LANDSLIDE MORPHOLOGY AND ITS INSIGHT INTO THE TIMING AND CAUSES OF SLOPE FAILURE: CASE STUDY OF POST-GLACIAL LANDSLIDES IN YELLOWSTONE NATIONAL PARK

by

Grace Ellen Nicholas

A thesis submitted in partial fulfillment
Of the requirements for the degree

of

Master of Science

in

Earth Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2018

ACKNOWLEDGMENTS

First and foremost, I would like to thank my graduate advisor Dr. Jean L. Dixon for her constant guidance throughout my Master's program. Jean, thank you for all the field trips and pushing me to always be a better scientist.

I owe a huge thanks to my committee member, Dr. Ken L. Pierce for taking me into the field on several occasions and imparting his wisdom on the geomorphology of Yellowstone National Park onto me. And thank you Ken for our countless meetings where you guided me through every scientific tangent I had throughout this program. Thank you, Dr. Dave Lageson for serving on my graduate committee, your structural expertise was critical for this project.

To the wonderful members of the Earth Surface Process Lab past and present, Aaron Feldhaus, Colin Quinn, and Kailey Adams, thank you for your constant support. This program would have been much different without all of you by my side.

Thank you, Tobias Goldman for your assistance during fieldwork. I don't think I would have been able to auger that many sites without your help!

I would also like to acknowledge my wonderful friends and family for always believing in me and supporting me in all my adventures.

And last but certainly not least, I would like to acknowledge my husband, John H. Nicholas, none of this would have been possible without your love and guidance.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ABSTRACT

Landslides are ubiquitous to post-glacial landscapes worldwide. Withdrawal of glacier ice exposes over-steepened landscapes that may be unstable, and consequently susceptible to landsliding. Glacial debuttressing may directly destabilize slopes; however, seismicity and transitions to interglacial climates associated with greater effective moisture and subsequent degradation of permafrost may also play a role. Here, we explore disparate potential mechanisms of slope failure in a set of post-glacial landslides in northwest Yellowstone National Park. We quantify spatial relationships, topographic metrics, and relative age of eight landslides within the north entrance to the park, a system traversed by over 700,000 visitors every year. Analysis of high-resolution topography indicates increasing surface roughness of non-active landslides southward. These roughness values in ancient slides are roughly half those of the active Slide Lake Landslide, and suggest younging ages along the retreat path of the Yellowstone Ice Cap, consistent with glacial debutressing as the likely trigger for these slides. However, roughness values and their application for relative age dating are strongly confounded by topographic biases such as gullying, fluvial erosional contacts, and anthropogenic features (e.g., roads, structures). Once roughness biases are removed, we find roughness differences between landslides decrease, and do not support younging ages along the path of ice retreat. Thus, glacial debuttressing most likely only had a preparatory influence on slope failure, and was not the direct trigger. Analysis of subsurface soils at landslide toes indicate a >17 plasticity index, pointing to highly expansive clays that are sensitive to moisture addition. Stratigraphic relationships between post-glacial terraces and soil analyses suggest a late Pleistocene (~13 – 11.5 ka) timing for slide initiation, a period coincident with high available moisture. Stream power analysis indicates that Holocene incision of the Gardiner River is focused at a knickpoint locally coincident with the toe of the active Slide Lake Landslide, providing a mechanism for modern, local reactivation of the ancient slides. Together, our findings broadly show how quantifying the temporal and spatial patterns of landslides can be diagnostic of the controls on slope failure, and can be used to understand risk. They also highlight the importance of careful site calibrations and bias removals in roughness analysis.

INTRODUCTION

Landslides are ubiquitous to post-glacial landscapes worldwide (Ballantyne, 2017; Booth et al., 2017; McColl, 2012), and multiple mechanisms can explain their occurrence. Determining the temporal and spatial distribution of post-glacial landslides can aid in determining the perturbation in which caused these slopes to fail. Paraglaciation (the conditioning of landscapes by prior glaciation) is widely considered a major predisposing factor in landslide occurrences (Cossart et al., 2014; Darnault et al., 2012; Ivy-Ochs et al., 2009). Throughout a glacial cycle, factors influencing slope stability change such as ice distribution, hydrologic conditions, and seismicity (McColl, 2012). Landslides may have several preparatory factors but only one trigger, the external stimulus that causes a near-immediate failure (Varnes, 1978). Because landslides are discrete events, their timing and spatial distribution can provide important insight into what factors are responsible for preparing or triggering slope failure (Booth et al., 2017).

As glaciers advance, they widen and deepen valleys, creating the classic "U" shape thereby steepening adjacent hillslopes. Glacial retreat debuttresses these slopes, increasing shear stresses (E. Cossart et al., 2008), and possibly triggering slope failure upon removal of ice (McColl, 2012). As a preparatory factor, the unloading of ice would reduce the stability of the slope without triggering failure, and other processes would reduce stability further, triggering failure sometime after the removal of ice.

Climatic transitions (i.e. change into a warmer interglacial cycle) are also widely linked to landslide occurrence (Crozier, 2010; Fischer et al., 2013; Huggel et al., 2012), and the redistribution of the water table during deglaciation may provide a climate-driven

trigger for slope failure in post-glacial systems (Gruber & Haeberli, 2007). The triggering effect of the relocation of water across the landscape during warm periods is primarily through the rapid generation of meltwater and consequent production of high pore pressures, decreasing the normal force holding materials on the slope and producing a lubricated failure surface (Wang et al., 2014). Thus, an increase in temperature leading to deglaciation would be associated with slope failures. Furthermore, the increased discharge caused by meltwater and high-post glacial sediment load would increase the local fluvial networks' ability to incise (Jansen et al., 2014); changing local base-level for hillslopes and furthering the potential for failure.

The unloading of glacial ice and isostatic uplift has also been linked to changing seismicity, providing an additional potential trigger for landslides over glacial cycles (Etienne Cossart et al., 2014; Darnault et al., 2012; Hormes et al., 2008). A large potential control on seismically induced post-glacial landslides is due to the influence of topography and ice cover on intensity of seismic shaking (Mccoll et al., 2012). Generally steep slopes, like those left by retreating glaciers, amplify seismic shaking. Additionally, ice cover has been found to dampen seismic shaking, protecting the slopes during glacial periods and leading to increased likelihood for seismically-triggered landslides following glacial retreat (Mccoll et al., 2012)

Distinguishing a triggering versus preparatory role for these direct and indirect controls on post-glacial slope failure is challenging. We take advantage of a series of eight landslides following the retreat of the northern Yellowstone outlet glacier, which pose a unique opportunity for comparing slope failure. Here, we use surface roughness,

derived from high-resolution topography, as a quantitative proxy for age. We compare the temporal and spatial distribution of these landslides with the known pattern and timing of glacial retreat determined through cosmogenic ^{10}Be exposure dating of glacial moraines in this system (Licciardi & Pierce, 2008), and to regional climatic changes determined from paleo-climate reconstructions of lake sediment cores (Krause & Whitlock, 2013; Whitlock et al., 2006). Disparate mechanisms for slope failure are associated with unique timing relationships between slides; slides triggered by glacial debuttressing should be coincident with and follow patterns of ice retreat, while climatic wetting or seismic induced triggers would cause regional landslides of comparable ages.

Our work unveils both the promises and pitfalls of using surface roughness as a proxy for depositional age. Past studies have used this method for both relative and absolute ages in disparate systems (Booth et al., 2017; Frankel & Dolan, 2007; LaHusen et al., 2016a). Here, we have found that if surface roughness is to be used as an indicator of landslide age, both the post-depositional geomorphological changes and historical anthropogenic modifications to the landscape must be intimately investigated to account for all roughness biases that could bias the inferred landform ages. Yellowstone National Park has been heavily influenced by humans for the last 200 years and archival evidence has shown many modifications to the landscape that could not have been identified remotely or in the field, thus, robust historical accounts must be included in surface roughness investigations in landscapes that have been under human influence over historical timescales.

APPROACH AND METHODS

Field Site & Geologic Setting

Our field site is located just within the North Entrance of Yellowstone National Park, south of Gardiner, MT. We focus on a series of eight landslides that trend north to south along the Yellowstone (landslides 1-3) and Gardiner Rivers (landslides 4-8) (Figure 1), historically referred to as the Gardiner landslides (Fraser et al., 1969). The Gardiner landslides are sourced from Sepulcher Mountain and occur within the late Cretaceous Landslide Creek Formation, composed of dark-gray, commonly conglomeritic sandstone interbedded with varicolored mudstone and claystone, and bentonite (Fraser et al., 1969).

Along this roughly 8 km transect, stream elevations rise ~150 m southward towards the Central Plateau of Yellowstone National Park. The Yellowstone Plateau is defined by a region of high-uplift associated with the current location of the Yellowstone Hotspot; a thermal mantle plume, 350 km in diameter located underneath Yellowstone National Park (Nelson & Grand, 2018). The North American plate has migrated southwest over the mantle plume, beginning near the Oregon-Idaho-Nevada border ~16 Ma until reaching its current location near the Idaho-Montana-Wyoming border. This succession of caldera forming volcanism along the migration path has resulted in the distinct topography of the East Snake River Plain, a ~90-km wide trench bounded by two arms of high-uplift. At the apex of this tectonic parabola is the Greater Yellowstone area in which the Gardiner landslides are located (Figure 1). This arcuate shape of high-relief is coincident with increased seismic activity much of which occurred during the

Holocene (Pierce & Morgan, 1992; Russo et al., 2017). North of the Gardiner landslides (~50 km), the Emigrant fault offsets Quaternary alluvial fans by about ~4-6 m (Personius, 1982), evidence that seismic activity has occurred in our study region following the Pinedale glaciation.

