ranging from weeks to months and that there is a direct relationship between overserved pre-failure deformation and rockfall volume. Moreover, Kromer et al (2017b) demonstrate that discrete movement events can be observed when near-real-time data collection is deployed.

Another application for change detection is the delineation of failure masses for the purpose of runout modelling. Sala et al (2018) demonstrates how the output of
change detection can be used to identify the block volume and shape as well as commenting on the structural characteristics of the mass. This information can then be used as an input to a runout model.

2.6 Context of Current Work

This research fits predominantly into the Searching and Prioritization, and to a lesser extent, the Characterization phase of the framework in Figure 2-1. Slope angle and colour derived from LiDAR and orthoimagery were used to classify terrain in order to address the question of where rockfalls are possible and identify evidence which could be used to determine where they’re likely. Forest cover was also assessed as it helps to establish the spatial probability of a hazard reaching the element at risk. If there is ample protection offered by forest cover it can likely be prioritized lower than a similar slope with less forest cover. Portions of the work also fit into the Characterization phase. Specifically, coarse runout modeling was used to define a zone of influence for a rockfall source (thereby defining the extents of the slope scale); this information was then used to determine if a rockfall from a given source is likely to reach the track and to characterize evidence of rockfall at the slope scale.
2.3 References


photogrammetry for mapping differential slope change in mountainous terrain, 140(June 2014), 129–140.


26


Chapter 3: Classifying terrain using the colour values found within colourized 3D pointclouds derived from LiDAR and orthoimagery

3.1

3.2 Introduction

Rockfall poses a risk to railway infrastructure in British Columbia where railway corridors traverse rugged mountainous terrain. Topographic and thematic data collected through terrestrial and aerial remote sensing platforms have proven useful in identifying, characterizing and monitoring existing rockfall hazards at the slope scale (e.g. Jaboyedoff, 2010; Lato et al, 2015; van Veen et al, 2016; Abellan et al, 2016; Eitel et al, 2016; Kromer et al, 2017). Rockfall susceptibility has been quantified using regional scale datasets collected using ALS (e.g. Cloutier et al, 2015; Li & Lan, 2015; Fanos & Pradhan, 2016). Slope angle from such datasets has been used to identify rockfall sources and deposits at this scale (e.g. Loye et al, 2008; Loye et al, 2009; Lan et al, 2010; Horton et al, 2013; Cloutier et al, 2015). Lan et al (2010) note that, in addition to slope angle, identification of rockfall sources should include analysis of surficial materials. To date, this has mostly involved the extraction of topographic parameters such as surficial roughness, kernel geometric signature (e.g. Jasiewicz and Stepinski, 2013), contour profiles (e.g. Leschinsky et al, 2015) and the decomposition of the distribution of slope angles (Loye et al, 2009). Hartzell et al (2014) and Miller et al (2017) have demonstrated the utility of hyper and multi spectral satellite data in classifying surficial features such as outcroppings and colluvial deposits.

Increases in computational and data storage capabilities have led to the increased use of advanced statistical analyses on large datasets. In response, these
Capabilities have been leveraged within the geosciences in order to extract detailed spatial information at unprecedented scales from data collected by aerial and satellite-based platforms (e.g. Van Den Eeckhaut et al; Yusof, 2015; 2012; Goetz et al, 2015; Li et al, 2015).

A novel technique for the automated classification of colourized 3D point clouds in terms of topographic features such as areas of exposed rock or vegetation was developed for this study. Our objective was to identify any distinct differences in spectral reflectance in the visible range (colour, here forth) of different surficial features relevant to rockfall hazard which can be identified through statistical analysis of colourized point cloud data. We explored two different algorithms for this: K-Means clustering (MacQueen, 1967) and Random Forest classification (Breiman, 1996). K-Means is an unsupervised clustering algorithm which is simplistic in principle, though can be computationally complex. This algorithm partitions data by iteratively placing a given number (k) of points (cluster centroids) within a dataset and measuring the distance between all points and their closest cluster centroid. The iteration which limits the distance between all points and their closest cluster centre is kept and the dataset is segmented on this basis (each point is clustered with the k-value of its closest cluster centroid). Conversely, the Random Forest Algorithm is a supervised classification technique which requires a training dataset in which each element’s classification is known. The algorithm selects random samples of a dataset and applies randomly compiled decision trees to each sample and tests the success of each tree in classifying the random sample of data. Given the fundamental differences between these two algorithms, we assess their usefulness in classifying terrain using colour, and in doing
so developed preliminary guidance on what type of classification algorithm would be most effective. However, a direct comparison between the two methods cannot be explicitly carried out as they have fundamental differences; specifically, K-Means is a parametric algorithm while Random Forest is non-parametric.

We applied both algorithms to colourized airborne LiDAR data and validated the results by means of a manually classified dataset of random points within the study area. We used this dataset to test these algorithms and assess their accuracy in classifying points as belonging to a given surficial material type.

We assessed the utility of using colour on its own to accurately identify geological features which are relevant to the assessment of rockfall susceptibility in 3D point clouds of large areas. True-colour orthoimagery is usually collected simultaneously with airborne LiDAR. However, such information is not widely used in an analytical way; mostly, imagery is used for visual interpretation. Leveraging the colour information could be useful when aiming to perform screening level assessment of topographic data for indications of geohazard activity or susceptibility. Moreover, full-feature LiDAR is often not available and is costly to collect; the data requirements for this approach are bare earth LiDAR data and orthoimagery within the same extents. The ability to understand not only the 3D geometry, but also the surficial material type through colour will allow for further refinement of data thereby increasing the efficiency of further hazard assessment at finer scales using terrestrial remote sensing or other means.

In forestry sector studies based on remote sensing, the presence of vegetation is often detected by finding the distance between bare earth and non-bare earth returns. This is mostly used as means of quantifying tree canopy but has been used to identify
areas of exposed ground (e.g. White et al, 2010). For this study, we tested the use of spectral reflectance in the visible range (colour) in 3D point clouds as a means of generating information regarding surficial cover. This is carried out using a bare earth LiDAR point cloud which is colorized based on orthoimagery. This work was done to determine if this alternative method offers benefit compared to the differencing of bare earth and non-bare earth returns. The use of colour in this manner, if proven to be a reasonable approach, would be very useful when working with photogrammetry-based structure from motion models, as they do not offer the benefit of multiple returns to be used as a basis for classification. Moreover, this method would also be practical in the event that only bare earth (as opposed to full feature) LiDAR data is available.

### 3.2.1 K-Means Classification

The K-Means algorithm technique is ubiquitous in image classification as it is computationally efficient for use with large datasets and is conceptually simple for non-statisticians to understand (Dhanachandra and Chanu, 2015). K-Means leverages the assumed internal central tendencies of k (user defined) groups in a blended dataset. These tendencies often manifest themselves in the form of statistical attributes allowing for the identification of features possessing these attributes by means of partitioning data in this space. The K-Means algorithm does exactly this; it iteratively places k user-defined centroid points until the average distance from each data point and its cluster centre is minimized. One drawback of the conventional K-Means algorithm is that it requires enough knowledge of a dataset to justify the number (k) of clusters to be used. Nayak et al (2014) found that the K-Means algorithm is effective for segmentation of aerial imagery provided that there are inherent relationships between feature types and
their spectral signature, e.g. between light coloured rock and dark coloured vegetation etc.

### 3.2.2 Random Forest Classification

Many studies have demonstrated the utility of Random Forest classification in the assessment of landslide susceptibility using predictor variables derived from remote sensing data (e.g. Galiano et al, 2012; Guan et al, 2013; Micheletti et al, 2014; Yousef et al, 2016). For rockfall, this technique has not been widely explored. Fanos et al (2018) demonstrated how Random Forest classification can be used to quantify rockfall susceptibility and to determine likely failure modes. However, their work involves extensive use of field mapping and historical event data whereas in this study, our goal is to derive all inputs to the process from a single airborne LiDAR and orthoimagery dataset. This is a supervised classification technique which is specifically designed to limit bias in data. The algorithm inherently utilizes a two-tiered system of randomness in order to identify the logical path to the best classification. The first tier involves picking a random sample of data for iteration of the process, the second involves randomly configuring the decision tree for each iteration. The Random Forest is made up of randomly compiled decision trees, each of which being compiled of random branches (decisions).

### 3.2.3 L*a*b Colour Space

The L*a*b colour space was designed by Hunter (1942) to resolve the subtle differences perceived by the human eye between colours, and is therefore ideal for computer vision and machine learning studies (Bora et al, 2015). Figure 3-1 is a simple conceptual
diagram which illustrates the L*a*b space. “L” defines the luminescence or intensity of a colour while and “a” and “b”, define the colour’s location on the red-green axis and blue-yellow axis respectively (Rathore et al, 2012). Because the colour component is defined through only the “a” and “b” axes, 2D analysis measuring the Euclidean distance between two or more points in “a*b” space is suitable for colour comparison and segmentation of data. Bora et al (2015) present a methodology which utilizes K-Means clustering to the a and b values for the purposes of image segmentation. We use the L*a*b colour space in this paper.

Figure 3-1: Depiction of L*A*B colour space, adapted from Bora et al (2015)

3.3 Study Area

The study area for this site is Mile 28 to 31 of the CN Yale subdivision. The corridor is oriented predominantly north/south, roughly parallel to the east bank of the Fraser river. The southern-most point in the study area is roughly 13 km north of Hope, British Columbia. The study area is shown in Figure 3-2. As can be seen, the study area comprises mostly forested, natural slopes but also contains many sporadic steep rocky cliffs and talus deposits. Clearings within the vegetation can be attributed to active rockfall processes preventing regrowth of vegetation in recent history, or occurrence of
large-magnitude events such as rock avalanche which have deposited large blocks which have prevented the slope from being vegetated since the time of the event (Figure 3-3). Such clearings could be attributed to rockfall activity or coarse rockslide deposits which prohibit the growth of vegetation. In addition, the location of streams and avalanche paths would also generate clearings within vegetated areas.

There is a history of rockfall events within the study area, which have been recorded as incidents in an event database, maintained by CN railway. Since 1996 there have been 55 reported rockfalls, of which 30 have reached the track; this gives an average rockfall frequency of approximately 2.5 rockfalls per year, with 54% reaching the track. Detailed block size information is not available for all reported incidents, however, according to the database, 15 incidents were reported as being <1 m in diameter, while 15 were reported as being > 1 m.
Figure 3-2: Plan map of study area and extents of airborne LiDAR and orthoimagery data used for this study (highlighted in pink). Background image is ESRI base map satellite data.
3.4 Methods and Data

All data used for this study were taken from a single airborne LiDAR and orthoimagery survey in October of 2015, contracted by CN Railway and carried out by McElhanney Consulting Inc. LiDAR data were collected using a helicopter-mounted Leica ALS 70 aerial LiDAR system. The data were collected at a target resolution of 20 points per m\(^2\) with a horizontal accuracy of <10 cm and vertical accuracy of 15 cm in open areas and 50 cm in heavy vegetation cover, to a 95% confidence level. Orthoimagery was collected at a resolution 20 cm per pixel. The extents of the survey relevant to this study are highlighted in transparent pink in Figure 3-2. The methods employed here require a
colourized 3D point cloud. To create this, we combine the orthoimagery with the LiDAR using ArcGIS; each point within the LiDAR point cloud is given the RGB value of the pixel which it overlaps with in the orthoimagery. The discrepancy in resolution between the LiDAR and the orthoimagery may cause instances of misclassification, however, qualitative assessment of the data showed that this was mostly effective. The result is a point cloud where each point contains the 3D location in space as well as a colour represented as an RGB value.

The methodology can be broken up into five phases (Figure 3-4). The process consists of:

1) *Creation of manually classified sample dataset*—as noted in the previous section, we carry out the majority of our analysis on a sample of 45696 points extracted from the colourized point cloud for the study area.

2) *Testing of automatic classification methods using manually classified data*—we test the K-Means and Random Forest algorithms on the manually classified dataset in ‘a*b’ space in order to determine if there are distinct differences in the colour signatures of different surficial features.

3) *Measurement of classification success*—we carry out an in-depth analysis of the automatic classification techniques in order to determine if colour is a useful metric to use for classifying surficial materials in colourized point clouds. This is done by comparing the manual classification to the automated results for both techniques.
4) **Application of automatic classification to dataset for entire study area**—We then apply the classification algorithms to the entire study area dataset and manually inspect the results in order to determine if the strengths and weaknesses of each algorithm identified through testing on a sample dataset can be identified when applied to a larger dataset.