The greater Yellowstone area has also undergone extensive modification under the Pleistocene Bull Lake (~200-130 ka) and Pinedale (~23-14 ka) glaciations. Uniquely, the Yellowstone Hotspot has greatly influenced the geographical extent of these two glacial periods. Bull Lake terminal moraines reach further down valley than the Pinedale moraines in the areas south, southwest, and west of Yellowstone. This is due to the southern region of Yellowstone National Park occurring at higher elevations during the Bull Lake glaciation because this area was coincident with the uplifting leading margin of the Yellowstone hotspot, whereas, during the Pinedale glaciation the migrating hotspots uplifting leading margin was located in the northern region of Yellowstone National Park while the lower elevation tailing margin of the hotspot was located in the southern region of the park (Pierce et al., 2014). Glacial ice originated in the mountains surrounding Yellowstone and flowed onto the plateau. As glaciers on the Yellowstone Plateau built up to ~1000 m glaciers then flowed down the major valleys that drain Yellowstone National Park. One such feature was the Northern Yellowstone Outlet glacier that flowed past the Gardiner landslides, 65 km down Paradise Valley to its terminus (~ 16 ka) south of Livingston, MT (Licciardi & Pierce, 2008; Pierce, 1979; Pierce et al., 2014) (Figure 1).

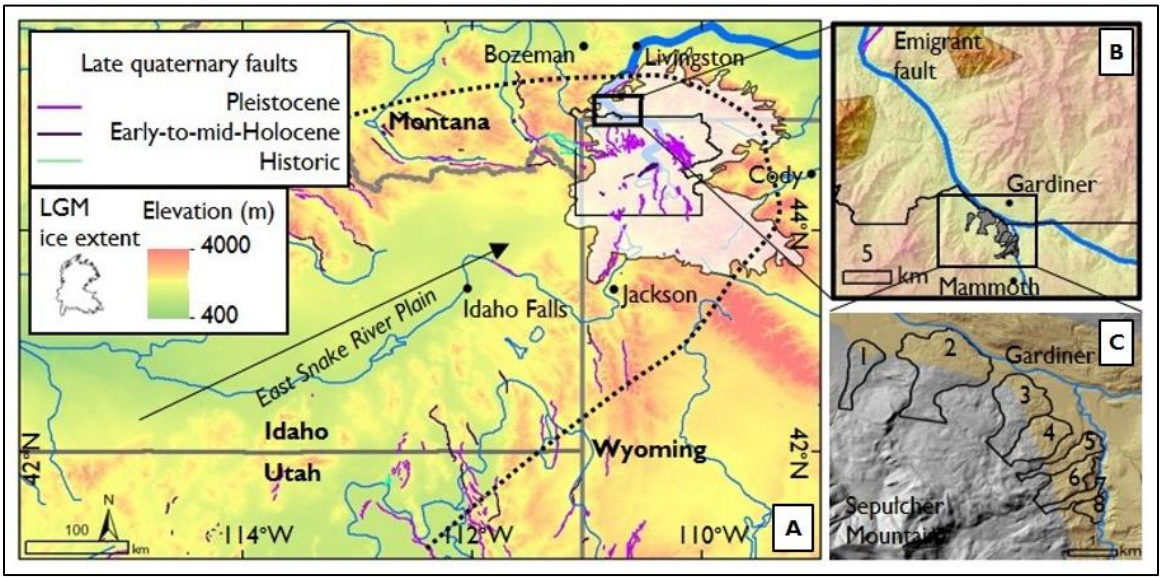

Figure 1. (A) Shows the study area in relation to the Yellowstone crescent of high terrain (dotted line). White transparency represents the maximum extent of glacial ice during the Last Glacial Maximum (LGM) (Licciardi & Pierce, 2008). Arrow represents the direction of migration of successive caldera forming volcanics and the Yellowstone hotspot. Colored lines represent late quaternary faults. (B) Area showing the Gardiner Landslides (grey) in the south and the Emigrant fault to the north. (C) Close look at the Gardiner landslides (black outlines) sourced from Sepulcher Mountain, numbers represent landslide units, these will remain the same throughout, and the tan shading is the available LiDAR in which we used to calculate roughness.

Surficial Mapping and Sampling

The Gardiner landslides were originally mapped as moraines, and their form was ascribed as molded under glacial ice (Weed, 1893). Waldrop and Hyden (1963) revised this classification, describing and mapping the overall structure and topography of these features as mass movements. We build on their original coarse-scale mapping of, and refine landslide boundaries through remote mapping of the area using high-resolution (1-meter) LiDAR bare-earth imagery (Figure 1) provided by Yellowstone National Park and made available through the Montana Bureau of Mines and Geology ArcGIS server (MBMG Geographic Information Systems, 2018). Although LiDAR data is not

continuous over the entire study site, coverage is sufficient to encompass all landslide toes.

Initial field observations noted that landslide margins were coincident with small drainages and topographic depressions. We utilized topographic wetness index (TWI) to aid in initial mapping of landslide boundaries and separating of adjacent deposits. TWI, which combines local upslope contributing area and slope, is commonly used to quantify topographic control on hydrological processes (Sorensen et al., 2006):

$$TWI = \frac{lnA_i}{tan\theta_i} \qquad \text{(Equation 1)}$$

where, θ_i is the local slope and A_i the upslope accumulation area for each cell i. Higher TWI values represent drainage depressions, while lower values represent crests and ridges.

TWI was calculated using 10-meter USGS digital elevation data, processed in ArcGIS. Raster cells of high TWI were targeted for field mapping, to validate boundaries based on exposed deposit stratigraphy and surface morphology. Landslide morphology was compared across the set of 8 deposits, including length of runout, vertical height, height to length ratio, mean slope, mean elevation and landslide area (Table 1). Here, height is defined as the elevation relief (uppermost minus lowermost elevations) for each slide, and length is defined as the linear distance from the top of the deposit to the distal edge of the landslide toe.

Field expeditions consisted of mapping all Pleistocene and Holocene features including landslides, glacial deposits, glacial landforms, and glacially molded surfaces and were informed by previous studies in the area (Pierce, 1979; Pierce et al., 2014;

Waldrop & Hyden, 1963). We explore material mobilized by the landslides to investigate if these landslides failed in glacial till by augering to resistance at three sites within the toe of each of eight landslides. Coring sites were coincident with areas of low slope on the apex of hummocks near the toes of landslides. These sites were selected to minimize local erosion. In addition to the coring sites, areas exposed by fluvial incision and road cuts were also investigated to understand landslide composition. Other geomorphic landforms, including glacial deposits and stream terraces were mapped to help identify timing relationships.

Our original research design included measuring thicknesses of time-dependent carbonate rinds on lithics collected from the landslide deposits to be used as a proxy for age in addition to surface roughness. However, landslide deposits consisted mainly of shales, with no available cobble-size materials for carbonate rinds to accumulate. Thus, surface roughness serves as our primary quantitative proxy for relative landslide ages. Samples were also collected from road-cuts for geotechnical analysis of rock and soil properties to help mitigation efforts by the Yellowstone Park associated with the North Entrance road that lies along the toes of both ancient and active landslides (Dewey et al, in-press). Cut slope samples were collected from two of the landslides (5 and 6) and analyzed for water content, particle size distribution, conductivity, resistivity, and pH (Figure 2).

Surface Roughness Analysis

We determine surface roughness to be used as a proxy of relative landslide age. Here, surface roughness is measured as the standard deviation of slope (SDS; Equation 1; (Berti et al., 2013; Frankel & Dolan, 2007; Lahusen et al., 2016), using 1m-resolution LiDAR-derived topographic data:

$$SDS = \left[\frac{1}{N^2}\sum_{i=1}^{N^2}(m_i - m_m)^2\right]^{0.5} \qquad \text{(Equation 2)}$$

N = the width of the roving window (m)
m_i = slope of the local cell (i) centered in the moving window
m_m = mean slope within the moving window

Available LiDAR data (Figure 1) covers the distal end of the landslides in our study area, providing sufficient area for our surface roughness analysis. Elevation data was extracted separately for each landslide, using digitized landslide polygons as an extraction mask, and each landslide was analyzed separately to avoid any bias in processing from adjacent cells at landslide boundaries. Standard deviation of slope was calculated as a focal statistic for each cell within the landslide, using several roving window sizes. Mean roughness was then calculated as the average SDS for each landslide. All data processing was performed using ArcGIS and MATLAB software.

Applications of SDS as a landslide age proxy assume that the active landslide processes create roughness in topography that become muted by colluvial and sheet-wash properties following emplacement. Applying this relative age proxy assumes that initial roughness of fresh deposits are similar, and that the diffusion rate is similar across all landslides (Booth et al., 2017; Lahusen et al., 2016). These basic assumptions are likely

sound in our study system as these slides are sourced from the same underlying lithology, and aspect and elevation differences are trivial.

While roughness applications assume landslide-systems smooth over time due to erosional processes (Booth et al., 2017), other unintended processes may also influence roughness (Lahusen et al., 2016). We recognized several such processes within our study system, and term the associated regions as 'roughness biases'. We noted two major bias classes: hydrological and human infrastructure. Observed hydrological biases include gullying, fluvial undercutting, and lakes, while human infrastructure biases include buildings, roads & road cuts, and parking-lots.

To account for roughness biases, we map their individual extent in each of the landslides. One set of hydrological biases were associated with >2m deep gullies that developed within landslide deposits. These were identified and removed as 10m buffers around stream features (pixels identified by with > 600,000 m^2 accumulation area threshold). Biases associated with steep undercutting of the Gardiner and Yellowstone Rivers at the toes of several landslides were removed as 100m buffers along both rivers. Slide Lake, a 3500 m^2 feature within landslide 7, was also removed, and slopes adjacent to its shoreline were removed using 5m buffers. Human infrastructure, including roads, buildings, and associated or other modified topography were hand digitized and buffered to varying lengths (5m for buildings and parking-lots, and 10m for roads). All biases were removed from the LiDAR digital elevation data prior to terrain processing.

Statistical analysis of roughness data was processed in MATLAB, and we compare the contrast of landslide roughness both prior to and following bias removal.

Two statistical analyses were used to test variance of landslide morphology. First, following the methods of Frankel and Dolan (2007), we test the likelihood that individual landslide roughness values were derived from unique populations using the Kolmogorov-Smirnov test. This test does not require prior assumptions regarding the distribution of sample populations. This is useful in comparing large data sets and data sets of different sizes, as is the case in landslide roughness data (Frankel & Dolan, 2007). This test quantifies the maximum difference between the cumulative frequency distributions of two sample populations. We also compared landslide roughness using a one-way, fixed effects, normal analysis of variance (ANOVA) test with post-hoc Tukey-Kramer test using Tukey's Honestly Significant Difference procedure (HSD). We utilize this test to determine whether the data from the eight landslides have a common mean. ANOVA tests are used to determine if mean landslide roughness's are distinct (Stoline, 2016). All tests were performed using Statistics and Machine Learning toolbox in the MATLAB software.

Stream Power Analysis

We explore the relative gradients along rivers within the study area to understand the broader geomorphic context for the Gardiner Landslide system. Rivers typically display concave longitudinal profiles that are well-described by a power-law relationship between local channel slope (S) and the contributing upstream drainage area:

$$S = K_s A^{-\theta}. \quad \text{(Equation 3)}$$

Here, S is the local channel slope, K_s is the steepness index, A is the upstream drainage area (m^2) and θ is a concavity index. The perspective that the fluvial systems of erosional landscapes are in a dynamic equilibrium allows for perturbations in channel form to be detected (Kirby & Whipple, 2012). We calculate normalized steepness indices (K_{sn}) for the Gardiner and Yellowstone Rivers as:

$$K_{sn} = K_s A^{(\theta ref)}, \quad \text{(Equation 4)}$$

to investigate the relationship between increased channel steepness and location of landslides in our study area.

Normalized stream steepness was calculated using ArcGIS- and MATLAB-based tools of Whipple et al. (2007). For this analysis, K_{sn} was extracted at a 2.5km window from 10m elevation data, using 0.45 as the reference concavity index (θ_{ref}). We were primarily interested in relative channel steepness across the landslide region, identification of which was not strongly influenced by the parameters chosen.

RESULTS

Surficial Mapping & Sampling Analysis

Mapped landslides are characteristic of earthflows, following the classification of Varnes (1970) based on their characteristic hourglass and elongate shape, and the fact that mobilized material that is predominately fine-grained and sourced from clay-bearing rock. All flows are sourced from the north and east flanks of Sepulcher Mountain, and flowed down-gradient towards the Yellowstone and Gardiner River (Figure 1). A set of rotational landslides were mapped south of our southernmost slide (8). However, these features are distinct in their rotational block-slide morphology, with arcuate defined headwall scarps, whereas, the landslides in the northern region exhibit hummocks and closed depressions within the well-defined earthflow shape described above. In addition, this area marks the transition to a region dominated by travertine deposition. We exclude these features from our roughness analysis as they violate necessary assumptions of uniform lithology and initial landslide morphology.

The set of landslides (1-3) in the northern region flow towards the Yellowstone, while the southern slides (4-8) flow toward the Gardiner River, its adjoining tributary. All remaining slides appear to extend to the modern base level set by the current river, except slides 1 and 4 which are deposited atop the upper surface of a Pinedale outwash terrace. This outwash deposit is visible across the Gardiner River from flows 4 and 5, with two terrace levels (T1) (Figure 2 and 6) comparable on either side of the river.

Accommodation space for the landslides within the river valley decreases southward along the confinement of the Gardiner Canyon, resulting in narrower slide widths for the southern slides (5-8) subduing the hour-glass shape that occurs when the toe of a slide can distribute laterally. The Gardiner Canyon is characterized by vertical cliffs of Cretaceous shales and sandstones on the east side of the river while landslides occupy the entire western side. Two southern slides (6 & 7) currently interact with the river and have been historically accounted as being major problems for the North Entrance road. Along the toes of these landslides, fluvial incision has exposed a boulder deposit (B1) (Figure 6) atop Cretaceous bedrock related to the east walls of Gardiner Canyon that was over-run by landslide 6 at the time of failure. The deposit is exposed ~10 m above the current Gardiner River (elev. ~1670 m) and continues north beneath the landslide (6) for ~50 m until lost under landslide material. Here, the Gardiner River has incised through the landslide and flood boulders, and into underlying shale bedrock. Another boulder deposit (B2) (elev. 1742 m) (Figure 6) was observed across from these two slides (6 & 7) on the east side of the river. This boulder deposit is also atop Cretaceous bedrock but overlain by glacial outwash and appears to be pro-glacial deposits with stone-on-stone fabric composed of cobble to boulder sized stones of mostly volcanic basalts and Precambrian rocks. Just south of B2 at roughly the same elevation (1745 m) this boulder deposit appears again (B3) preserved atop a resistant bedrock layer that facilitates base-level for current deposition of alluvial fans (~50 m) above the current Gardiner River and 35 m above the southernmost mapped outwash terrace (T2) (Figure 1 and 6).

Traversing atop the landslide deposits we observed extensive fluvial erosion at the lateral margins of the landslides, creating deep gullies (>3m) along landslide boundaries in addition to local gullies (~2 m in depth) incised into landslide deposits. Remobilization along the toe of slides 6 and 7 occurred in 1915 during a time of high-precipitation and Slide Lake landslide has been documented as a problem since the onset of construction of the North Entrance road in 1903 (Pitcher, 1903).

Deposit material analysis through observation of slide material cored from the toes of each landslide confirm Cretaceous shales and bentonite beds as the dominate material mobilized during failure. The three samples collected for geotechnical material analysis presented in this study are S-4, S-5, and S-6 (Figure 2). Samples consist primarily of lean brown clay, with lesser sand content. Water content was high, averaging 38.9% (24.5, 62.1 and 30.1% for samples S-4, S-5, and S-6). All samples follow classify as highly plastic clay (CH), with high plasticity indexes (PI) ranging 25-50, following the Unified Soil Classification System (USCS).

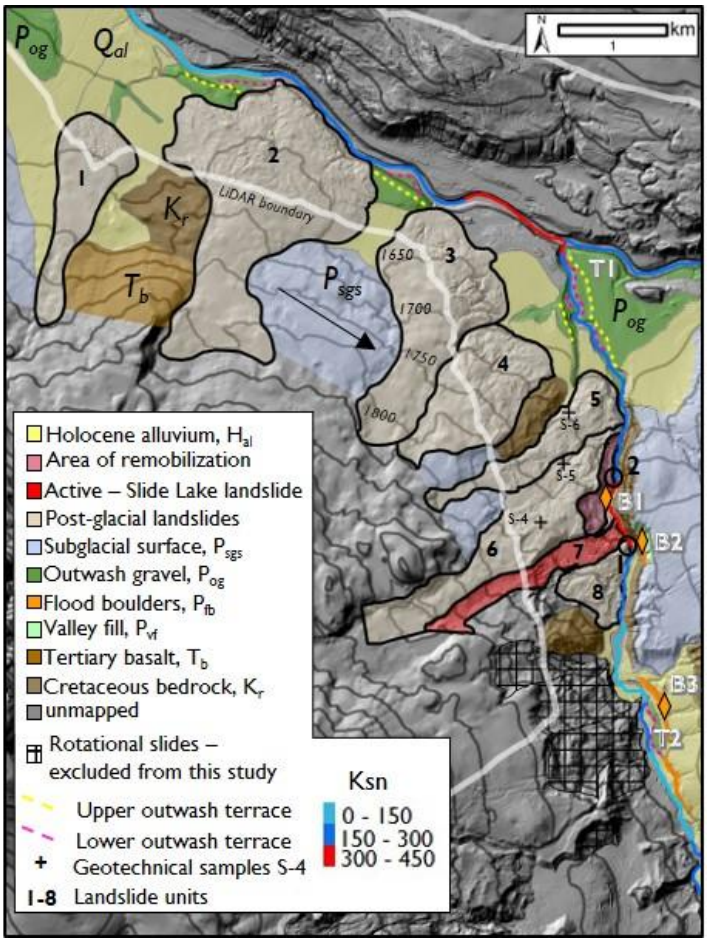

Figure 2. Surficial geologic map of the study area and results from our stream power analysis (Ksn). Arrow shows the direction of retreat of the northern Yellowstone outlet glacier. Crosses are locations of cut-slope samples (S-4 to S-6) for material analysis. T1 & T2 represent upper and lower outwash terraces and diamonds represent boulder deposits (B1-B3) plotted in profile space in figure 6.

Morphological & Surface Roughness Analysis

Landslide morphological analysis (Table 1) reveals on slide footprints between 0.15-1.37 km^2, averaging ~0.5 km^2. These variable areas do not appear to have a discernable pattern N to S, and also do not co-vary with landslide runout lengths of 435-1163 km (averaging ~800 m). Runout lengths are all similar with lengths of 806-1163 m, except for slide 1 and 8 which have lengths of 457 m and 435 m, respectively. Height to

length ratios, and average elevations and slope gradients are similar (Table 1). These physical characteristics could influence surface roughness values, our quantitative proxy for age. While we fine some correlation between mean SDS and landslide area (r^2 = 0.34), importantly we found no correlation with mean standard deviation of slope (SDS post-biases removal) and landslide gradient or H: L ratio (r^2 = 0.06).

Landslide ID	Area (km^2)	Length (m)	Height (m)[1]	Height:Length ratio	mean slope (°)	mean Elevation (m)	mean SDS (°)[2]	mean SDS (°)[3]
1	0.16	457	45	0.10	6.3	1679	3.3	3.2
2	1.37	985	85	0.09	6.6	1685	2.9	2.4
3	0.46	467	65	0.14	8.4	1698	3.2	2.9
4	0.62	806	105	0.13	10.7	1739	2.8	2.8
5	0.31	1163	143	0.12	9.1	1725	3.8	3.3
6	0.45	965	149	0.15	10.6	1821	3.7	3.1
7	0.15	963	133	0.14	10.5	1806	5.3	5.1
8	0.21	435	69	0.16	9.1	1736	4.4	3.2

[1] Height defined by the uppermost elevation minus the lowermost elevation of each landslide.
[2] Average standard deviation of slope (SDS) calculated on landslide data prior to removal of roughness biases.
[3] Average standard deviation of slope (SDS) calculated on landslide data following the removal of roughness biases.

Table 1. Morphological attribute of the Gardiner landslides measured from 10m DEM.