![Diagram](image)

**Figure 3-4:** a diagram highlighting the main components of methodology employed in this study. The inputs are derived from the airborne LiDAR and orthoimagery survey mentioned above.

### 3.4.1 Creation of manually classified sample dataset

The LiDAR point cloud for our study site contains roughly 7 million points, but the majority of the analysis here is carried out on a sample dataset from the above-mentioned survey containing 45696 points, each point being manually classified as
either exposed rock, vegetation or shadow. First, we randomly selected points from the entire dataset, and manually classified each. We continued until the manually classified sample contained approximately 15,000 tree points, and subsequently only added rock or shadow points until each class were equally represented. For our study this is preferable despite the unbalanced proportion of each in the population (i.e. there are far more trees than rock or shadow), as the automated methods we test respond to the ‘signature’ of each group, and are therefore not sensitive to the relative prevalence in each group. The sample dataset is used in order to test the accuracy and success of the K-Means and Random Forest classification techniques. In the case of Random Forest classification, this sample dataset is used to build a classification model to be used in classifying the dataset for the entire study area.

3.4.2 Automated classification of sample dataset

The first algorithm tested here is K-Means clustering using 3 clusters. The method deployed in this study mostly conforms to the workflow proposed by Bora et al (2015). Three clusters were used in running the K-Means algorithm (rock, vegetation and shadow) and iterating fit 1000 times. The iteration which minimized the distance between each point and its cluster mean was used to cluster the dataset. The input to this process is an ASCII text file containing one row for the coordinates of each point in the cloud, with appended scalar fields containing the colour code in L*a*b. A scalar field containing the manual classification code is also present. The output of the process is the same point cloud with an additional scalar field containing an integer value ranging from 1-3 which defines the cluster number of each point based on the K-Means classification.
The clustering process was carried out on the manually classified sample dataset and the resulting classification scalar was compared to the manual classification scalar. If both the manual classification scalar and the K-Means cluster scalar contain the same classification, that point was treated as a true (correct) classification. However, a discrepancy between classifiers necessarily means that one point is accurately classified. Knowing the location of such cases was useful in that it provided a systematic mechanism for identifying and characterizing error.

The Random Forest algorithm is an iterative, decision-tree based approach to classification. A set number (user-defined) of decision trees are generated using random samples of a dataset in order to determine a logical path to accurate classification. The Random Forest approach, originally referred to as tree bagging, was first presented by Breiman (1996) where the concept of repeatedly sampling predictor variables into random subsets of data in order to build decision trees was first used. It was further developed in Breiman (2001) where partitioning of data occurs en masse with the prediction for a given data point being based on the most common result to come out of the trees for that point; all trees, cumulatively, make up the “Random Forest”. The selection of data points for each iteration is random.

The Random Forest classification is carried out using the Tree Bagging algorithm found in MATLAB. The manual classification scalar field is used as the predictor value in order to train the Random Forest algorithm on the manually classified dataset. The inputs for this step were identical to those described for the K-Means classification process. The manually classified dataset was used to train the Random Forest algorithm using incrementally increasing proportions of the dataset. For each proportion,
the model is applied to the entire manually classified dataset (N=45696) and the results are recorded in the same manner as the output of the K-Means classification; the Random Forest classification scalar is compared to the manual classification to quantify true classification and false positive rates.

Upon testing analysis methodology, cursory inspection revealed a large overlap in the L*a*b colour values for shadow and trees, therefore the decision was made to still classify on three classes/clusters but validate the results as a binary classification of rock/not rock. With the goal of identifying rockfall-prone areas in mind, this distinction is sufficient. Therefore, for all proceeding analyses a value of 1 represents exposed rock while a value of 0 represents everything else.

3.4.3 Measurement of classification success

We ran the K-Means algorithm using different proportions of the sample dataset and subsequently measure the movement of each cluster centre for each proportion. This will provide some insight as to how sensitive the process is to sample size and allow for a semi-quantitative assessment of the likely accuracy when applied to the entire dataset. Being an unsupervised technique, it does not offer the benefit of building a model with a sample dataset which can then be applied to a larger dataset with some understanding of the relative success rate. To address this, we generate random samples of the manually classified dataset and measure the difference in cluster centres when K-Means is applied to that proportion of the sample. This will provide some insight into how sensitive the process is to different data. With each run of the K-Means algorithm we also measured the accuracy by comparing the manually classified scalar field to the scalar field generated by the K-Means process.
The goal here was to measure the success of the K-Means algorithm in highlighting the central internal tendencies (discussed in Section 3.1.1) within the data. To do this we measured the direct accuracy by comparing the automated results to those generated by the K-Means algorithm. We also tested the sensitivity of the procedure to sample size by assessing the classification accuracy at different sample sizes. We also measured the distance between cluster centroids at each sample size.

A similar exercise was carried out for the Random Forest classification. Here, we were looking to test the accuracy of the algorithm in the same way as K-Means by directly comparing the results of the classification to the manually classified data. The secondary goal was to determine how the sample size used to create the model effected the results. We did this by building the model at each step using incrementally increasing proportions of the manually classified dataset and subsequently measure the success.

For each iteration of both K-Means Clustering and Random Forest Classification we made note of the number of correctly and incorrectly classified points. This is used to draw conclusions on the validity of our sample set as well as the relative success of each classification technique, which is the ultimate goal of this work. Appendix A contains more in-depth results of the statistical testing carried out here.

3.4.4 Application of automatic classification to dataset for entire study area

The results of this testing on the manually classified dataset are then applied to the entire study area. We run the K-Means clustering algorithm and measure the
distance between these cluster centres and those calculated for the manually classified sample dataset. For the Random Forest classification, the training dataset was applied to the point cloud for the entire study area. In both cases, the output is an ASCII text file containing the colour information, 3D location in space as well as one scalar field containing the classification results for each classification technique. Another scalar field is added which contains the sum of the two classification scalars. This allows us to highlight and better understand the sources of classification error in a qualitative way. In this scalar, a value of zero (K-Means = 0 and Random Forest = 0) suggests a high confidence that a point classified as not rock is correct. A sum of two (K-Means = 1 and Random Forest = 1) means that both algorithms classified a given point as rock and that there is high confidence in this being correct. A sum of one means that the two approaches returned conflicting classifications. This allows you to visually inspect where discrepancies exist, and determine which algorithm is responsible for them. The results are investigated qualitatively to comment on the viability of the process for the purposes of geomorphic classification of 3D colour data.

To compare this technique to the conventional method of calculating the difference between bare earth and non-bare earth returns, we carried out a GIS analysis to determine the vertical difference between the bare earth and non-bare earth returns in an effort to identify areas of bare rock. These results were examined in a qualitative manner in comparison to the results of the classification using colour.

3.5 Results

3.5.1 Sample dataset
A scatter plot of the validation dataset showing the classification of each point is presented in Figure 3-5 while Figure 3-6 shows the validation dataset with each point coloured using its true colour value from the orthoimagery. There is clear separation between the rock section from both vegetation and shadow. However, the boundary between shadow and vegetation is less clear. This is likely due to the presence of dark green coloured vegetation as well as shadow being cast against vegetation, creating a green tone on the shadow. As noted above, the decision was made that rather than attempting to discern between shadow and vegetation, the analysis should be carried out by discerning between exposed rock and all other surficial materials.

Table 3-1 is a contingency table containing the results of the K-Means classification run on the manually classified dataset; each table contains the results for a different proportion of the manually classified dataset in order to test sensitivity. Here the 'n' value represents the number of points used in the clustering. The results are consistent independent of sample size. Of interest here is the rate of true classification for rocks, as well as the false positive rate for vegetation. The conditional probability associated with these metrics are shown in the final two columns. Table 3-2 contains the same information for the Random Forest classification on the manually classified dataset. Here the 'n' value represents the number of points used to generate the classification model, however, the model was applied to the entire manually classified sample dataset each time. The intention of this exercise was to demonstrate that our sample dataset can be used as a test case for the entire study area. There is little variability in the overall results of the K-Means clustering with changes to the sample size. With the Random Forest Classification, the results varied moderately when a
sample size of 229 to 4570 points was used to build the model; sample populations above this range achieved success rates of 98-100% for classifying rocks.

### 3.5.2 Application to Study Area Dataset

Figure 3-7A shows the true-colour 3D point cloud (as shown in Figure 3-2) while B and C show the results of the K-Means and Random Forest classification respectively. In all three figures, 3 boxes are highlighted; these are used to examine the results of both classification techniques on different topographic features. Figure 3-8 contains close up views of the boxes highlighted in Figure 3-7. Both classification techniques capture the near vertical rock slope at Location 1. The exposed talus deposit at location 2 is clearly delineated in both cases. It appears that the shadow points along the edges of the exposed talus deposit are classified as rock in the K-Means results and not rock in the Random Forest Classification results; the former is likely more appropriate, though this is difficult to test. The discolored vegetation at Location 3 are clearly classified more accurately using the Random Forest.
Figure 3-5: scatter plot of the manually classified validation dataset showing their 2D location along the A and B colour scales along the X and Y axes respectively.
Figure 3-6: Results of K-Means classification carried out on manually classified dataset. Point colours represent the true colour of the point extracted from the orthoimagery. The coloured areas represent the boundaries delineated through the K-Means classification.
Table 3-1: Contingency table from testing K-Means clustering with different proportions of manually classified dataset

<table>
<thead>
<tr>
<th>Sample Size (n)</th>
<th>Class</th>
<th>Samples per Class</th>
<th>Correct Classification</th>
<th>Incorrect Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>4570 (10%)</td>
<td>Rock</td>
<td>1442</td>
<td>1384 (96%)</td>
<td>58 (4%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>3128</td>
<td>3025 (97%)</td>
<td>103 (3%)</td>
</tr>
<tr>
<td>11424 (25%)</td>
<td>Rock</td>
<td>3509</td>
<td>3375 (96%)</td>
<td>134 (4%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>7915</td>
<td>7663 (97%)</td>
<td>252 (3%)</td>
</tr>
<tr>
<td>22848 (50%)</td>
<td>Rock</td>
<td>10443</td>
<td>10021 (96%)</td>
<td>422 (4%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>23829</td>
<td>23068 (97%)</td>
<td>761 (3%)</td>
</tr>
<tr>
<td>34272 (75%)</td>
<td>Rock</td>
<td>7004</td>
<td>6728 (96%)</td>
<td>276 (4%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>15844</td>
<td>15316 (97%)</td>
<td>528 (3%)</td>
</tr>
<tr>
<td>45696 (100%)</td>
<td>Rock</td>
<td>13919</td>
<td>13368 (96%)</td>
<td>551 (4%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>31777</td>
<td>30726 (97%)</td>
<td>1051 (3%)</td>
</tr>
</tbody>
</table>

Figure 3-10 contains a scatter plot of all 7 million points in the study area. The orange and red points represent cases which have competing classifications between K-Means and Random Forest (highlighted using the scalar field which sums the two classification fields). The orange points are cases where K-Means classified a rock and Random Forest classified not rock whereas red points show points where Random Forest classified rock and KMeans classified not rock. The inset table provides the contingency percentages for this test.
Table 3-2: Contingency table from Random Forest classification using different proportions of the manually classified dataset to train the model. Each model is applied to all 44690 points. Here, 'n' refers to the number of points used to train the model.