Surface roughness (SDS, both the analysis before-and-after removal of roughness biases) of the active Slide Lake landslide (slide 7) is roughly 2x higher than that of the assumed ancient landslides for (Table 1). Mean SDS values of the pre-biases removal

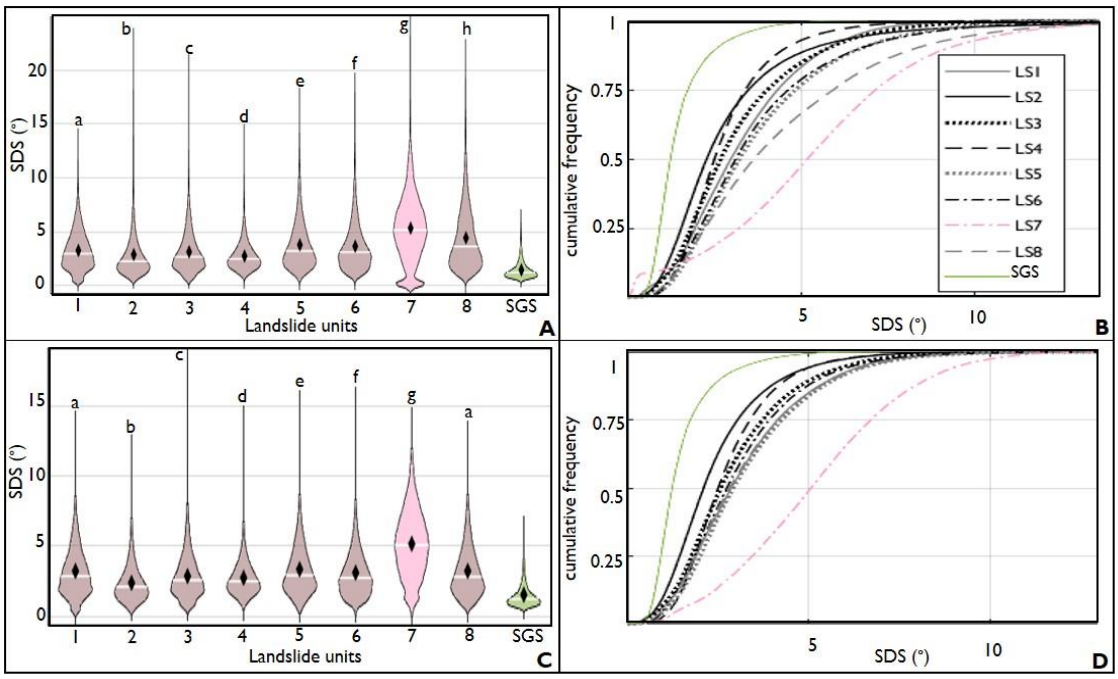

Figure 3. Distribution of standard deviation of slope (SDS) for eight landslides before removal of roughness biases (A) and after removal of biases (C). White lines within violin plots represent the median of the data while the diamonds show the mean value. Violin plots highlight data density at SDS values, and extend to the most extreme data points. Letters above violins represent groups of landslides highlight statistically distinct groups, as determined by post-hoc Tukey analysis. Cumulative distribution plots of SDS before (B) and after the removal of biases (D) highlight the importance of roughness biases. The green line represents the subglacial surface (SGS) while the pink line is the active Slide Lake landslide (7).

data show a greater data distribution spread and an increasing trend of roughness sequentially with landslide units. Conversely, roughness results after removal of biases show no systematic increase in roughness moving south (Table 1, Figure 3). The roughness values for the subglacial surface represent topography smoothed by basal sliding during glaciation and show a marked decrease in roughness compared to that of the modern and ancient slide surfaces (Figure 3). Mapped landslides were statistically distinct based on both the Kolmogorov-Smirnov test and one-way ANOVA (for both, $p<0.01$). The post-hoc Tukey-Kramer Honestly Significant Difference test found that

mean roughness values for all landslide pairwise comparisons (28 comparisons for 8 landslides) were significantly different ($p<0.01$) prior to removal of biases. Post-biases analysis was similar, except slide 1 and slide 8 were not distinct ($\rho=0.17$) (Figure 3).

Analyzing the roughness of the various biases categories we found that fluvial incision along the toes of landslides, gullying, and road-cuts present with higher average

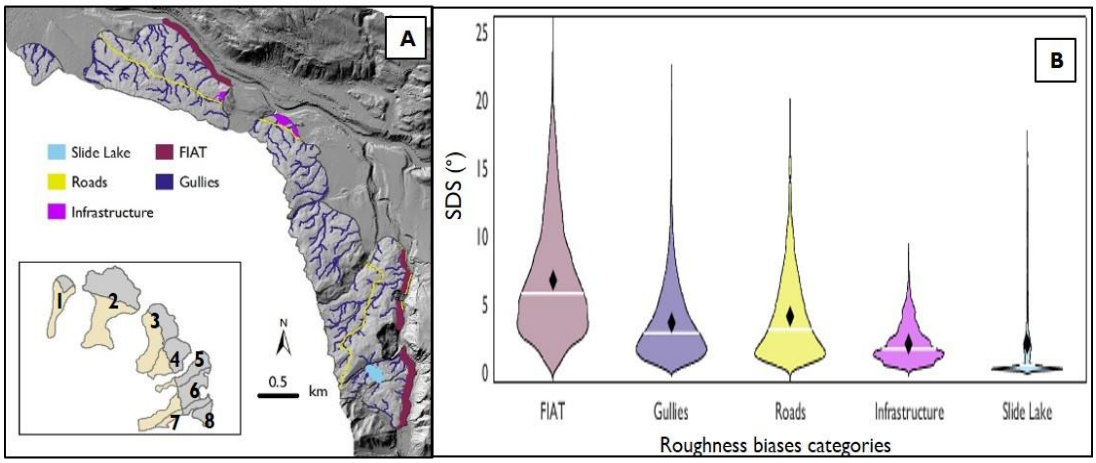

Figure 4. (A) Map of roughness biases. (B) Data distribution of surface roughness for the five biases categories shown in (A). Black diamonds represent mean SDS, and white lines represent median values. Colors correspond to map colors in (A).

roughness values than the landslides while infrastructure (in the field these are dominantly parking lots) and lakes present with lower roughness values roughly to that of the ancient in-active landslides (Figure 4). Of the biases categories, gullying accounted for bias area of greatest spatial extent, averaging 17% of the available LiDAR coverage removed from each landslide. Fluvial incision along the toe of landslides accounted for an average of 11% landslides area. Roads, infrastructure and lakes represented on average less than 5% landslide area.

The size of the roving window used influences derived roughness metrics because they capture different topographic components of the landscape. We tested the influence of selected roving windows and confirmed that a 15 x 15 m roving window was ideal to capture the scale of hummocky topography of the landslides in this system.

By plotting mean roughness versus varying window sizes (Figure 5). The dominant surface wavelengths are represented by the steepest part of the curve, with the greatest

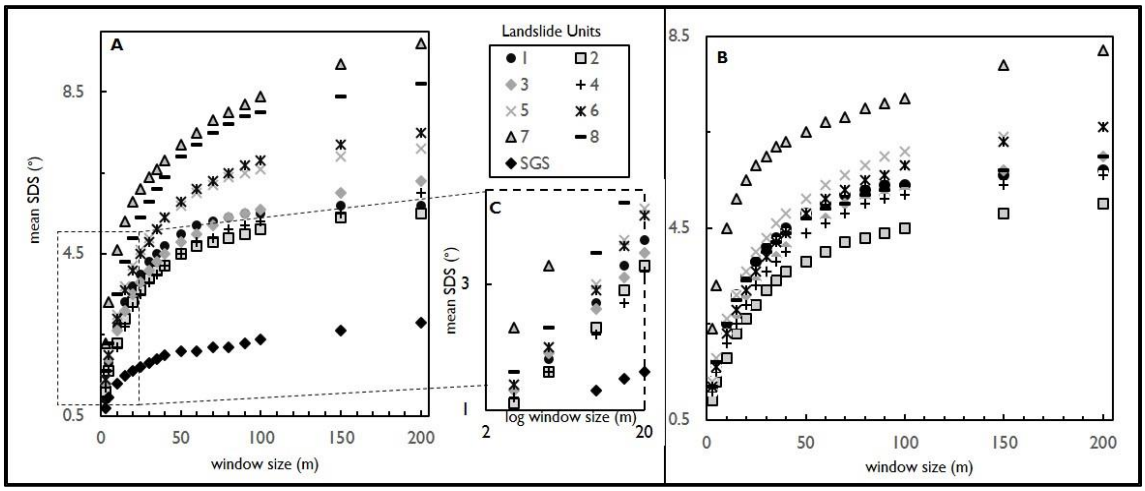

Figure 5. Plots A-C showing mean surface roughness versus window size used to calculate the standard deviation of slope values. Panel (A) shows results before biases were removed and panel (B) shows results following removal of biases. Window sizes range from 3 m by 3 m to 200 m by 200m. Dominant surface roughness wavelengths are represented by the steepest parts of the curve. SGS represents roughness for the subglacial surface, the surface in which the landslides over-run. Panel (C) is a zoomed-in view of the variability of roughness values at smaller roving windows. Notice panel (C) has logarithmic axes. The landslide unit symbols correspond to landslide numbers presented in figure 2.

spread in roughness values occurring around a 15-m window size (Figure 5c). Changes in average surface roughness with changing window sizes was investigated for both datasets; before-and-after the removal of roughness biases (Figure 5). We expected to see a convergence in mean roughness at the longest length scales, however, the pre-biases

removal data results show distinct grouping of mean roughness for landslides at larger window sizes (~40 to 200 m). This grouping phenomenon diminishes after biases are removed and roughness values essentially converge at the longer length scales, aside from the active Slide Lake landslide. Additionally, we measured the mean roughness from a subglacial surface, a planar, moderately sloping topographic end member to which the irregular landslide topography is working to attain. Average surface roughness for the subglacial surface is around half that of the ancient in-active landslides.

Stream Power Analysis

Stream power analysis indicates indicate notable variation in normalized stream steepness indices (K_{sn}) along the Gardiner and Yellowstone Rivers. We find increased stream steepness coincident with the toe of the active Slide Lake landslide and portions of remobilization along slide 6 (Figure 2). This steep section along the Gardiner River coincides with a knickpoint visible in both the river longitudinal profile and plot of local stream gradient versus drainage area (Figure 6). A second, segment of increased steepness occurs at and downstream of the confluence of the Gardiner and Yellowstone Rivers. This is most likely due to a change in relative base-level for the Gardiner River due to the higher erosive power of the larger trunk stream; the Yellowstone River (Kirby & Whipple, 2012). Along this section, the river passes through the town of Gardiner, MT, and river banks have been heavily armored with rip-rap to reduce erosion.

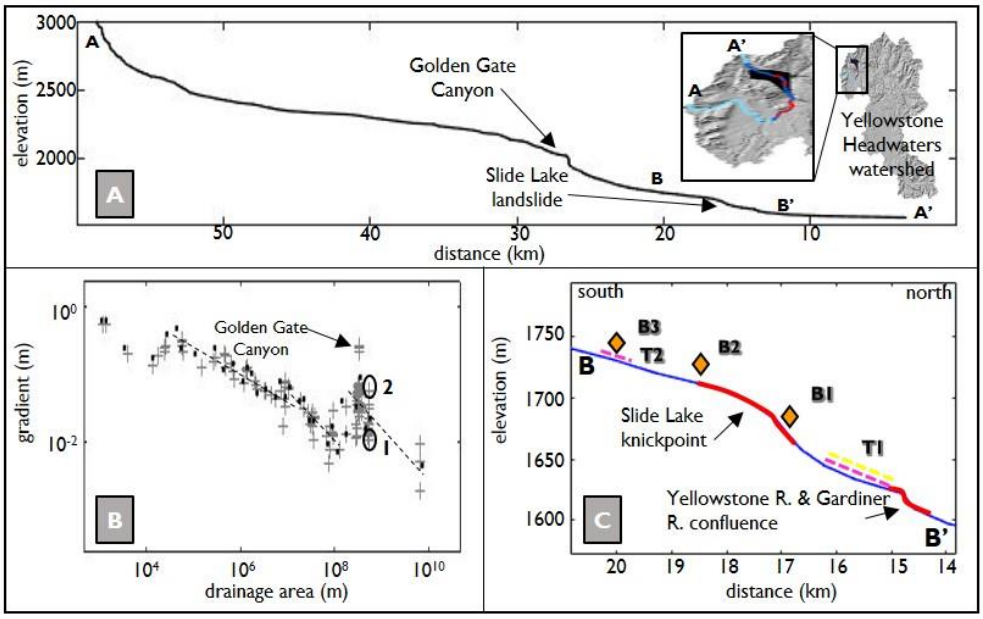

Figure 6. (A) Longitudinal profile across Yellowstone Headwaters watershed (A-A'); highlighted in inset map). The river section B-B' represents the reach coincident with the southern slides (4-8). Golden Gate Canyon, a well-known attraction in Yellowstone National Park displays a major knickpoint. The Slide Lake landslide is also coincident with a lesser knickpoint. (B) Plot showing channel slope versus drainage area (crosses), gray squares are log-bin averages of the slope-area data. Dashed lines represent the profiles predicted by the regressed channel concavity, θ_{ref}. (C) Longitudinal profile of the section of the Gardiner River that interacts with the southern slides (B-B'). The symbols correspond to symbols mapped in Figure 1. B1-B3 are boulder deposits mapped in figure 2 and T1-T2 represent northern and southern outwash terrace surfaces mapped in figure 2.

Mapped boulder deposits occur at similar elevations to the upper Pinedale outwash terraces (T1). The lower outwash terraces were incised sometime after the deposition of the boulders and incision of the upper outwash terraces (Figure 2).

DISCUSSION

Importance of Site Calibration for Surface Roughness Analysis

Slide Lake landslide, a known active landslide resulted in an increased roughness twice that of the ancient dormant landslides for both roughness analyses, before and after the removal of roughness biases. This shows that quantifying roughness remotely is useful for identifying active landslides and areas within landslides that have been remobilized. Roughness age-proxies rely on the assumption that depositional topography subdues over time (Booth et al., 2017; Frankel & Dolan, 2007; LaHusen et al., 2016b), however post-emplacement landscape modifications may confound this analysis. However, our analysis demonstrates the importance of first understanding the post-depositional modifications within a site of interest before surface roughness analyses can be useful for age determinations. We identify multiple potential biases in roughness analyses, and find that mean roughness values for each landslide differ significantly between pre- and post-biases removal, with changes of between 0.1 to 1.2 degrees (Table 1). Considering that most biases resulted in increases in observed roughness, these biases could result in younger inferred ages, or incorrect comparison of relative ages if biases are not equally represented in each landslide.

Equally as important to identifying and removing roughness biases is the size of roving window used to calculate roughness (Figure 5). Different window sizes integrate over and capture different landscape domains (Frankel & Dolan, 2007; Milodowski et al., 2015). For example, large window sizes (e.g., 30 x 30 m) fail to capture the wavelength

of the hummocky surface topography of landslides, and instead record longer wavelengths reflecting the convexity of hillslopes incised by tributary channels (Frankel & Dolan, 2007). The rate of initial increase in roughness with increasing window size differs between the pre- and post-biases removal analyses. Roughness values calculated with windows varying from 3m to 50m increase more rapidly prior to bias removal than following biases removal (Figure 5). This is most likely due to biases increasing roughness dominantly at smaller wavelengths. Thus, it is important to understand landscape signals generated by roughness biases at sites that are to be investigated using surface roughness as a relative age proxy.

Surface Roughness Variations across the Gardiner Landslides

Statistical analyses show significant differences in surface roughness, determined by SDS. All slides have higher average roughness than the subglacial surface, the glacially-smoothed pre-failure surface of Pinedale age (~15 ka) (Licciardi & Pierce, 2008) the surface from which the landslides are sourced. The active, Slide Lake landslide has observably highest surface roughness compared to all other slides, consistent with its younger age. Among the inferred ancient landslides, the northernmost and southernmost flows were found to have similar mean roughness values and roughness distributions, while others were distinct (Figure 3).

When determining the temporal distribution of landslides using surface roughness it is important to consider what a significant difference among average roughness suggests for this system. Both the Kilmogorov-Smirnov and ANOVA tests show that

surface roughness of landslides can be distinguished statistically (Frankel & Dolan, 2007). These topographic datasets (before and after removal of roughness biases) have large populations (> 100,000 SDS measurements), giving a high power to detect differences, however, this also leads to smaller differences being considered as significantly different. This is important because determining the chronological history for the Gardiner landslides requires not just assessing if roughness differences are statistical, but also if they are meaningful.

Chronology for a set of Holocene landslides in Washington State using surface roughness in combination with absolute dating of carbon-14 observed SDS values ranging from 4 – 8 degrees corresponding to 50 to 12,000 years in age, determined using a best-fit exponential function calculated from both observed and modeled age versus roughness curves (Lahusen et al., 2016). This range of four degrees in average SDS corresponds to an age range covering 11,950 years, and suggests that very small differences in average surface roughness can represent large differences in landslide age. Specifically, in the Washington system differences of 0.6 and 1.3 in mean SDS were associated with 1600 and 6300 year differences in landslide ages, respectively (Lahusen et al., 2016). If these calibrations are transferable, then the 0.9-degree SDS difference observed between landslides 2 and 5 (mean SDS of 2.4 and 3.3 respectively) could represent a difference of upwards of thousands of years in age. Though determining such absolute age differences confidently requires site-specific calibration, this comparison highlights how quantifying surface roughness provides a statistically significant method for distinguishing relative ages for ancient landslides with large data populations.

Timing of Slide Initiation

Field mapping made clear that all Gardiner landslides date to after the removal of glacial ice from the region containing the Gardiner landslides, as they are detached from a glacially sculpted surface of Deckard Flats age (15.1 ± 1.2 ka; Pierce et al., 2018) and are mostly graded below late Pinedale age terraces to the current level of the Gardiner and Yellowstone Rivers. Although landslides 2 and 4 do not reach the current river level, this is likely because of shortened runouts over Pinedale terraces, not because they are older and graded to a higher base level pre-Holocene incision (Figure 2). Just north of landslide 1 are mid-channel flood deposits that occur at an elevation of ~1600 m and are dated to 14.4 ± 1.2 ka (Pierce et al., 2014) and recalibrated using updated production rates from (Lifton et al., 2015). A flood of this magnitude would have modified the distal ends of the Gardiner landslides, however, we so no flood modification to the toes indicating the Gardiner landslides were emplaced sometime after 14.4 ± 1.2 ka. These superposition relationships also indicate that the Gardiner landslides occurred following the major pulse of post-glacial incision, and therefore once most, if not all, glacial ice had retreated. Combined these relationships provide an uppermost constraint on landslide age of ~13 ka. Based on surface roughness differences between ancient slides and the Slide Lake system, slides are older than (likely far older than) the 150 years over which the Slide Lake system has been observed as repetitively active.

Controls on Post-Glacial Slope Failure in this System

Our analyses described above provide quantification of surface roughness to determine relative age difference amongst the Gardiner landslides, stratigraphic relationships between the landslides and glacial deposits to determine age boundaries for these slides, and stream power indices to better understand the influence of fluvial interaction on landslide occurrence. But how can these relative age relationships inform use about preparatory factors and triggers that initiated these slides.

Our surface roughness results show little evidence for a systematic increase in roughness of landslides as you move south along the path of retreat of the Yellowstone Outlet Glacier which you would expect with glacial debutressing as a trigger. In addition, these slides have been deposited atop post-glacial terraces suggesting that most if not all glacial ice had retreated from this system and the current river levels assumed before these landslides occurred. However, glacial debuttressing probably served as a preparatory factor for these landslides, by oversteepening the slopes preparing them for failure but not causing a landslide.

Instead, we find a small range in roughness (2.4-3.2) over the ancient Gardiner landslides that suggest these landslides are of distinct, but similar ages. This area is characterized by increased seismicity associated with the Yellowstone Hotspot and there has been Holocene surface offset of ~4-6m along a fault ~50 km north of the Gardiner Landslides dated to 12,000-10,000 years using diffusional scarp degradation methods (Personius, 1982; Pierce and Morgan, 1992). The most active seismic region of Yellowstone National Park occurs on the northwest margin of the Yellowstone caldera

along the Montana and Wyoming border (Figure 1). Western Montana has experienced at least 18 earthquakes of magnitude >5 from about 1809 to present (Wong et al., 2004). The epicenter for the Hebgen Lake earthquake (magnitude 7.3) is ~60 km southwest of our study area. This was the only historic earthquake in Montana to rupture at the surface (Wong et al., 2004), and triggered a large landslide that damned the Madison River forming Earthquake Lake (Wallace, 1980). Thus, seismicity is a likely candidate for triggering landslides in this system.