<table>
<thead>
<tr>
<th>Sample Size (n)</th>
<th>Class</th>
<th>Samples per Class</th>
<th>Correct Classification</th>
<th>Incorrect Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>229 (0.5%)</td>
<td>Rock</td>
<td>14624</td>
<td>13526 (92%)</td>
<td>1098 (8%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>31072</td>
<td>30014 (97%)</td>
<td>1058 (3%)</td>
</tr>
<tr>
<td>457 (1%)</td>
<td>Rock</td>
<td>14624</td>
<td>14168 (97%)</td>
<td>456 (3%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>31072</td>
<td>29966 (96%)</td>
<td>1106 (4%)</td>
</tr>
<tr>
<td>4570 (10%)</td>
<td>Rock</td>
<td>14624</td>
<td>14319 (98%)</td>
<td>305 (2%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>31072</td>
<td>30862 (99%)</td>
<td>210 (1%)</td>
</tr>
<tr>
<td>11424 (25%)</td>
<td>Rock</td>
<td>14624</td>
<td>14463 (99%)</td>
<td>161 (1%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>31072</td>
<td>30848 (99%)</td>
<td>224 (1%)</td>
</tr>
<tr>
<td>22848 (50%)</td>
<td>Rock</td>
<td>14624</td>
<td>14474 (99%)</td>
<td>150 (1%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>31072</td>
<td>30947 (100%)</td>
<td>125 (0%)</td>
</tr>
<tr>
<td>34272 (75%)</td>
<td>Rock</td>
<td>14624</td>
<td>14509 (99%)</td>
<td>115 (1%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>31072</td>
<td>30951 (100%)</td>
<td>121 (0%)</td>
</tr>
<tr>
<td>45696 (100%)</td>
<td>Rock</td>
<td>14624</td>
<td>14517 (99%)</td>
<td>107 (1%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>31072</td>
<td>30975 (100%)</td>
<td>97 (0%)</td>
</tr>
</tbody>
</table>
3.6 Discussion

These results suggest that classification of point cloud data using colour offers value to conventional hazard identification techniques. The percentage of misclassified vegetation is low, therefore there is confidence that points classified as rock are indeed rock. One critical consideration in this, however, is shadow; because it is so prominent it’s important to acknowledge that shadow points may actually be rock. Therefore, if this information is to be used in assigning rockfall probability for a given location, the presence of rock identified through this process should be used to increase probability based on the presence of a given feature, however, its absence should not rule that location out as a potential source or deposit. This is because there is a high level of confidence that points classified as “rock” are rocks, but points classified as “not rock” can be shadow, which may be obscuring rock. This means that the absence of discernable rock at a given point should not adversely affect any type of probabilistic rating. Put simply; the uncertainty associated with the “not rock” classification should be taken into account when using these results as inputs into probabilistic hazard assessment.

More work is necessary in order to deploy this concept broadly with confidence. More attention could be paid to building an unbiased test dataset. A random sample dataset was initially created. However, it was found that a useful number of points could not be practically classified manually in a truly random way, because the site contains so much vegetation and the point cloud for the study area contains 7 million points; a random sample of even 5% would be 350 000 points, which would be an extremely time-consuming endeavor to manually classify. Instead, an even number of points for
each class were created and classified. Effort is required to understand the inherent bias created in this approach. The fact that results stayed relatively consistent independent of the number of points used suggests that the impact of this bias is minimal.

Figure 3-7: A-RGB image of one section of track within the study area. B-results of unsupervised K-Means classification. C-results of Random Forest classification.
Both processes yielded favorable results which were relatively consistent independent of the sample size used in the process; Random Forest returned results which are marginally more accurate than the K-Means classification. This is due to the fact that K-Means clustering is a linear process and, as demonstrated in Figures 3-5 and 6, the boundaries in 'a*b' space between features is not. The Random Forest algorithm is nonlinear and is therefore able to identify individual trends within the data and segment accordingly.

Figure 3-11 contains a profile from the study area. The profile contains points coloured by their classification. This demonstrates that there are noticeable clusters of points classified as rock with different slope angles. Using the methodology from Loye et al (2009), we can identify these clusters in order to identify slope angle thresholds which represent rockfall sources and deposits. This figure is a conceptual demonstration of how the analytical techniques explored here can be applied to a conventional hazard identification methodology, e.g. slope angle can be added to the process in order to not only understand the properties of the surficial materials, but also the characteristics of the topography as they relate to rockfall hazard. The added benefit here is that upon identification of potential rockfall sources and talus deposits using slope angle, the colour analysis can be used to focus in on areas of exposed rockfall deposits and outcrops. In the following chapter this will be carried out; we will integrate the output of the analysis proposed here into a hazard identification exercise for Miles 20 to 30 in the Yale subdivision.
Figure 3-8: Zoomed in images of the boxes highlighted in Figure 8; for each box, an RGB image as shown in addition to the results of both the K-Means and Random Forest classification analyses.
Figure 3-9: This plot shows the point cloud for the entire study area in A*B space. In red, are points where Random Forest classified a rock and K-Means classified not rock. In orange are points where K-Means classified a rock and Random Forest classified not rock. The orange points in the top right corner are mostly attributed to lighter coloured vegetation. The cause for the error elsewhere is harder to pinpoint, but the linear nature of these patterns suggest that there is some overfitting happening in the Random Forest classification.
Figure 3-10: a-Colourized LiDAR point cloud of study site, with line denoting plan-view of section A-A’ (shown in Figure 10b. b-Section A-A’ consisting of terrain classification and non bare earth LiDAR returns. Points 1 and 2 are highlighted in both A and B and demonstrate how colour classification paired
with information regarding slope angle can be used to interpret geomorphic setting for specific sites. The inset in the top left corner of Figure 9B zooms in on area 1 to further highlight the impact of slope angle on superimposing 2D colour into 3D space. The horizontal line at the top of 3-9b is the 2D representation of the colour classification.

The testing dataset was sufficient for our preliminary testing, however, the Random Forest model showed definite signs of overfitting; the sporadic and non-linear classification in Fig 3-9 is indicative of overfitting, where the model fails to generalize, instead simply matching the variance in training data. It should be noted, however, that the impact of overfitting, though visible in the Random Forest model, had a minimal impact on the final output; Figure 3-9 shows that the two algorithms only disagreed on 8% of points. Of these points, some are likely due to the differing boundaries between classification in the two models, these are represented in the cluster of points in the lower right of Figure 3-9. The points most likely effected by overfitting are the ones which fall along the line extending vertically, starting at roughly 5,12 in ‘a*b’ space.

The use of 2D orthoimagery in 3D space is inherently problematic; in cases where geometry is near vertical, colour may not be truly representative of topography, for example, a cliff could be assigned the colour of a tree. Due to this limitation, this process should be attempted using helicopter-based structure from motion (SfM) photogrammetry. Using this method of data collection, colour would inherently be collected in 3D space, thus eliminating the limitations discussed associated with extracting colour from 2D imagery into 3D space. However, as SfM necessarily requires images captured at varying incidence angles, limiting shadow can pose a challenge, but, given the uncertainty in classification of points within areas of shadow there is
interest in taking precaution in order to reduce the prominence of shadow in data collected for these purposes.

From the onset of this study, the analysis was carried out ignoring the luminosity value of a colour in L*a*b space. A further study could benefit from understanding what can be gained by adding this third dimension to the analysis. It is possible that some of the ambiguity between classification boundaries could become more clear by adding this element to the analysis.

These results suggest that the process could be iteratively carried out to further classify data into sub categories such as lithology or a measure of surficial “freshness” which could be used to indicate recently active talus slopes or rockfall scarps. That is; the output of this analysis, specifically the rock points, could be used as an input to another iteration of the analysis which aims to pick up on important differences in the colour value of different rock types, lithologies and/or conditions in order to offer more insight into the geological setting of the area.

The work done here shows promise as a standalone tool in broad land cover classification of topographic data. However, it may also have a place within a multi-criteria evaluation aimed at quantifying rockfall susceptibility. Colour value could be used alongside geometric parameters such as surficial roughness, curvature, slope angle or aspect as well as many other parameters to more thoroughly classify terrain.

With access to extremely high-resolution data, it is plausible that the analysis of colour of just exposed rock could be used iteratively to segment data further. The workflow presented here can be used to isolate areas of exposed rock. A dataset containing just these points could be put through this process again in order to identify
subtle differences in the colour value of exposed rock in order to identify different lithologies, assess weathering and general surficial conditions of identified outcrops.

Figure 3-11a – c show an area of slope in plan view in order to demonstrate how an analysis using the vertical distance between bare earth and non-bare earth returns (Figure 3-11b) compares to the analysis carried out here (Figure 3-11c). Using the vertical difference allows for the isolation of areas lower vegetation cover, however, a decision needs to be made as to what distance represents bare ground; thin vegetation and surface roughness can generate non-bare earth returns at locations where the surface is essentially bare. Figure 3-12 provides a cross section to exemplify cases where the bare earth model is missing ground return points and the non-bare earth point cloud includes points which appear to be part of the bare earth. Moreover, areas of high gradient and high surficial roughness can result in negative difference values between bare earth and non-bare earth returns; this complicates the decision about the above noted thresholds. The colour analysis, on the other hand, is a binary classification which inherently eliminates the need for the above-mentioned decision. Moreover, this classification technique is useful in the event that non-bare earth data is unavailable, which is common in practice.
Figure 3-11: a) orthoimagery of an area of slope in the Yale subdivision. b) slope angle. c) the vertical distance between bare earth and non-bare earth returns. d) the results of the colour analysis carried out here. e) instances of negative vertical differences resulting from surficial roughness, mostly occurring on areas
of high gradient. The purple areas shown in each image are to provide reference for comparison.

3.7 Conclusions

The a and b values of the L*a*b colour space were used to carry out tests regarding the classification of colourized point cloud data by means of colour. These values proved useful in highlighting the central internal differences in colour value between features. We tested two semi-automated approaches for classifying colourized 3D point cloud data; K-Means clustering algorithm and Random Forest classification. They achieved success rates of 97 and 94% respectively. Table 3-1 shows that there was little variability in the results of the K-Means clustering. Table 3-2 shows the same for Random Forest. Classification when the model was built using 25% of the sample or greater. This suggests that, although our sample dataset has inherent bias, the findings of our classification tests on a manually classified dataset can be extrapolated to a larger dataset with confidence. Manual inspection of a dataset of 7 million points classified using both algorithms furthered this argument. We found that, though classification accuracy was high, there are inherent problems with extracting colour from 2D space into 3D which could be addressed by selecting a true-3D dataset at the onset of the project.

We showed that there are distinct differences between the colour values of exposed rock and other surficial features within a*b space. We presented two potential approaches to identifying and leveraging these differences in the interest of image classification. Although the boundary between shadow and vegetation in 'a*b' space is not well defined, the boundary between these features and exposed rock is; thus
allowing for our analysis to look simply at the binary classification of exposed rock and all other features.

Figure 3-12: Cross section demonstrating cases where bare earth model is inconsistent and may create confusion when differencing the bare earth and non-bare earth returns

The Random Forest classification is the more accurate of the two techniques explored here; though the 97% accuracy rate of Random Forest is only marginally
higher than that of K-Means at 94%. However, Random Forest would be the preferred method if only one technique is employed. This is due, firstly, to the marginally higher success rate, but also the fact that error is much more contained. The K-Means errors are mostly due to lighter coloured vegetation which falls towards the top right of the plot in Figure 3-9, though there are some errors closer to the middle of the x axis where there is a cluster of red points which fall in the darker areas of a*b space (See Figure 3-6) which makes it difficult to discern between feature types in a visual manner. The approach of combining the results of both techniques provides otherwise unachievable insight into classification error for exploratory studies such as this one.

The use of colour as a means of identifying outcrops and rockfall deposits in colourized point cloud data has been assessed. We have demonstrated the ability to highlight areas of exposed rock through this analysis but have not developed means of making the distinction between types of rock. Surficial parameters such as slope angle can be used to further refine these results in this way.
3.8 References


4.1 Introduction

Rockfall hazards can cause disruption to the consistent and safe operation of transportation corridors. In western Canada and elsewhere, managing risk from slope hazards within these corridors is particularly challenging as they necessarily traverse rugged and steep, mountainous terrain. Therefore, operators have interest in recognizing and quantifying slope hazards to reduce the risk they create. For rockfall hazards, several hazard or risk rating systems have been developed (e.g. Pierson, 1990; Abbot et al, 1998; Budetta, 2004; Pritchard et al, 2005; Franklin et al, 2012) and have been reviewed by Ferrarri et al (2016). These are mostly tailored to the management of rockfall hazards directly adjacent to the right of way, which means that hazard sources not visible from the base of the slope, where the right of way is normally located, may be overlooked. The use of high resolution point cloud data has proven to be an effective platform for identifying rockfall hazards, regardless of whether or not they are visible from track level and assessing their failure mechanisms and potential volumes (e.g. Kromer et al, 2017; van Veen et al, 2016; Lato et al, 2015; Abellan et al, 2016; Eitel et al, 2016; Jaboyedoff, 2010). Topographic data of high enough resolution to acquire such information has mostly been demonstrated through terrestrial-based platforms, which cannot practically be deployed over large areas. Therefore, novel tools for the identification and characterization of rockfall hazards using regional scale datasets are of value in that they would allow for the deployment of high resolution data.
collection in an efficient manner. In this study, a GIS-based methodology was proposed for the identification and prioritization of rockfall prone areas. Semi-automated techniques were used to ensure that the methods developed here are deployable at a large scale. We considered the geometry of the slope, as well as the protective capacity of forest cover to quantify the level of hazard present for a given slope. To do so rockfall sources were identified, extents of runout zones were delineated, and forest characteristics downslope of a given source and within its runout zone were quantified. All data used for this study were collected in a single Airborne LiDAR and orthoimagery Survey in October of 2015, contracted by CN Railway and carried out by McElhanney Consulting Inc. LiDAR data were collected using a helicopter-mounted Leica ALS 70 aerial LiDAR system. The data were collected at 20 points per m² with a horizontal accuracy of <10 cm and vertical accuracy of 15 cm in open areas and 50 cm in heavy vegetation cover, to a 95% confidence level. Orthoimagery was collected at a resolution of 20 cm per pixel. The survey was carried out for the CN Rail corridor on the Yale and Ashcroft subdivisions.