Most factors contributing to shear strength and stress of a hillslope are adversely affected during or a result of increased hydrologic conditions. These include: cohesion (water affecting soil suction), bulk density (water affecting weight), and pore-water pressure (decreasing the normal force holding the material on the slope) (Wang et al, 2014). Cohesion and internal friction may also be affected by climatic factors promoting weathering processes (wetting, drying, shrinkage, expansion, dissolution, etc.), which cause deterioration of material structure and mass strength but are generally more preparatory than triggering factors in slope failure (Crozier, 2010). Geotechnical analyses of landslide material collected from road cuts traversing landslides 5 and 6 are characterized by high Plasticity Indexes (>25) indicating highly plastic expansive clays. The Plasticity Index is the difference between the measured liquid and plastic limits. The liquid limit is regarded as the water content that is sufficient to allow clay particles to slip past one another, decreasing shear strength of the material (Warkentin, 1961) and increasing probability of failure. The liquid limit of the three samples (S-4, S-5, and S-6) (Figure 2) was 55, 67, and 52 respectively. This means that these landslides only need a

minimum of 50% water content to reach a liquid state. While the plastic limit of the landslide material is low 25, 25 and 50 for samples S-4, S-5, and S-6, respectively. Thus, these landslides need even less water (~25%) to deform plasticity by encroaching down-gradient.

Paleoclimate reconstruction studies using climate proxy data collected from Lake Core sediments located southeast of our study area, just downgradient from the wasting Northern Yellowstone outlet glacier found increasing summer insolation during the late Pleistocene causing slight increases in summer temperatures but cooler than present as indicated by low $^{\delta 18}O$ suggesting low evaporative summer losses from lakes. Additionally, pollen records suggest that from ~13-11.5 ka the landscape was dominated by a Spruce Parkland suggesting a cooler and wetter climate than present (Krause & Whitlock, 2013; Whitlock et al., 2006).

During the last glacial maximum, ice thicknesses in our study region likely reached ~3000 m (Licciardi and Pierce, 2008). The flowing mass of ice atop the Gardiner landslides would have provided overburden pressures of ~27 MPa, compacting Cretaceous shales and Bentonites, driving out water held within the deposits, and decreasing pore pressures. During deglaciation, overpressures would be released, allowing the shales and Bentonites to rebound, and water to be re-incorporated (Hantschel and Kauerauf, 2009; Phanikumar & Amrutha, 2014). This unconsolidation, subsequent reincorporation of water into the clay rich deposits, and increasing pore-pressure was likely a slow process, possibly providing a mechanism for the lag between deglaciation and landslide initiation.

Significant differences in mean roughness were found between landslides (4-5) while, landslides 1 and 8 have means that are not significantly different suggesting these slides are the youngest features after the Slide Lake landslide. Because we have found that small variations in mean SDS can result in large age distinctions we believe these landslides were not triggered in one event, rather they were perpetuated by increased hydrologic conditions found 13,000 and 11,500 years ago. Movement, rapid and slow, of the Slide Lake landslide has been recorded as being triggered by increased precipitation has been recorded in 1906, 1915, and 1917. July 1906 the Slide lake landslide failed twice each time producing upward of 2,000 cubic meters of material deposited in the Gardiner River. These incidents were both recorded as being triggered by high water due to increased precipitation (0.64 inches) for the month of July in 1906 (NWCC Report Generator, 1901-2018). And on April 23, 2018 the toe of landslide 8 was reactivated after a high precipitation event lasting 72-hours and resulting in 0.72 inches of rainfall, recorded by the National Oceanic and Atmospheric Administration. Increased roughness values in the southern region are attributed to Holocene localized incision maintained by the reactivation of the distal ends of landslides that interact with the Gardiner River. Thus, we suggest that the Gardiner Landslides were likely facilitated by increased moisture along with increased seismicity during the transition from glacial to interglacial climate and that paraglacial processes acted as preparatory factor for these slopes to fail.

CONCLUSION

We determine the temporal and spatial relationships of eight landslides that occur along the path of retreat for the Yellowstone Outlet Glacier to understand causal relationships between paraglaciation and slope failure. Our surface roughness analyses show that caution should be shown when using roughness as an indicator of depositional age without first identifying and removing potential sources of bias. However, this method shows great promise for identifying active landslides remotely and could help aid in mitigation and hazard protocols, as the active slide in this system consistently show significantly rougher surfaces than ancient slides, regardless of bias removal. Slope material analysis results show that the Gardiner landslides occur within deposits of high clay content with high plasticity indices, indicating the landslide is comprised of highly expansive materials that are sensitive to water. Stream power analysis indicates that localized Holocene incision of the Gardiner River is focused at a knickpoint locally coincident with the toe of the modern and active Slide Lake Landslide, providing a mechanism for reactivation of the ancient slides. Our combined results provide a strong case that no single mechanism is likely responsible for causing these landslides to fail. We find strong evidence that greater effective moisture combined with the high-clay content of mobilized material and increased seismicity during the transition from glacial to inter-glacial climates together led to slope failure sometime during the late Pleistocene (~13-11.5 ka) suggesting that glacial debuttressing was not a direct cause for slope failure, rather that the Gardiner landslides were initiated by a complex interplay between glacial debutressing, increased effective moisture and seismicity.

REFERENCES CITED

Ballantyne, C. K. (2017). Paraglacial geomorphology, *21*(February 2001), 1935–2017.

Berti, M., Corsini, A., & Daehne, A. (2013). Comparative analysis of surface roughness algorithms for the identification of active landslides. *Geomorphology*, *182*, 1–18. https://doi.org/10.1016/j.geomorph.2012.10.022

Booth, A. M., LaHusen, S. R., Duvall, A. R., & Montgomery, D. R. (2017). Holocene history of deep-seated landsliding in the North Fork Stillaguamish River valley from surface roughness analysis, radiocarbon dating, and numerical landscape evolution modeling. *Journal of Geophysical Research: Earth Surface*, *122*(2), 456–472. https://doi.org/10.1002/2016JF003934

Cossart, E., Braucher, R., Fort, M., Bourlès, D. L., & Carcaillet, J. (2008). Slope instability in relation to glacial debuttressing in alpine areas (Upper Durance catchment, southeastern France): Evidence from field data and 10Be cosmic ray exposure ages. *Geomorphology*, *95*(1–2), 3–26. https://doi.org/10.1016/j.geomorph.2006.12.022

Cossart, E., Mercier, D., Decaulne, A., Feuillet, T., J??nsson, H. P., & S??mundsson, T. (2014). Impacts of post-glacial rebound on landslide spatial distribution at a regional scale in northern Iceland (Skagafj??rur). *Earth Surface Processes and Landforms*, *39*(3), 336–350. https://doi.org/10.1002/esp.3450

Crozier, M. J. (2010). Deciphering the effect of climate change on landslide activity: A review. *Geomorphology*, *124*(3–4), 260–267. https://doi.org/10.1016/j.geomorph.2010.04.009

Darnault, R., Rolland, Y., Braucher, R., Bourlès, D., Revel, M., Sanchez, G., & Bouissou, S. (2012). Timing of the last deglaciation revealed by receding glaciers at the Alpine-scale: Impact on mountain geomorphology. *Quaternary Science Reviews*. https://doi.org/10.1016/j.quascirev.2011.10.019

Dewey, C.S., Pierce, K.L., and Nicholas, G. E. (2018). *Geological studies on the North Entrance Road: Phase 3-Collaboration on alternatives for the road corridor, Yellowstone National Park, Slickenside Consulting, LLC.*

Evans, S. G., & Clague, J. J. (1994). Recent climatic change and catastrophic geomorphic processes in mountain environments. *Geomorphology*, *10*(1–4), 107–128. https://doi.org/10.1016/0169-555X(94)90011-6

Fischer, L., Huggel, C., Kääb, A., & Haeberli, W. (2013). Slope failures and erosion rates on a glacierized high-mountain face under climatic changes. *Earth Surface Processes and Landforms*, *38*(8), 836–846. https://doi.org/10.1002/esp.3355

Frankel, K. L., & Dolan, J. F. (2007). Characterizing arid region alluvial fan surface roughness with airborne laser swath mapping digital topographic data. *J. Geophys. Res*, *112*. https://doi.org/10.1029/2006JF000644

Fraser, G. D., Waldrop, H. A., & H.J., H. (1969). Geology of the Gardiner Area Park County Montana. *Geological Survey Bulletin*, *1277*, 1–118.

Gruber, S., & Haeberli, W. (2007). Permafrost in steep bedrock slopes and its temperatures-related destabilization following climate change. *Journal of Geophysical Research: Earth Surface*, *112*(2), 1–10. https://doi.org/10.1029/2006JF000547

Hantschel, T. and Kauerauf, A.I., 2009. *Fundamentals of basin and petroleum systems modeling*. Springer Science & Business Media.

Hormes, A., Ivy-Ochs, S., Kubik, P. W., Ferreli, L., & Maria Michetti, A. (2008). 10Be exposure ages of a rock avalanche and a late glacial moraine in Alta Valtellina, Italian Alps. *Quaternary International*, *190*(1), 136–145. https://doi.org/10.1016/j.quaint.2007.06.036

Huggel, C., Allen, S., Deline, P., Fischer, L., Noetzli, J., & Ravanel, L. (2012). Ice thawing, mountains falling-are alpine rock slope failures increasing. *Geology Today*, *28*(3), 98–104. https://doi.org/10.1111/j.1365-2451.2012.00836.x

Ivy-Ochs, S., Poschinger, A. v., Synal, H. A., & Maisch, M. (2009). Surface exposure dating of the Flims landslide, Graub??nden, Switzerland. *Geomorphology*. https://doi.org/10.1016/j.geomorph.2007.10.024

Jansen, J. D., Codilean, A. T., Stroeven, A. P., Fabel, D., Hättestrand, C., Kleman, J., et al. (2014). Inner gorges cut by subglacial meltwater during Fennoscandian ice sheet decay. *Nature Communications*, *5*(May), 1–7. https://doi.org/10.1038/ncomms4815

Kirby, E., & Whipple, K. X. (2012). Expression of active tectonics in erosional landscapes. *Journal of Structural Geology*, *44*, 54–75. https://doi.org/10.1016/j.jsg.2012.07.009

Krause, T. R., & Whitlock, C. (2013). Climate and vegetation change during the late-glacial/early-Holocene transition inferred from multiple proxy records from Blacktail Pond, Yellowstone National Park, USA. *Quaternary Research*, *79*(3), 391–402. https://doi.org/10.1016/j.yqres.2013.01.005

Lahusen, S. R., Duvall, A. R., Booth, A. M., & Montgomery, D. R. (2016). Surface roughness dating of long-runout landslides near Oso , Washington (USA), reveals persistent postglacial hillslope instability, (m), 111–115.

https://doi.org/10.1130/G37267.1

Licciardi, J. M., & Pierce, K. L. (2008). Cosmogenic exposure-age chronologies of Pinedale and Bull Lake glaciations in greater Yellowstone and the Teton Range, USA. *Quaternary Science Reviews*, *27*(7–8), 814–831. https://doi.org/10.1016/j.quascirev.2007.12.005

Lifton, N., Caffee, M., Finkel, R., Marrero, S., Nishiizumi, K., Phillips, F.M., Goehring, B., Gosse, J., Stone, J., Schaefer, J. and Theriault, B., 2015. In situ cosmogenic nuclide production rate calibration for the CRONUS-Earth project from Lake Bonneville, Utah, shoreline features. *Quaternary Geochronology*, *26*, pp.56-69.

MBMG Geographic Information Systems. (n.d.). Retrieved March 25, 2018, from http://www.mbmg.mtech.edu/gis/gis.html

Mccoll, S. T., Davies, T. R. H., & Mcsaveney, M. J. (2012). The effect of glaciation on the intensity of seismic ground motion. *Earth Surface Processes and Landforms*, *37*(12), 1290–1301. https://doi.org/10.1002/esp.3251

McColl, S. T. (2012). Paraglacial rock-slope stability. *Geomorphology*, *153–154*, 1–16. https://doi.org/10.1016/j.geomorph.2012.02.015

Milodowski, D. T., Mudd, S. M., & Mitchard, E. T. A. (2015). Topographic roughness as a signature of the emergence of bedrock in eroding landscapes. *Earth Surface Dynamics*, *3*(4), 483–499. https://doi.org/10.5194/esurf-3-483-2015

Nelson, P. L., & Grand, S. P. (2018). Lower-mantle plume beneath the Yellowstone hotspot revealed by core waves. *Nature Geoscience*, *11*(4), 280–284. https://doi.org/10.1038/s41561-018-0075-y

NWCC Report Generator. (n.d.). Retrieved April 13, 2018, from https://wcc.sc.egov.usda.gov/reportGenerator/view/customWaterYearGroupByMonthReport/monthly/9905:WY:COOP%7Cid=%22%22%7Cname/POR_BEGIN,POR_END/PRCP::value?fitToScreen=false

Personius, S. F. (1982). *Geologic setting and geomorphic analysis of Quaternary fault scarps along the Deep Creek fault*.

Phanikumar, B.R. and Amrutha, K., 2014. Effect of overburden pressure and degree of saturation on compressibility characteristics. *Geomechanics and Geoengineering*, *9*(1), pp.52-62.

Pierce, K. L. (1979). History and dynamics of glaciation in the northern Yellowstone National Park area. *Professional Paper*, 90.

Pierce, K. L., & Morgan, L. A. (1992). Chapter 1: The track of the Yellowstone hot spot: Volcanism, faulting, and uplift. *Geological Society of America Memoirs* , *179*(2), 1–54. https://doi.org/10.1130/MEM179-p1

Pierce, K. L., Licciardi, J. M., Krause, T. R., & Whitlock, C. (2014). Glacial and Quaternary geology of the northern Yellowstone area, Montana and Wyoming. *Exploring the Northern Rocky Mountains*, *37*(9), 189–203. https://doi.org/10.1130/2014.0037(09)

Pitcher, J. (1903). *Report of the Acting Superintendent of the Yellowstone National Park to the Secretary of the Interior, for the Year 1903*.

Le Roux, O., Schwartz, S., Gamond, J. F., Jongmans, D., Bourles, D., Braucher, R., et al. (2009). CRE dating on the head scarp of a major landslide (Séchilienne, French Alps), age constraints on Holocene kinematics. *Earth and Planetary Science Letters*, *280*(1–4), 236–245. https://doi.org/10.1016/j.epsl.2009.01.034

Russo, E., Waite, G. P., & Tibaldi, A. (2017). Evaluation of the evolving stress field of the Yellowstone volcanic plateau, 1988 to 2010, from earthquake first-motion inversions. *Tectonophysics*, *700–701*, 80–91. https://doi.org/10.1016/j.tecto.2017.02.009

Sørensen, R., Zinko, U., & Seibert, J. (2006). On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences*, *10*(1), 101–112. https://doi.org/10.5194/hess-10-101-2006

Stoline, M. R. (2016). The Status of Multiple Comparisons : Simultaneous Estimation of All Pairwise Comparisons in One-Way ANOVA Designs Author (s): Michael R . Stoline Source : The American Statistician , Vol . 35 , No . 3 (Aug ., 1981), pp . 134-141 Published by : Taylor , *35*(3), 134–141.

Varnes, D. J. (1978). Slope movement types and processes. *Special Report - Transportation Research Board, National Research Council*, (176).

Waldrop, H. A., & Hyden, H. J. (1963). Landslides near Gardiner, Montana. *U.S. Geological Survey Professional Paper*.

Wallace, R. W. (1980). Degradation of the Hebgen Lake fault scarps of 1959. *Geology*, *8*(5), 225–229. https://doi.org/10.1130/0091-7613(1980)8<225:DOTHLF>2.0.CO

Wang, S., Qi, J., Yin, Z., Zhang, J., & Ma, W. (2014). A simple rheological element based creep model for frozen soils. *Cold Regions Science and Technology*, *106–107*, 47–54. https://doi.org/10.1016/J.COLDREGIONS.2014.06.007

Warkentin, B. P. (1961). Interpretation of the upper plastic limit of clays. *Nature*, *190*(4772), 287–288. https://doi.org/10.1038/190287a0

Weed, W. H. (1893). The glaciation of the Yellowstone Valley north of the Park. *U.S. Geological Survey Bulletin*, *104*(41), 4.

Whitlock, C., Fritz, S. C., Bracht, B., Stevens, L., Power, M., Whitlock, C., et al. (2006). The last 2000 years in Northern Yellowstone National Park Based on Multiproxy Data from Crevice Lake, (February).

Wong, I., Olig, S., Dober, M., Wright, D., Nemser, E., Lageson, D., et al. (2004). Earthquake Ground Shaking Hazard Maps for the state of Montana. *13th World Conference on Earthquake Engineering*, (1013).

APPENDICES

APPENDIX A

SURFICIAL MAPPING AND SAMPLING

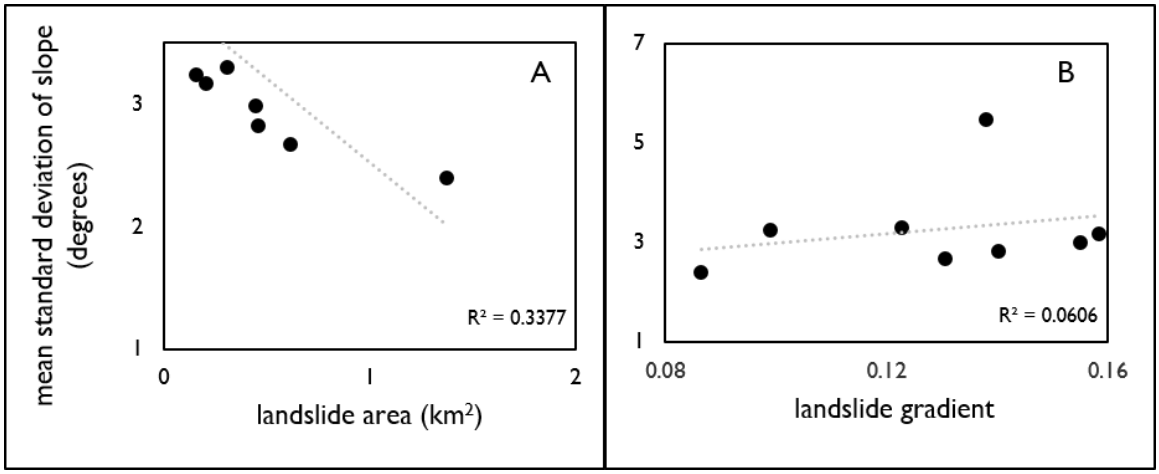

Figure A1. Mean standard deviation of slope (SDS) for each of 8 landslides versus (A) landslide area and (B) mean gradient.

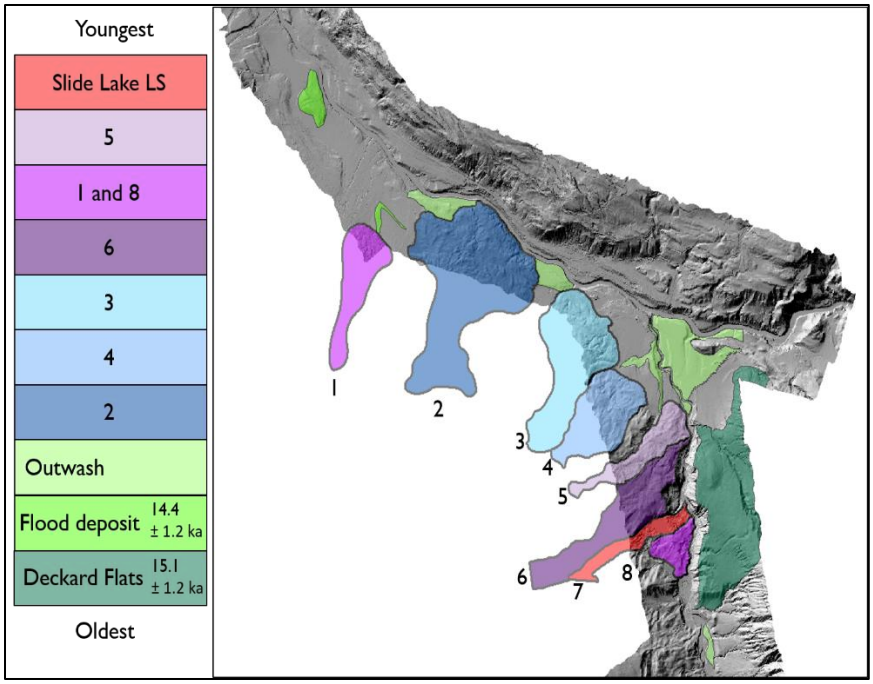

Figure A2. Stratigraphic relationships providing upper and lower age constraints for the Gardiner landslides. Landslides are ordered based on their roughness values. Ages for glacial features are from (Pierce et al, 2018). Mapped stratigraphic relationships provide upper age constraints for the Gardiner landslides.