Identifying rockfall sources at a regional scale is a complex task. Furthermore, Lato et al (2014) note that, due to occlusions resulting from the angle of incidence between a sensor and a target surface (common to any remote sensing technique), features dipping more than 80° are likely to be occluded from data collected from downward-facing vantage points (e.g. airborne LiDAR). This complicates the task of identifying and properly characterizing potential rockfall sources in that rockfalls are most likely to occur on steep or vertical slopes (Hoek and Bray, 1981; Dorren, 2003). Thus, many studies have used slope angle as a means of identifying rockfall prone
areas (e.g. Cloutier et al, 2015; Li & Lan, 2015; Fanos & Pradhan, 2016). We used the Slope Angle Distribution (SAD) method proposed by Loye et al (2009) to classify the terrain geomorphically using slope angle, thus allowing for the identification of rockfall sources and talus deposits. In addition to the geomorphic classification using slope angle, analysis of colour derived from orthoimagery was used to differentiate between vegetated slopes, bare outcrop and accumulations of talus.

Many authors have explored the protective capacity of forests and developed methods for quantifying it (e.g. Brauner et al, 2005; Berger and Dorren, 2007; Bourrier et al, 2012; Fuhr et al, 2015; Dupire et al, 2016). For a given rockfall source the downslope gradient, length of forested slope and the size and distribution of trees are the most crucial factors controlling the passage of rocks through forested terrain. We followed the approach of Berger and Dorren (2007) and considered these factors at the slope scale to estimate the likelihood that a rockfall produced by a given source area will be arrested during passage through the forest, and thereby be prevented from reaching elements at risk (e.g. railway, highway) at the base of the slope.

Rockfall simulation has proven to be an effective means assessing the path and velocity of a rockfall given the size and shape of a detached block (e.g. Dorren, 2003; Lan et al, 2010; Sala, 2018). However, because of the regional nature of this study and the noted limitations of downward facing aerial LiDAR, the ability to accurately assess block volumes and shapes is not a practical goal. That is, the necessary information to assess shape and volume is not achievable at a regional scale because the limitations of airborne LiDAR make it impossible to assess this information at any scale in steep terrain using ALS. The inability to capture vertical slopes through the vantage point of
traditional ALS (e.g. mostly downward-looking) makes identifying and measuring discontinuities along vertical faces impossible, thereby not allowing for accurate measurement of the shape and dimensions of rock masses likely to fail. Therefore, a process-based model (Flow-R), developed by Horton et al (2013) was used here. In this model, rockfall propagation is assessed by means of a raster cell-based analysis of probabilistic spreading as well as a simple calculation of energy which does not require information regarding the volume or shape of a detached block. The terrain and spreading parameters alone dictate the lateral extents and total runout distance of rockfall events. Though initially designed to model debris flow propagation, Michaud et al (2011) demonstrated that the model can be adequately parameterized to characterize rockfall runout. They found that the use of Flow-R and the Slope Angle Distribution method along with high-resolution topographic data (as well as detailed geological information, if available) offers a low-cost means of generating valid regional-scale rockfall hazard assessment to identify areas where detailed investigation and hazard management strategies might be appropriate. These findings are in line with the objectives of our study; leveraging regional scale datasets to focus further investigation in areas which exhibit the characteristics of slopes with high probability for rockfall that will impede railway operation in any way.

A framework for the study of rockfall hazards through collection and analysis of high resolution 3D data, which is largely based on work by Gauthier et al (2017), is presented in Figure 4-1. Our work fits into the searching and prioritization and, to a lesser extent, the characterization section of the framework. We evaluated slope conditions to identify where rockfalls are possible through the identification of rockfall
sources. Of these, we determined where rockfalls are most likely by using the runout zone of each source to assess the characteristics of the downslope terrain and forest cover. The terrain provides indication of rockfall activity (or a lack thereof) and forest cover will help to determine the likelihood that a rockfall from a given location will interface with the track. Together, these parameters can be used to prioritize areas of slope in terms of where time, effort and financial resources will be most efficiently used in the further assessment of rock slope hazards and/or the implementation of mitigation strategies.

We have developed a novel approach for the identification and characterization of rockfall hazard at a regional scale. This is achieved through the creation of analytical techniques and tools which leverage modern computer processing capabilities and, where possible, make use of the rich 3D information provided by LiDAR data. The use of colour as an additional input (as presented in Chapter 3) to conventional terrain classification using slope morphometrics of LiDAR data allows for an increased level of understanding of the terrain without direct human intervention and adds to the practicality of the approach in that it maintains the goal of being semi-automated and therefore applicable over large areas. The assessment of forestry parameters as a means of further refinement of regional prioritization is a novel approach; many studies have looked at the protective capacity of forest cover for rockfall protection and the subsequent reduction in risk, however, none of them have attempted to quantify it at this scale.
The workflow presented here fits mostly on the first two terms of the risk equation defined by Porter and Morgenstern (2013) and represented in Equation 1:

Figure 4-1: A process developed for the identification, characterization and monitoring of rockfall hazard through the use of remote sensing. The information derived within each step of the process informs the best decision for the next. Broad suggestions for spatial and temporal resolution required to achieve various outcomes are also noted.

The workflow presented here fits mostly on the first two terms of the risk equation defined by Porter and Morgenstern (2013) and represented in Equation 1:
\[ R = P_H \times P_{S,H} \times P_{T:S} \times V \times E \]

where:

\( R \) = risk;

\( P_H \) = annual probability of a rockfall occurring;

\( P_{S:H} \) = spatial probability of a rockfall reaching the element at risk, which, from a railway perspective, is ultimately the train although it is more likely to be impacted by, for example, impassable track. Therefore, when we refer to this variable, we are referring to the probability of a rockfall reaching the track.

\( P_{T:S} \) = temporal probability that a train will be present when a rockfall occurs

\( V \) = the probability of a loss of life in the event of a rockfall reaching a train

\( E \) = the number of people at risk (equal to 1 for individual risk).

We simplified the \( P_H \) term by distributing it evenly through space. We provided insight as to how the information derived in this workflow can be used to inform \( P_H \) in a more rigorous way on the basis of the presence of talus deposits being indicative of rockfall activity, however, this was not explicitly carried out. Figure 4-2 demonstrates how this workflow fits into a complete risk assessment in terms of quantifying overall risk. Further, this diagram shows how any component of the risk chain can be added to this process to reduce uncertainty. For example, triggers are not being considered in this study, however, if information regarding the spatial distribution of triggering probability became available, this information could be included in the analysis of rockfall occurrence probability, and thereby increase confidence in the final probabilistic outputs. This diagram is designed in such a way as to demonstrate the entire risk chain.
(on the left), explain the fundamental question being answered by each parameter (middle) and make note of how, if at all, that question was addressed in this study.

Figure 4-2: The risk equation as defined by Porter and Morgenstern (2013). In the middle of the diagram, for each component of the risk equation, there is a single question related to the evaluation. On the right, an explicit link between the parameters of the risk equation and this study is made. This study works to inform the risk equation up to the point of a rockfall reaching the track, impact and consequence are not considered here.
4.2 Engineering Geology of the Fraser Canyon

The study area is roughly Mile 21 to Mile 25 of the CN Yale subdivision, which runs through the Fraser River canyon between Yale and Boston Bar, British Columbia, shown in Figure 4-3. The Fraser River canyon is oriented North/South within this reach and is largely characterized by steep scarp-like rock slopes, talus deposits and vegetated slopes. The canyon was forged by glaciation as well as erosion by the river (Monger, 1970). It is marked by numerous steep, rocky cliffs, mainly made up of metamorphic rocks. These features have required steep cuts to accommodate the CN alignment, which runs mainly along the east bank of the river. There is evidence of geohazard activity in the area such as landslides, rock avalanches, rockfall and debris flows (Pratt et al, 2018). Steep slopes, oriented near parallel to the Fraser and Yale faults (Piteau, 1977) actively release rocks downslope, resulting in notable accumulations of talus lower-down on these slopes. Failures controlled by local-scale faulting in this same orientation has resulted in bench features in many locations; these are often marked by large talus blocks which were sourced by the steep rock slope directly above.

Rockfall activity in the area is predominantly sourced from these steep, natural slopes, however, Pratt et al (2018) note that the presence of rock cuts has a major influence over rockfall activity along the CN alignment. The presence of vegetation and its prominence on these talus benches can be used in assessing the short term rockfall fall frequency for a given location.
Figure 4-3: Plan map of study area– Mile 20 – 25 of the CN Yale Subdivision, British Columbia, Canada --

The Fraser River represents the boundary of the Coast and Cascade Mountains (Monger, 1970). The geology along this frontier is quite diverse due to a geologic history which includes several orogenic episodes (McTaggart and Thompson, 1967). Most involved folding and faulting and many involved metamorphism and intrusion. The presence of many orogenic events makes the geology difficult to interpret as each event is obscured by the next; though McTaggart and Thompson (1967) note that there are certain discernable fold geometries and configurations which are indicative of specific events. According to Pratt et al (2018) the presence of faulting creates a zone of foliated and highly metamorphosed rock characterized by high levels of shearing between units.

Piteau (1977) identified several post-glacial rock avalanches in the area between Boston Bar and Yale, specifically between Miles 24 and 26 on the CN corridor, directly
adjacent to the town of Yale. He also noted that the entire area is characterized by a pronounced faulting zone which trends roughly parallel to the Fraser River and is marked by an abundance of heavily fractured rock along the two dominant faults in the area—the Fraser and Yale faults. From a geomorphic standpoint, the physiography of the Fraser Canyon points to post-glacial landslides as being the most influential process in the area along with fluvial processes which have ultimately shaped the landscape. This observation is reinforced by the prominence of steep rock cliffs on the upper reaches of many of the slopes which are likely scarps of rockslides (Piteau, 1977; McTaggart and Thompson, 1967).

4.3 Methodology

A basic outline of the methodology employed for this study is shown in Figure 4-4. Broadly, the process can be broken up into five components:

- **Terrain classification**: Regional scale semi-automated process which made use of colour from orthoimagery and slope angle from the ALS to determine slopes which are likely to be releasing rockfalls.

- **Assessment of forest cover**: Regional scale identification of individual trees as well as their dimensions using the Dorren (2017) approach. These were used as inputs to the subsequent steps aiming to quantify the probability of a rock reaching the track given forest cover.

- **Delineation of maximum runout boundary for each potential rockfall source**: A rockfall propagation model (Flow-R; Horton et al, 2013) was used to identify the limit of credible reach of a rockfall from a given location. This process was carried out for each rockfall source zone identified through the terrain.
classification process. We determined where a rockfall from a given source is likely to end up in order to assess terrain downslope of a given source as well as the impact of forest cover on rockfall runout within the same reach.

- **Slope-Scale analysis**: the maximum runout boundaries were used to move from the regional scale to the slope scale; the forest parameters and terrain downslope of each identified source was assessed, and protective capacity offered by forest cover was quantified. Here, for each source, we answered the question of, given forest cover, what is the probability that a rockfall reaches the track.

- **Regional scale prioritization**: a simple probabilistic assessment is employed which considers the probability of rockfall occurrence as well as the reach-reduction potential offered by the forest cover for each slope. Here, we spatially quantify all of the identified hazards in terms of where they intersect the railway corridor and use the density of such occurrence to assign a priority level for any given point on the track.

Each component of this methodology will be described in detail in the following sections.
Figure 4-4: High-level overview of the workflow developed in this study. The analysis starts at the regional scale and uses the individual runout zones to bound analysis at the slope scale for each potential rockfall source.

4.3.1 Terrain Classification

This section outlines the methodologies employed for classifying the terrain into geomorphological units using slope angle and for identifying surficial material using the colour from orthoimagery. Figure 4-5 demonstrates basic principles used to characterize
rockfall hazard. In this example, there is a fresh rockfall scarp as well as a runout zone which almost perfectly adheres to the observations made by Toppe (1987) which suggest that rockfall runout generally stays within +/-20° of the slope direction downslope of the source. The downslope terrain is mostly marked by talus accumulation. Moreover, in this case the vegetation is very sparse and likely has very little effect on the probability of a rockfall reaching the track.

Conventional rockfall susceptibility studies make use of slope angle to identify relevant features such as rockfall sources and talus slopes (e.g. Cloutier et al, 2015; Li & Lan, 2015; Fanos & Pradhan, 2016). Many also use a combination of field work, air photo interpretation, examination of LiDAR data and analysis of event databases in order to identify rockfall sources and rate their level of activity (e.g. Guzzetti et al. 2003; Frattini et al. 2008). Lan et al (2010) note that the source identification process should consider slope angle, surficial materials and vegetation to methodically focus attention onto slopes which are actively releasing rocks.
We used colour in addition to slope angle to hone in on specific features within the data with more confidence. For example, using the methodology proposed by Loye et al (2009), one can identify areas with a slope angle which satisfies the criteria for a given geomorphic classification; however, whether or not the surficial material within the area is indicative of rockfall activity is unknown. In the case of rockfall deposits, this is particularly pertinent; sources located above talus slopes containing fragments of rock
are likely to have a history of relatively higher activity (Toppe, 1987; Dorren and Seijmonsbergen, 2003; Guzzetti et al., 2003; Jaboyedoff and Labiouse, 2003; Frattini et al., 2008) and should be treated as such in any effort to prioritize slopes in terms of their level of rockfall hazard. It is important to recognize that other processes, such as rock avalanche, could also be responsible for creating talus deposits, therefore a means of discerning between deposit types is critical when using surficial material to inform a measure of rockfall probability; this concept is addressed in section 4.5. Simply stated, slope angle alone only answers half of the question at hand—we’d like to know where slope geometry is indicative of rockfall sources and talus and subsequently determine whether or not rockfall processes are actually at play by defining talus accumulations which are sourced by rockfall and appear to be active. The former can be answered by slope angle alone while colour can be used to answer the latter.

Ortho-imagery for the study area is first combined with the bare earth LiDAR data using a simple GIS overlay technique where each point within the bare earth point cloud is given an RGB value from the corresponding pixel of the orthoimage. The next step is to assess the colour value of each point and place it in the appropriate classification bin (rock, vegetation and shadow); this is done using the methodology described in Chapter 3 and broadly described in Figure 4-6.

To classify terrain into morphological units, we used the methodology from Loye (2009) which decomposes the histogram of slope angles of an area into a single Gaussian distribution for each geomorphic unit present. This approach is based on Strahler (1950) who suggests that slope angles within a given geomorphic unit will be distributed around a significant mean value. Loye et al (2009) developed a process for
identifying the distribution of slope angles around these means which can subsequently be used to find slope angle thresholds for each geomorphic unit. In our study area, four dominant geomorphic units were delineated:

1) flats/benches: low-gradient features, benches refer to flat features below steep rock slopes which are likely the result of local scale fault-dominated failures, described in Section 4.2. Flats refer to flat areas at the foot of slopes as well as plateaus.

2) fans: high concentrations of colluvium which have been deposited by means of fluvial processes.

3) talus slopes: moderate gradient, usually immediately below steep rock slopes.

4) potential rockfall sources: steep rock slopes capable of releasing rocks downslope.

Figure 4-7 contains an example of each classification type, as defined for this study, found within the airborne LiDAR dataset.

The Loye (2009) method can produce broad distributions (low-kurtosis) within geomorphic domains which can lead to overlap between distributions increasing the uncertainty in the classification of points within these overlapping ranges. We applied Loye’s method as it was originally intended (as an unsupervised technique). But to determine whether these areas of overlap are of concern, we used the distribution of slope angles within a manually classified dataset of 45000 points (discussed in Chapter 3) to validate the results from Loye’s method.
Figure 4-6: Visual depiction of the process of combining 3D information from ALS and 2D information from orthoimagery in order to create a colourized 3D point cloud which can be used as the input to the unsupervised colour analysis.
Figure 4-7: bare earth LiDAR DEM with annotations which define the features identified through the Slope Angle Distribution method

4.3.2 Calculation of Forest Parameters

Dorren (2017) introduced a software package called FiNT, which uses a bare earth digital elevation model (DEM) as well as a digital surface model (DSM) as inputs and uses a moving kernel to identify individual trees that exceed a given height threshold. Hanus et al (1999) examined the relationship between tree height and stem diameter and developed empirical equations for calculating such; we used this equation to infer stem diameter for each tree point identified by FiNT. This process requires several input parameters which vary based on tree species; we used the constants recommended for Douglas Fir trees as this is the predominant tree species in the study area (British
Columbia Ministry of Forests, 2002). An important thing to note here is that FiNT has not been calibrated for use in North America and calibration data for our study site is not available for testing. We have used FiNT on the basis that our study area is composed of similar terrain and vegetation to those where FiNT has been tested.

This information can then be used as an input for calculating the protective effect offered by forest cover which is expressed as the probability of a falling rock reaching an element at risk, given passage through a protective forest cover. Berger and Dorren (2007) developed a system for quantifying this based on several parameters, which are listed in Table 4-1. The method requires the calculation of a profile line representing the likely path of a falling rock from a source. To generate this, we projected a line from each source to the lowest point within its runout zone (delineated through process described in Section 4.3.3).

**Table 4-1: Parameters required for the calculation of risk reduction offered by forest cover as suggest by Berger and Dorren (2007).**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Process of extraction of parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block Size (m)</td>
<td>Size of falling block</td>
</tr>
<tr>
<td>Block Density</td>
<td>Density of falling block</td>
</tr>
<tr>
<td>Profile Line</td>
<td>Maximum runout length derived from runout boundary generated through process outlined in section 1.3.5</td>
</tr>
<tr>
<td>Slope length (m)</td>
<td>Extracted for profile line using DEM</td>
</tr>
<tr>
<td>Gradient (deg)</td>
<td>Extracted for energy line, using slope angle raster (derivative of DEM created in ArcGIS)</td>
</tr>
<tr>
<td>Forested slope length (m)</td>
<td>Combination energy line and terrain classification results.</td>
</tr>
<tr>
<td>Average tree diameter at breast height (DBH) (cm)</td>
<td>Individual trees are generated through FiNT, average for a given slope is calculated by averaging DBH for all trees within runout boundary of a given source.</td>
</tr>
</tbody>
</table>
Tree density (trees per Ha) Individual trees are identified through FiNT, density for a given slope is calculated using all trees within runout boundary of a given source.

4.3.3 Delineation of Maximum Runout Boundaries

The next step in the process is to delineate a maximum credible runout limit for each source identified through the terrain classification process. Conventionally, runout modelling is used to determine the (typically 2D) path a detached block may take in the event of a rockfall and to assess its energy. Here we determined the maximum likely extents (lateral and along-path runout distance) of a rockfall from a given location to link evidence of rockfall (talus deposits) to likely sources. This output provides the means to carry out analysis at the slope scale for each identified source. We used the sources identified through terrain classification as input to Flow-R. The model was run once for each source. Each source was given a unique integer ID which was used to link the Flow-R output to its source in later stages of this methodology. Michoud et al (2012) and Cassanova et al (2016) have demonstrated its use in the identification of rockfall runout zones.

Regarding rockfall, principles applied in Flow-R are similar to those of Conefall, a software developed by Loye et al (2008) for delineating rockfall runout zones. This software simply generates a cone of a given lateral extent (e.g., +/-20°) and modifies the cone based on the terrain in order to eliminate topographic barriers from the runout path. Flow-R gives more control over relevant parameters to more effectively delineate a runout zone—those parameters guide the propagation of energy and lateral spreading. Cassanova et al (2016) calibrated the Flow-R model for rockfall and found
good agreement using the modified Holmgren algorithm for probabilistic spreading, which was found to be the critical parameter in using Flow-R for rockfall modelling. The original Holmgren (1984) method is a raster-based multiple flow direction algorithm which allows flow to continue in multiple downward directions (as opposed to just one as is the case with the D8 algorithm) while taking gradient into account in order to distribute energy to each direction accordingly. Michoud et al (2009) added a parameter to this algorithm which adds a constant to the elevation of the raster cell being calculated to reduce the impact of minor surficial roughness. Simply put, this modification is designed to prevent a small relative topographic high from terminating propagation altogether. Cassanova et al (2016) successfully used a value of 1 m for this constant for modelling rockfall; we do the same. Table 4-2 provides an explanation of the parameters available within the Flow-R model and their function, including lateral spreading. Table 4-3 provides an overview of the parameters used for this study.

To reiterate the objective here: we are not attempting to identify specific rockfall paths through this modelling process, rather we are using Flow-R to calculate a zone of influence for each source, where, within those boundaries, all features can be attributed to the source(s) which cross that point in space. Using the spatial boundaries of the runout zones, we can determine which rockfall sources reach any given point within the study area. We are assuming here that a deposit of rocky material at a given point was necessarily deposited by a source which has a runout zone that crosses that point; we do not account for other mechanisms, natural (e.g. rock avalanches) or anthropogenic, which may be responsible for such deposits.
Table 4-2: Main functions of Flow-R model

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic</td>
<td>Spreading Each raster cell is rated based on the probability that it spreads to its neighbors. The probability is a function of the slope angle and direction of the cell in question until a null-probability is reached (Michoud et al, 2012).</td>
</tr>
<tr>
<td>Energy</td>
<td>Models the behavior of inertia; each cell is weighted based on it's slope angle and direction as well as the slope angle and direction of its neighboring cells. Energy loss is calculated using a simple friction model (Michoud, 2012)</td>
</tr>
<tr>
<td>Runout Distance</td>
<td>Achieved once an energy balance of 0 is reached</td>
</tr>
</tbody>
</table>

Table 4-3: Parameters used in Flow-R model for this study in order to delineate maximum runout boundaries for each potential source identified through the terrain classification process.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow Accumulation Method</td>
<td>D8</td>
</tr>
<tr>
<td>Spreading Direction Algorithm</td>
<td>Holmgren Modified 1m, exp. 1</td>
</tr>
<tr>
<td>Energy</td>
<td>Weights</td>
</tr>
<tr>
<td>Energy Loss</td>
<td>Travel Angle (30 deg)</td>
</tr>
</tbody>
</table>

4.3.4 Slope-Scale Analysis

The previous components of the workflow facilitated the generation of regional-scale information. In this section we used the spatial boundaries delineated using Flow-R to limit the extents of further analysis. That is, exposed talus deposits and other rockfall sources downslope of a source within those boundaries are identified and their proportion of the total area within the runout zone was calculated. Also, the same boundaries were used to quantify the protective capacity of downslope forest cover using the trees identified using FiNT as inputs to the method proposed in Berger and
Dorren (2007). The total basal area is calculated by summing the total diameter of all downslope trees; the average of this value is also taken and referred to as the average diameter at breast height (DBH). The terrain classification is used to identify the length of forested slope along the profile line generated using the method described in Section 4.3.3.

Figure 4-8 gives context to what is being measured here; on the top half of the figure the bare earth models of two slopes with obvious signs of rockfall activity are shown. In both cases, with only this information, both slopes would likely be assigned a high hazard rating. The addition of forest cover as a channel of information (shown on the bottom of Figure 4-8) provides a great deal of site-specific context. Though both slopes appear to be actively releasing rocks, the forest cover of the slope on the left is much more prominent than the one on the right and will therefore offer more protection from rockfall by limiting the extent and energy capacity of a rock traversing downslope. By using FiNT to identify all trees and their dimensions, and subsequently clipping that output to the extents of a given rockfall runout zone, it is possible to generate all of the parameters required for Berger and Dorren’s (2007) method aside from the dimensions and properties of the falling block. As this approach requires the volume of a falling block as an input, we used a block size of 1 m³. This decision was made for two reasons. First, cursory observations of the LiDAR data for the study area showed that block sizes within notable talus deposits are in this range. Second, Abbot et al (1998) defined the block sizes to be considered within the RHRA based on their impact on the train. In this system, they note that blocks of > 1m in the vertical dimension on the track have a high potential to cause derailment.
Figure 4-8: Top: Bare earth DEMs of two different slopes within the study area. Bottom: Non-bare earth Returns on top of bare earth returns for same two slopes. Both sites are shown at the same scale and range from roughly 10 to 700 m a.s.l.

All of the parameters used in this analysis were extracted directly from the LiDAR data or derived through the terrain classification outlined above. These parameters were extracted mostly by means of simple GIS tools. Source height was extracted by using the Zonal Statistics tool in ArcGIS Spatial Analyst with the topography for the study area as the input and the rockfall source polygon as the boundary.

Whether or not a runout zone reaches the track was determined by intersecting the track with the runout zone polygons; if any point within the runout zone touches the track, that runout zone was flagged as such.
The percentage area of each terrain classification was calculated by overlaying each runout polygon with the terrain classification results and calculating the proportion of the total area belonging to each terrain class. Slope length was calculated using the length of profile line generated in section 4.3.3. Forested slope length was calculated by determining the proportion of that line which intersects vegetated areas identified through the terrain classification. The parameters regarding tree dimensions and density are all direct outputs of FiNT, which was run once for the entire study area. For a given runout zone we clip the FiNT output to the extents of the runout boundary and extract these parameters at the slope scale. The reach reduction is calculated by applying the process outlined in Berger and Dorren (2007)—this is expressed as the percentage of rockfall events that are likely to be intercepted by the forest cover. Appendix B contains the python script used to apply this process to our data.

4.3.5 Prioritization

A simple probabilistic analysis was carried out using only the parameters generated in the process outlined above. The factors considered were the probability of rockfall occurrence for a given source and the reduction in reach probability provided by the forest cover.

The objective here was to highlight the workflow to extract useful parameters from a single LiDAR and orthoimagery dataset for the purposes of rockfall hazard assessment. It is therefore important to note that the probability of occurrence is oversimplified here. We assumed a uniform distribution of rockfall probability over the entire study area. Therefore, the probability of occurrence for a given source identified through the terrain classification process was quantified simply using its dimensions; larger
sources will have a proportionally larger probability of occurrence than smaller sources. This is over-simplified in that other factors influence rockfall probability, however, for the purposes of this study a uniform spatial probability was used simply to illustrate how the parameters generated through this process can be used to inform the $P_H$ and $P_{S:H}$ factors in the risk equation shown in section 4.1. The reach reduction is quantified using the formula outlined in Berger and Dorren (2007). Finally, the probability of occurrence is multiplied by the reach reduction to calculate our final output, which is the probability of the hazard reaching the track given hazard occurrence ($P_{S:H}$ from the risk equation).

At this point, the process moves from the slope to the element at risk, which is the railway track in this case. The hazard priority at track level is a function of how many runout cones cross that point, this can be defined as the density at track-level. Hazard density is quantified by summing the $P_{S:H}$ when two or more runout zones overlap at a given location on the track.

4.4 Results and Discussion

4.4.1 Terrain Classification Results

Figure 4-9 shows the Gaussian decomposition of slope angles within the study area while Figure 4-10 shows the results spatially. Each Gaussian curve represents one geomorphic unit; the slope threshold for each unit is extracted by finding the intersection between a unit and the next unit below it, as outlined in Loye et al (2009). For example, the slope angle threshold for potential rockfall sources is found by determining the intersection between the steep slope curve and the source curve. This point was found
to be at 50°; therefore, this is the slope angle used to identify potential sources within our study area.

**Figure 4-9**: Results of the slope angle distribution decomposition. The intersection points shown represent the cutoff angle for each morphological unit (Flats/Benches: 0 -17°; Fans: 17 -30°; Talus Slopes: 30 -50° and Rockfall Sources: 50 -90°).
Figure 4-10: Results of morphometric classification using the decomposition of slope angles.

Figure 4-11 is a histogram of the distribution of slope angles within the classified sample dataset from the study area. This distribution shows that exposed rock is dominant at slope angles above 60 degrees, while the 'not rock' classification has an upper tail within the 50-60-degree range. This distribution sheds light on why Loye’s method produced such a wide distribution for rockfall sources; the process involves fitting a normal distribution to each domain but, for rockfall sources, the distribution is uniform at angles above 60 degrees. The wide distribution is not a direct reflection of the data, but an artifact of fitting a normal curve to a uniform distribution, which is an inherent component of Loye’s method.
Figure 4-11 Histogram showing the distribution of slope angle values within the manually classified dataset of 45696 points, discussed in Chapter 3. The black dashed lines and labels represent the results of the Loye et al (2009) method.

Figure 4-12 shows the results of the terrain classification method which combines the use of slope angle and colour. Here, we add a useful piece of information on top of the geomorphic classification. For example, using the Loye et al (2009) method, only
slope angle is considered, therefore talus deposits are identified with no knowledge of surficial material. Colour analysis provides valuable context to this analytical approach; by using colour you’re able to determine if the slope angle meets the criteria of a given morphological unit, and, by using colour, you’re able to know whether or not a recently active talus deposit actually exists. The combination of both the Loye et al (2009) method and the colour analysis proposed here is a powerful combination. Particularly on lower gradient features such as talus deposits and fans, there is a high level of confidence in the colour analysis, therefore, using these results to highlight fans and talus deposits which are void of vegetation would work towards the overarching goal of guiding further data collection at higher resolution for slopes which are of interest in terms of hazard management.

4.4.2 Assessment of Forest Cover Results

Figure 4-13 shows the output from FiNT; each point represents an individual tree derived and measured from the LiDAR inputs. As no control datasets exist for the study area, a thorough validation of these results could not be carried out. However, based on visual interpretation using the results from Figure 4-13, the results appear to align with the imagery.

4.4.3 Slope-Scale Analysis Results

Using four runout cones picked at random, shown in Figure 4-14, Table 4-4 summarizes the parameters being generated for each source identified through the terrain classification process, bounded by the results of the runout zone delineation. The
parameters shown in Table 4-4 were calculated for every runout zone delineated through the modelling exercise. These parameters provide all of the information necessary to calculate the effect of forest cover on rockfall runout as per Berger and Dorren (2007). The additional parameters regarding the terrain classification results could also prove useful in probabilistic hazard assessment, though this was not carried out for this study. Specifically, the amount of exposed talus within a runout cone could be used to comment on the level of rockfall activity on a given slope; this could be done qualitatively or probabilistically. Moreover, the amount of area within a runout cone belonging to the rockfall source classification could be used to identify reaches of slope that have many sources at play and may provide high return in terms of knowledge to be gained from slope-scale data collection using terrestrial LiDAR, for example.

4.4.4 Hazard Prioritization Results

The overarching theme of this work was to demonstrate the ability to extract many useful parameters from a single dataset in the interest of identification and characterization of regional rockfall hazard. To demonstrate how the parameters being derived here can be incorporated into a risk-based hazard management framework (Figure 4-2), we considered two factors. First, we extracted the spatial probability of occurrence for a given source. To do this, we simplified the problem and distributed probability evenly over space. This means that the probability of occurrence \( P_H \) from the risk equation discussed in Section 4.1 is a function of the size of a source. This is an oversimplification in that there are many other factors which can and should be considered when calculating regional rockfall frequency. However, it is sufficient, given the focus of this study. The next component we consider is the protective capacity of
forest cover at the slope scale. Here, we’re using many parameters discussed above as inputs to the process outlined by Berger and Dorren (2007). The final hazard level is the result of multiplying the probability of occurrence by the reduction factor from forests. This result gives us the spatial probability of a hazard reaching an element at risk ($P_{S:H}$ from the risk equation discussed in Section 4.1). The probability of hazard occurrence ($P_H$) tells us the probability of a hazard occurring from a given source while the reduction factor informs the likelihood of that event reaching the track. As noted, this is a simplified approach to assessing these probabilities using only parameters generated from the LiDAR and orthoimagery. This information could easily be paired with frequency magnitude data, event histories, expert knowledge and other information in order to more thoroughly inform probability of occurrence. Figure 4-15 shows the results of this process as a graph next to the classified terrain map for Mile 20 to 25 of the Yale subdivision. The red line represents $P_{S:H}$ while the black line shows the multiplier used to assess the protection offered by forest cover. We can clearly see that the level of hazard varies substantially over space in our study area. There is an inherent bias to larger cliffs due to the method which probability of occurrence was calculated. Moreover, there is a general correlation between the prominence of vegetation and the level of protection from forest cover, as expected.
Figure 4-12: Top-raw orthoimagery for subset of study area around Mile 30 on the CN Yale Subdivision. Bottom-results of classification methodology
combining the Slope Families concept with the use of colour for identifying areas of exposed rock. This figure demonstrates how colour can be used in tandem with the Slope Angle Distribution Method to further refine the terrain classification; this allows for the identification of exposed outcrop and exposed talus deposits rather than just areas which meet the slope angle criteria for a given classification.

Table 4-4: summary information derived through downslope analysis for four rockfall runout zones shown in Figure 4-11.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cliff Height (m)</td>
<td>4.5</td>
<td>16.2</td>
<td>14.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Reaches Track</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Percent Rockfall Source</td>
<td>4</td>
<td>19</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>Percent Exposed Talus</td>
<td>21</td>
<td>15</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>Average Gradient (degrees)</td>
<td>51</td>
<td>57</td>
<td>54</td>
<td>50</td>
</tr>
<tr>
<td>Profile Length (m)</td>
<td>234</td>
<td>1325</td>
<td>1133</td>
<td>565</td>
</tr>
<tr>
<td>Forested Slope Length (m)</td>
<td>31</td>
<td>782</td>
<td>375</td>
<td>320</td>
</tr>
<tr>
<td>Number of Trees</td>
<td>222</td>
<td>10714</td>
<td>4628</td>
<td>2073</td>
</tr>
<tr>
<td>Total Basal Area (m²)</td>
<td>58</td>
<td>4035</td>
<td>1250</td>
<td>945</td>
</tr>
<tr>
<td>Average Tree Diameter (cm)</td>
<td>38</td>
<td>38</td>
<td>27</td>
<td>45</td>
</tr>
<tr>
<td>Average Tree Height (m)</td>
<td>17.5</td>
<td>17.5</td>
<td>13.1</td>
<td>20.7</td>
</tr>
<tr>
<td>Percentage Reach Reduction from Forest Cover</td>
<td>N/A</td>
<td>31</td>
<td>50</td>
<td>15</td>
</tr>
</tbody>
</table>
Figure 4-13: a-Output of FiNT (Dorren, 2018). Each point represents a tree, for which FiNT outputs a height and stem diameter. b-cross section B-B’, black dots represent points placed by FiNT, green points represent non-bare earth LiDAR returns.
Figure 4-14: Results of the runout zone delineation process for 4 cliffs identified through the terrain classification process. Runout ID is the unique ID given to each source at the end of the terrain classification process (Top image is over a hillshade DEM derived from ALS; bottom is over orthoimagery).
Figure 4-15: On the left, the results of the colour classification, right, a graph showing the probability of hazard reaching a track given forest is shown in red while the reduction factor calculated using the Berger and Dorren (2007) method.
4.5 Discussion

In this study we carried out a terrain classification exercise using the analysis of differences in colour values to discern between different surficial material types; specifically, areas of exposed rock and areas of vegetation. These results were combined with the technique proposed by Loye et al (2009) which uses the decomposition of the cumulative distribution of slope angles in order to geomorphologically classify terrain. In combining the results of colour analysis with Loye’s morphometric analysis we’re now able to discern between not only talus slopes and slopes steep enough to be rockfall sources, but we also now know where talus deposits exist in the absence of vegetation. The latter of which is a good indicator of recent rockfall activity because the absence of vegetation hints that there has been either frequent rockfall activity preventing vegetation from growing or an event of high enough magnitude to create a clearing in the vegetation. This output alone can offer a great deal of insight as to the spatial distribution of rockfall activity in a given area. We identify all talus within the runout zone of each source, however, this information is not considered in the probabilistic assessment of rockfall hazard for the study area. As mentioned in Section 4.1, this is a critical piece of information when identifying high hazard areas for rockfall. The process outlined here uses colour and slope angle to identify areas of exposed talus but does provide insight as to how they were deposited. Piteau (1977) notes that our study area is marked by several major rock avalanche deposits; under our system, these deposits would be flagged as talus deposits and makes no distinction between the processes which deposited them; this is critical. Rock
avalanche deposits consist of large blocks of consistent sizes and shape and would likely be covered in mature vegetation such as moss (indicating little activity in recent history).

As a whole, the thresholds determined through Loye’s method correspond to the distribution of slope angles within the manually classified data, though, for rockfall sources defined by Loye’s method, the slope angle threshold is likely lower than necessary. Roughly 44% of all points above 50 degrees are classified as vegetation in Figure 4-11. A large portion of these are likely due to misclassification of surficial material based on colour. However, a small cluster of exposed rock is present at roughly 50 degrees. This could be the result another geomorphic domain not recognized through precursory investigation of the data. This could be a structurally controlled family of bedrock dipping at an angle shallower than most of the outcrop in the area, or perhaps a transitional zone adjacent to near-vertical faces that are common in the study area. The sides of large boulders may account for some of the data within this range.

Vertical occlusions due to limitations of airborne LiDAR data collection are also likely skewing the distribution of slope angles within the rockfall source domain identified through Loye’s method. If vertical slopes were included within this distribution, perhaps the mean value of the rockfall source domain would have been shifted to the right, increasing the kurtosis for that distribution and reducing the area of overlap between the distributions for talus slopes and rockfall sources determined from Loye’s method.

Rockfall deposits, on the other hand, are generally less uniform and, on active sites, will likely not be covered in vegetation. The system developed here could be used
as a starting point for further analysis that will identify the process which likely deposited a given talus accumulation delineated through the terrain classification process. Colour could likely be used as a metric to measure the freshness of the deposit while the presence of and dimensions of vegetation might also offer insight as to the recent rockfall history for a given site. Moreover, the application of work from Bonneau et al (2018) which leverages high-resolution topographic data in order to classify talus deposits in terms of their grain size distribution, could be used to comment on the characteristics of the talus deposits identified through the process developed here.

One could use this information to better inform PH by taking into account the amount and characteristics of downslope talus for a given source; areas of slope with large amounts of talus which has characteristics indicative of a rockfall deposit should be assigned a higher probability of occurrence than a slope with no exposed talus or which appears only to have blocks which were deposited by a single, high magnitude event (e.g. rock avalanche). The runout zone boundaries generated using Flow-R and the attribute information derived for each one, as displayed in Table 4-4, could be the basis for an exercise of this nature.

The runout zone extents were also used to bound further analysis regarding the protective capacity of the forest cover within the runout zone. In order to do so we first generated tree points using FiNT and extracted their diameters using the methodology put forth in Hanus et al (1999). We then extracted all the other parameters required, summarized in Table 4-1, in order to apply the workflow presented by Berger and Dorren (2007). This resulted in a percentage reduction factor offered by the forest cover below each source; simply put, this is a measure of what percentage of rockfalls of a
given volume can be expected to be absorbed by the vegetation downslope of the source. The reduction factor was multiplied by the spatial probability associated with each source to determine the final hazard rating for each source.

Figure 4-16A-D provides an example of how this system could be applied in an operational setting to back-analyse rockfall events. The pink point shown in each section of the figure represents the location of a hypothetical block which has landed on the track at the toe of a forested slope. In Figure 16A, the cyan coloured lines represent all the runout zones which intersect this point on the track. Figure 16B shows all potential sources that correspond to the runout cones mentioned above; these are found using the unique integer ID mentioned in Section 4.3.3. In Figure 16C a cone extending +/-20° of the slope direction is projected from the fallen block; this allows for the potential source location to be narrowed down further as some of these sources are quite wide and extend far beyond this range. Finally, Figure 16D shows the sources within the extent of this cone coloured by their lateral position relative to the fallen block, the idea here being that probability dissipates as lateral position relative to the block increases. A similar relationship could likely exist between the Euclidean distance between the fallen block and a given source. This information could all be ingested in order to provide a slope-scale probability for the precise location of the source for a fallen block. This is directly applicable and useful to railway operations in that it will allow for systematic guidance as to where inspection should be carried out. This is particularly useful on forested slopes where sources may not be visible from track-level or even helicopter inspection.
Figure 4-16: Illustration of how this system can be used in a backwards analysis in the event of a rockfall landing on track. A-the green lines are the extents of all runout cones which intersect the location of the fallen block. These are used in order to find the potential sources which could have released the block, shown in B. C shows the extents of a cone extending +/-20 degrees from the fallen block to all sources which could have been responsible for the block. D shows these
cliffs, coloured by lateral divergence from the fallen block; this is to demonstrate how further research could use the information

4.6 Conclusions

The addition of colour analysis of colourized point cloud data can refine the results of the morphometric analysis using slope angle proposed by Loye et al (2009). This information can be used to prioritize slopes based on the likelihood of rock fall occurrence. Runout zones, used as the zone of influence for a rockfall source defined by the terrain classification provides the necessary spatial bounding to generate a number of parameters from the terrain classification and forest cover data specific to that source. In particular, these spatial boundaries allow you to determine:

- if a rockfall from a given source will reach the element at risk
- how much of the terrain within that zone of influence is marked by evidence of previous rockfall activity
- the average gradient of the zone
- the profile length between the source and the element at risk
- the number, density and average size of downslope tree cover

Together, these parameters can be used to comment on the rockfall hazard associated with a given source. These parameters could be used to inform $P_H$ in a more robust way than demonstrated here (using a uniform probability of occurrence). The system developed here shows promise as a means of assessing rockfall hazard level at the regional scale, but also as a means of determining where fallen blocks of an unknown origin may have come from. Moreover, this system could be used to guide the
Characterization and Monitoring phases from the framework presented in Figure 4-1. Knowing where rockfalls are possible is achieved through the identification of rockfall sources. Determining where they are most likely can be done using the information derived through this process, though this was not explicitly achieved. This study provided a simple example of how parameters derived from a single airborne LiDAR and orthoimagery survey can be used to derive all of the inputs necessary for a regional scale rockfall hazard assessment and demonstrated how these parameters can be incorporated into a probabilistic assessment aimed at quantifying the spatial distribution of rockfall hazards within the study area.
4.7 References


British Columbia Ministry of Forests (2002). Vegetation Resources Inventory: The B.C Land Cover Classification Scheme (v 1.3)


(www.ecorisq.org)

Vierling, K. T. (2016). Beyond 3-D: The new spectrum of lidar applications for earth 
https://doi.org/10.1016/j.rse.2016.08.018

https://doi.org/10.1007/s10706-016-0049-z


susceptibility by integrating statistical and physically-based approaches. 
Geomorphology, 94(3–4), 419–437. 
https://doi.org/10.1016/j.geomorph.2006.10.037

and risk assessment in the Yosemite Valley, 491–503.

Species in the Coastal Regions of the Pacific Northwest. Research Contribution, 
Forest Research Laboratory, Oregon State University, (August), 12.

Group, New York, NY


susceptibility mapping of debris flows and other Atmospheric, 869–885. 
https://doi.org/10.5194/nhess-13-869-2013

Assessment of rockfall risk along roads, (January).

on GIS data, (January).

Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M. H., Loye, A., Metzger, R., & 

113


Chapter 5: Summary and Future Work

5.1 Goals and objectives revisited

We carried out a study which explores the use of high-density 3D information at the regional scale for the purposes of classifying terrain and using that information as an input into a regional rockfall hazard assessment. We developed a classification technique which utilizes the colour values of points within a colourized 3D point cloud in order to classify that point in terms of surficial cover type. We also applied a number of tools and techniques developed by others. Specifically, we used the Slope Angle Distribution method (Loye et al, 2009) to geomorphically classify our LiDAR dataset. We also used FiNT (Dorren, 2017) and the workflow described in Berger and Dorren (2007) to comment on the role of forest cover in rockfall protection within the study area. The overarching goals of this work were:

1) To determine how much information can be derived from a single, regional-scale topographic dataset (we used LiDAR and orthoimagery collected on the same flight in 2015). Figure 5-1 demonstrates how a single airborne LiDAR and orthoimagery dataset was utilized in different ways to generate all the parameters required for this study.

2) Test viability of colour as a metric for classifying colourized 3D data for rockfall hazard assessment.

3) Comment on how results of colour analysis can be incorporated with conventional means of hazard identification and hazard assessment.
4) Incorporate results of colour analysis with the Slope Angle Distribution method from Loye et al (2009).

5) Use identified rockfall sources as inputs for the delineation of rockfall runout zones using Flow-R (Horton, 2013).

6) Use boundaries of rockfall runout zones to classify terrain within the runout zone (area of influence) of each identified source.

7) Use ideas from Dorren (2017) and Berger and Dorren (2007) to quantify the protective capacity of forest within the runout zone of each source.

8) Use the information derived through these processes to quantify the hazard level for rockfall on the CN Yale subdivision from Mile 20 to 25.

5.2 Colour analysis for the segmentation of 3D Data

Here, we tested two different statistical approaches for classifying colourized 3D point cloud data using the colour value; the K-Means clustering algorithm and Random Forest classification. Both approaches achieved accuracy of over 90% when tested using a manually classified control dataset. Both algorithms achieved exceptional results for classifying vegetation, but the Random Forest algorithm out-performed K-Means in classifying exposed rock. Upon investigation of the errors occurring it was determined that the linear nature of K-Means clustering was misclassifying all light coloured (in the top right of the AB space plots) trees; this accounted for the majority of the classification error when tested on a point cloud of 7133257 points; an example of the results for this analysis are shown in Figure 5-2.
Figure 5-1: A-Bare earth LiDAR model—this provides the 3D geometry for all terrain classification carried out. B-orthoimagery—this is where all colour information is derived from. C-the results of the colour analysis providing a fully classified 3D pointcloud.
Figure 5-2: Results of colour analysis from Chapter 3, applied to an area in the Yale Subdivision. Red areas are exposed rock, while green areas are not.

The classification error generated through the Random Forest approach was much less prominent, though also more difficult to characterize. Therefore, on the basis of classification accuracy alone, we concluded that the Random Forest classification approach is preferable to K-Means. Because we had carried out two different classification methods we were able to combine the results in order to reduce uncertainty associated with the classification. In order to do this we added a scalar field to the fully classified point cloud which was the sum of both classifiers—here a value of 0 meant that both classification techniques returned vegetation, a value of 2 meant that both returned exposed rock, and a value of 1 meant that the classifiers were conflicting. The results of this exercise are shown in Table 5-1 where the bottom left, and top right cells represent the number of points where the results between classification techniques
were conflicting. Because of the high level of success of both algorithms, there is a high level of confidence in the points with values of 0 and 2.

Table 5-1: Classification results of both approaches to classification of points

<table>
<thead>
<tr>
<th></th>
<th>Rock</th>
<th>Not Rock</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=7133257 Random Forest</td>
<td>Rock</td>
<td>Not Rock</td>
</tr>
<tr>
<td>K-Means</td>
<td>Rock</td>
<td>1248570 (17.5%)</td>
</tr>
<tr>
<td></td>
<td>Not Rock</td>
<td>39761 (0.6%)</td>
</tr>
</tbody>
</table>

5.3 Combination of Colour Analysis Results with Slope Angle Distribution Method

Loye et al (2009) developed a method for the geomorphic classification of topographic data through the disaggregation of the distribution of slope angles around a mean for each classification. This is based on the idea that geomorphic features within a homogenous area will have a consistent slope gradient. We used this method to classify the study area into flats/benches, fans, talus slopes and rockfall sources (Figure 5-3). Upon this, these were further classified using the results of the colour analysis from the previous section, this is shown in Figure 5-4. Here, for each point we know its geomorphic classification from the Loye et al (2009) method in addition to its classification from the colour analysis.
Figure 5-3: Results of Loye et al (2009) Slope Angle Distribution method

Table 5-2 Summary of results of slope angle distribution analysis

<table>
<thead>
<tr>
<th>Classification</th>
<th>Slope Angle Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flats/Benches</td>
<td>0-17</td>
</tr>
<tr>
<td>Fans</td>
<td>1730</td>
</tr>
<tr>
<td>Talus Slopes</td>
<td>30-50</td>
</tr>
<tr>
<td>Rockfall Sources</td>
<td>50-90</td>
</tr>
</tbody>
</table>

This results in a classification scheme which offers more context to the terrain. For example, now, instead of knowing whether a given area meets the slope angle criteria for a talus slope, we know whether it meets that criteria and is marked by exposed rock, which is indicative of rockfall activity. The caveat here is that the system requires refinement in order to comment on the origin of a talus deposit in order to exclude large
magnitude events which may skew the probability in assessing the frequency of rockfall.

Figure 5-3 shows the results of the colour analysis combined with the morphometric analysis using slope angle.

<table>
<thead>
<tr>
<th>Slope Angle Distribution Method Results</th>
<th>Colour Analysis Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rockfall Source</td>
<td>Rock</td>
</tr>
<tr>
<td>Rockfall Source</td>
<td>Not Rock</td>
</tr>
<tr>
<td>Flat/Bench, Fan, Talus Deposit</td>
<td>Rock</td>
</tr>
<tr>
<td>Flat/Bench, Fan, Talus Deposit</td>
<td>Not Rock</td>
</tr>
</tbody>
</table>

Figure 5-4: The results of the combined approach of using slope angle and colour to classify terrain.
5.4 Delineation of maximum rockfall runout boundary for each source identified through SAD/Colour analysis

We used Flow-R (Horton et al, 2013) to delineate runout boundaries for each rockfall source identified through the terrain classification process. Flow-R was chosen based on its simplicity, limited input requirements and flexibility in terms of parametrization. The output of this process was used to bound further analysis of terrain and forest cover at the slope scale for each rockfall source. We did not calibrate the model for use in our study area, however, we were able to use the parameters suggested by Michoud et al (2012) and Casanova et al (2016) for applying this model to the assessment of rockfall runout. There are several benefits to this model as opposed to physical models. Within the Flow-R model, the terrain exclusively dictates runout propagation without information about the falling block. This is a benefit in that the limitations of airborne LiDAR, noted in Chapter 2, impede the ability to accurately assess block dimensions on vertical slopes. The Flow-R model likely results in more conservative (larger) runout boundaries; this is preferable in our case as we’re using these boundaries, not to identify specific rockfall paths, but to delineate a boundary of influence for a rockfall from a given location. Appendix C contains the rockfall source data, rockfall runout zone boundaries as well as the terrain classification data.

5.5 Slope-scale assessment for the runout zone of each source

The runout zone is the final parameter required to carry out a slope-scale assessment of each rockfall source identified by the terrain classification process. They were used to spatially bound the analysis of terrain and forest cover. For each runout cone we calculated:
1) the percentage area within the runout zone of exposed talus deposits;

2) the percentage area within the runout zone of potential rockfall sources;

3) the average gradient within the runout zone;

4) the endpoint of the runout zone (does it reach the track?);

5) the track mileage extents of all runout zones which reach the track;

6) length of profile line extending from source to lowest point within the runout zone;

7) the size, number and density of trees within the runout zone; and

8) the length of forested slope along the profile line.

Using these parameters, we applied the process developed by Berger and Dorren (2007) in order to calculate the probable reduction in reach potential for rockfalls initiating from a given source. This parameter was used as a modifier in the probabilistic hazard assessment, discussed below.

5.6 Quantification of rockfall hazard at track level by means of parameters generated through this process

The purpose of this research is to demonstrate how a single high resolution, regional-scale LiDAR and orthoimagery dataset can be leveraged in order to generate all of the parameters required to assess rockfall hazard in a given area. Given this, the final element of the work consisted of synthesizing all the data generated through Chapters 3 and 4 into a quantitative hazard assessment for the study area. For the sake of simplicity, we made the assumption that the probability of rockfall occurrence is distributed evenly through space within the study area. This means that the probability
of occurrence associated with a given rockfall source is a direct function of its dimensions.

The probability of occurrence for each source was multiplied by the hazard reduction factor discussed previously. This gave us the final output, which is the probability of a rockfall reaching the track.

5.7 Future Work

The colour technique demonstrates utility as a means of classifying colourized point cloud data as rock or not rock. However, it would be beneficial to test the limits of this technique. The analysis employed here could be used as the initial step in a process aiming to perform surficial geology mapping using colour. That is, in the first step, we isolate rock, next, we use the identified rock areas into another iteration of the analysis which performs a similar analysis in order to find differences in lithology or surficial conditions of rock. This could potentially offer value in hazard assessment in that links between hazard occurrence and geological setting could be made.

Pairing the results of the colour analysis with morphometric analysis using slope angle is a logical direction. However, there is potential to further classify identified talus deposits in terms of block size and number; Bonneau et al (2018) demonstrate how this type of analysis may be performed. This information would provide insight into the frequency and size of rockfall events in a given area.

In Chapter 4 we propose using the data generated as the input to a backwards analysis process in the event of a fallen block from an unknown source being located on the track. The results of the modelling output are used first in order to determine any
runout zones which cross the point at which the fallen block is located. Next, we use the unique IDs of those runout cones to isolate the sources which could have been responsible for the fallen block, these are the polygons shown in blue. Any brown rockfall source polygons do not have a runout zone that crosses the point of the fallen block on the track. We then propose the idea of projecting a cone extending +/-20° from the source lateral from the direction of slope. This is done in order to further refine the results in terms of the most likely candidates for sourcing the fallen block; probability will dissipate with lateral extents. There is an opportunity for research into how to effectively assign probability based on these factors. This system would be extremely useful for railway operators in that it will allow for methodical identification of likely rockfall sources. In the case of forested slopes, this is a difficult task in the absence of a system such as this.

5.8 Conclusions

Regional scale datasets such as airborne LiDAR are capable of providing a multitude of information regarding the geology, geomorphology and forest cover of terrain within a given area. Though the resolution and vantage point are not equal to those of terrestrial data collection mediums such as TLS. The vantage point has traditionally been seen as a limiting factor in terms of the use of LiDAR for rockfall studies in steep terrain. However, there are techniques, demonstrated here, which leverage the strengths of airborne LiDAR, specifically when paired with colour derived from orthoimagery. The analysis of colour in this case provides a great deal of context to the already rich 3D data. The combination of slope angle and colour in terrain classification offer a novel approach to targeting features of interest when assessing
rockfall; talus accumulations and rockfall sources. The addition of runout extents generated by means of runout modelling allow for information derived from the LiDAR and orthoimagery to be assessed at the slope-scale as opposed to regional-scale, at which many of the parameters are initially generated in this study. Forest parameters are also assessed in a similar manner; initially generated at the regional scale but assessed at the slope scale using these boundaries. Ultimately; we have presented a new approach to the evaluation of terrain over large areas for the purposes of rockfall hazard assessment in mountainous terrain.
5.9 References


Appendix A: Colour Classification Testing

Figure A-1: Results of K-Means Classification on a randomly generated sample dataset of 2500 points

<table>
<thead>
<tr>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock</td>
<td>X (n = 13368)</td>
</tr>
<tr>
<td>Not Rock</td>
<td>O (n = 30726)</td>
</tr>
</tbody>
</table>

X (n = 1051)
Figure A-2: Scatter plot of Random Forest classification on randomly generated sample dataset of 2500 points, 95% total accuracy
Figure A-3 Results of Random Forest model built using incrementally larger sample data sets, applied to the entire point cloud of 7 million points
Figure A-4: Results of K-Means Classification on the full dataset for the study area. Cluster centres for both the validation dataset and the full dataset are shown.
Appendix B: Code for Calculating Protective Capacity of Forest

This python code is the author's interpretation of the workflow presented in Berger and Dorren (2007). This is the core process within Rockfor.net which is aimed at calculating the reduction factor in rockfall reach potential given forest cover and geometry of a slope. Use of this code should be referenced as follows:


```python
import arcpy, sys, os, math
from arcpy.sa import *

fc =

# Given Variables
cd = 30.0000 # total distance between curtains, fixed at 30
density = 2500.0000 # density kg/m3 see chart on rockfornet website
FE_ratio = 0.9000 # fracture ratio of tree species
g = 9.8100 # gravity (m2/second)
pi = math.pi
Rdiam = 1 # diameter of rock (m)
volume = 1 # rock volume
m = volume * density # mass of rock (kg)

# Calculated variables

cursor = arcpy.da.UpdateCursor(fc, ["profile_length_2D", "average_dbh", "Cliff_Height", "DBH_m2_per_Ha", "talus_percentage", "reduction", "cliff_percentage", "avg_slope", "DBH_m_per_Ha_Required"], 'objectid in (1260,3269,784)')

for i in cursor:
    try:
        # Loop variables
```
slope = i[7]
pct_unforested = (i[4]+ i[6])
pct_forested = (100 - pct_unforested)
slope_length = i[0] * (pct_forested/100)

## i[7] = slope_length## forested slope length is calculated by finding the 2d profile length from source to track, and then using the percentage forest cover ratio from terrain classification

d = i[0]# travelling distance of rock
DBHm = i[1] * 100# mean DBH of forest on slope cm
DBH = i[1]
Fh =i[2] ## height of cliff
G = i[3]# total basal area of slope--from data

print i[3]
E_pot = m*G*h # Potential Energy

max_E_diss = FE_ratio * slope * DBHm ** 2.31
vmax = math.sqrt(2*g*h) # cannot be higher than 0.64 * slope gradient if the the
## total basal area is at least 10m2. If G is less than 10m2 or if forest cover is
## absent, maximum velocity is assumed to be 0.8*slope gradient (0.8 and 0.64
were
## derived from rockfall experiments carried out by Berger and Dorren (2006)

## Energy Dissipation

E_totd = 0.5*m*vmax**2+m*g*0.25*Fh

## GenTheo is the basal area which a rock theoretically encounters after travelling
## a given distance through a forest with a given basal area
gen_theo = (d*Rdiam) * G/10000
Nr_tree_impacts = (gen_theo * 4) / (pi * (DBH)**2)

Nr_required_curtains = Nr_tree_impacts * (100/66)

## Max_E_Diss is the maximum energy that can be dissipated by one tree, given in J

## FE_Ratio the fracture energy ratio of a given tree species

max_E_diss = FE_ratio * slope * DBHm**2.31

## Dissipated energy factor

dEcfactor = E_totd / (Nr_required_curtains * max_E_diss)

Nr_req_curtains = E_totd / (max_E_diss * dEcfactor)

## print Rdiam, slope_length, G

Gavailable = ((Rdiam) * (slope_length))/ (10000) * (G)

Grequired = Nr_req_curtains * pi * 0.25 * DBH*DBH

Nr_req_curtains = E_totd / (max_E_diss * dEcfactor)

DBHr = (E_totd * cd/slope_length*FE_ratio*dEcfactor*slope)**1/2.31

PRH = 100-(Gavailable * 100 / Grequired)

print PRH

i[5] = PRH

cursor.updateRow(i)

except:
    ***
Appendix C: GIS Dataset Overview

A GIS database is available on request to accompany chapter 4 of this thesis. The data is in ArcGIS File Geodatabase format. The database includes a number of datasets described in the table below:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrain Classification</td>
<td>Raster dataset containing results of colour analysis and slope angle distribution analysis.</td>
</tr>
<tr>
<td>Sources</td>
<td>Polygon dataset containing rockfall sources within study area. Dataset contains a Cliff_ID field which can be used to relate information from runout zone to source</td>
</tr>
</tbody>
</table>
| Runout Zones           | Polygon dataset containing runout extents for all sources identified through terrain classification process. Dataset contains a number of attribute fields:  
  - Cliff_ID – unique ID to be used to link to cliff dataset  
  - Reaches_Track—binary classification which describes whether a runout reaches the track  
  - Slope length—length of a profile line extending from the source to the lowest point within the runout zone  
  - Profile gradient—gradient of line used to calculate slope_length  
  - Pct_source—percentage of area within the runout zone covered by rockfall sources  
  - Pct_exposed_talus—percentage of area within the runout zone which is covered by exposed talus as identified by the terrain classification process  
  - Avg_DBH—the average DBH of all trees identified by FiNT  
  - Num_Trees – number of trees identified by FiNT  
  - Tree_density—number of trees per m²  
  - Slope – average gradient of runout cone in degrees |