APPENDIX B

SURFACE ROUGHNESS ANALYSIS

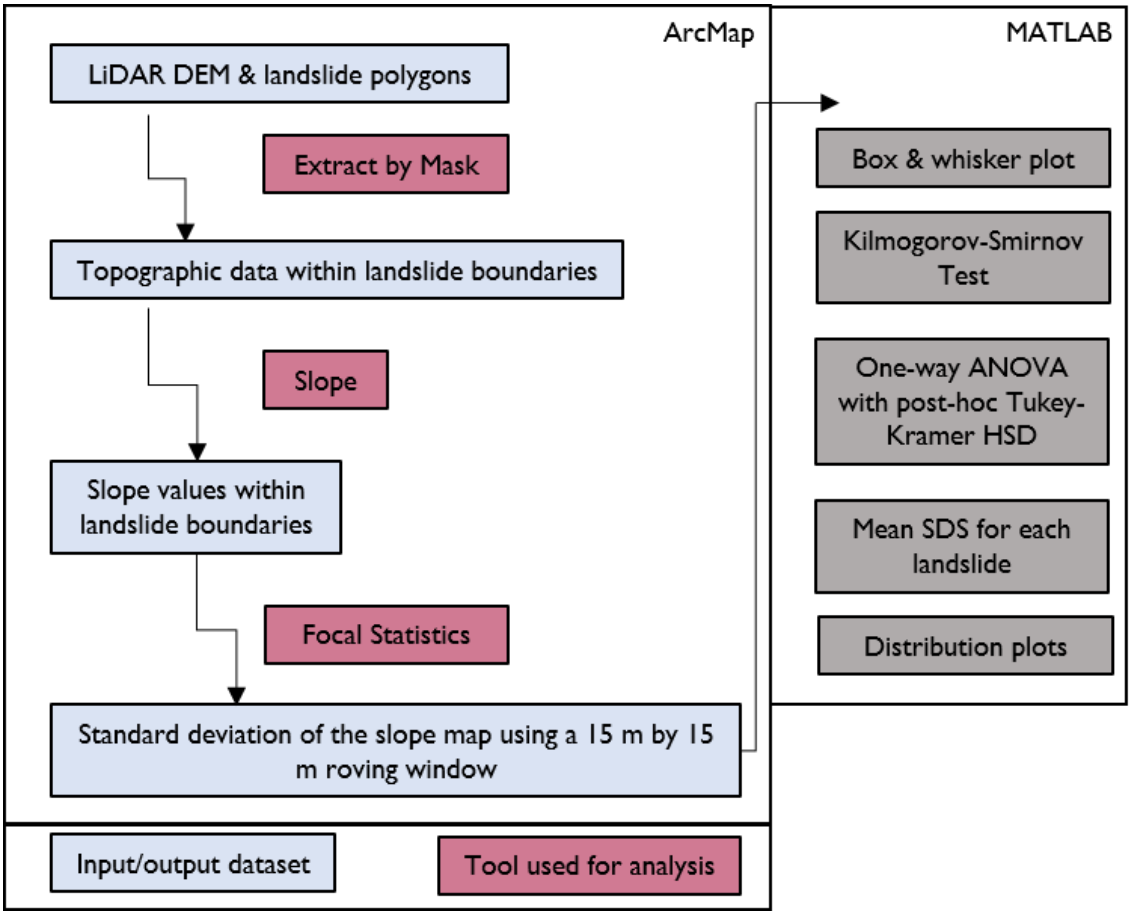

Figure 1B. Workflow for calculating standard deviation of slope for each landslides. Both ArcGIS and Matlab were used to process topographic data for surface roughness.

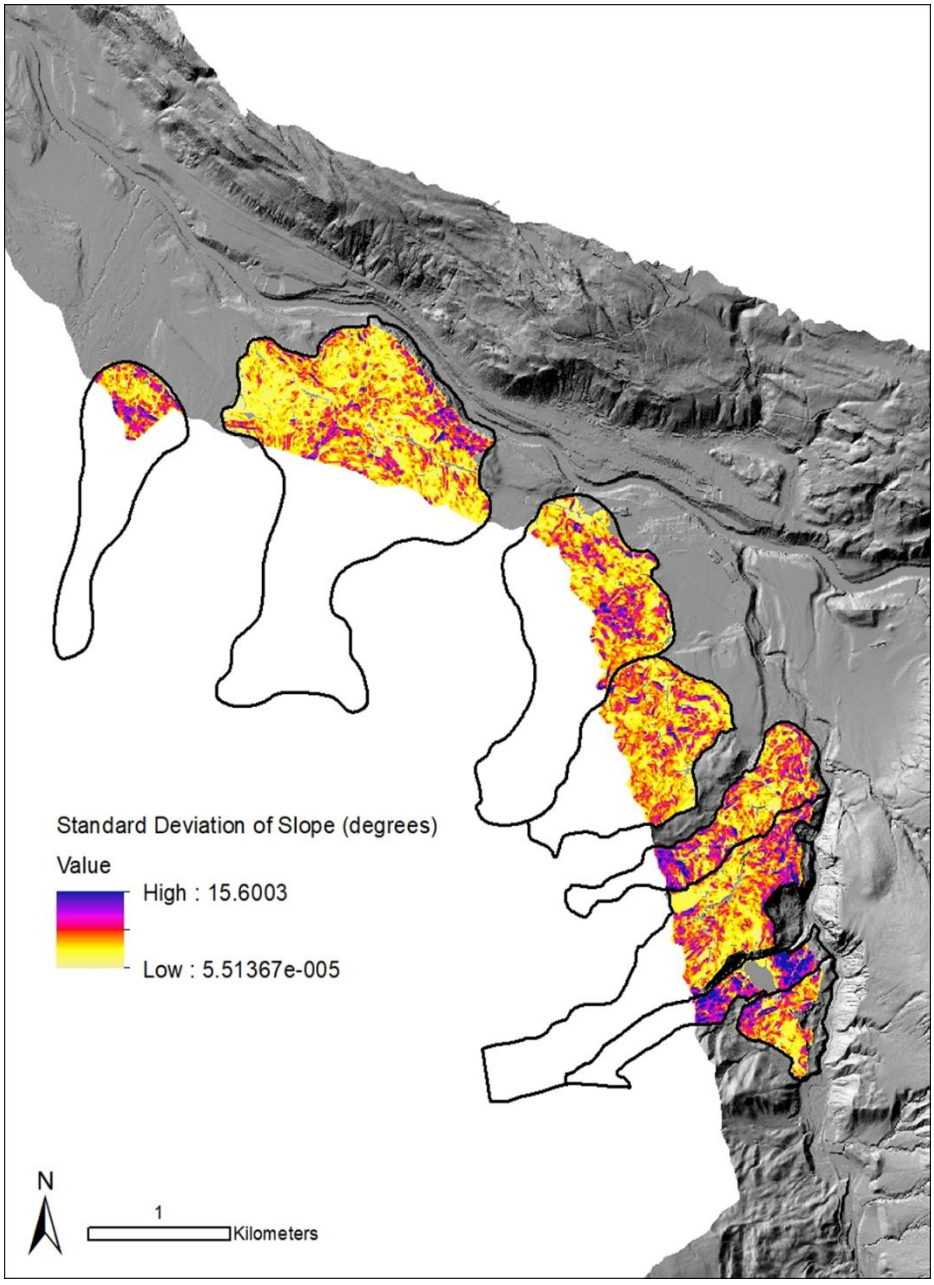

Figure 2B. Surface roughness map calculated using the available LiDAR that encompasses the toes of the Gardiner Landslides. This roughness map is the results after the removal of biases. Calculations were made with a 15 m by 15 m roving window.

Landslide Comparison	Significance Level α[a]	P Value[b]	D Statistic[c]	N n_1+n_2/n_1n_2[d]
LS1 versus LS2	0.01	0	0.1761	5.95424E-06
LS2 versus LS3	0.01	0	0.1140	2.18849E-06
LS3 versus LS4	0.01	0	0.1008	3.02998E-06
LS4 versus LS5	0.01	0	0.2150	3.41494E-06
LS5 versus LS6	0.01	3.50E-206	0.0281	3.32592E-06
LS6 versus LS7	0.01	0	0.3319	5.58826E-06
LS7 versus LS8	0.01	0	0.2101	7.38177E-06

Table 1B. Kolmogorov-Smirnov Test Results (pre-biases removal)
[a] This represents a confidence level of 99% (1-α)
[b] The P value is the probability that the null hypothesis (i.e. the likelihood that the two samples came from the same population) is true. If the P value is less than α, then the null hypothesis can be rejected.
[c] The D statistic is the maximum difference between the cumulative frequency plots of two populations
[d] This is the normalization coefficient for two population sizes, n_1 and n_2.

Landslide Comparison	Significance Level α[a]	P Value[b]	D Statistic[c]	N n_1+n_2/n_1n_2[d]
LS1 versus LS2	0.01	0	0.1761	5.95424E-06
LS2 versus LS3	0.01	0	0.1140	2.18849E-06
LS3 versus LS4	0.01	0	0.1008	3.02998E-06
LS4 versus LS5	0.01	0	0.2150	3.41494E-06
LS5 versus LS6	0.01	3.50E-206	0.0281	3.32592E-06
LS6 versus LS7	0.01	0	0.3319	5.58826E-06
LS7 versus LS8	0.01	0	0.2101	7.38177E-06

Table 2B. Kolmogorov-Smirnov Test Results (post-biases removal)
[a] This represents a confidence level of 99% (1-α)
[b] The P value is the probability that the null hypothesis (i.e. the likelihood that the two samples came from the same population) is true. If the P value is less than α, then the null hypothesis can be rejected.
[c] The D statistic is the maximum difference between the cumulative frequency plots of two populations

[d] This is the normalization coefficient for two population sizes, n_1 and n_2.

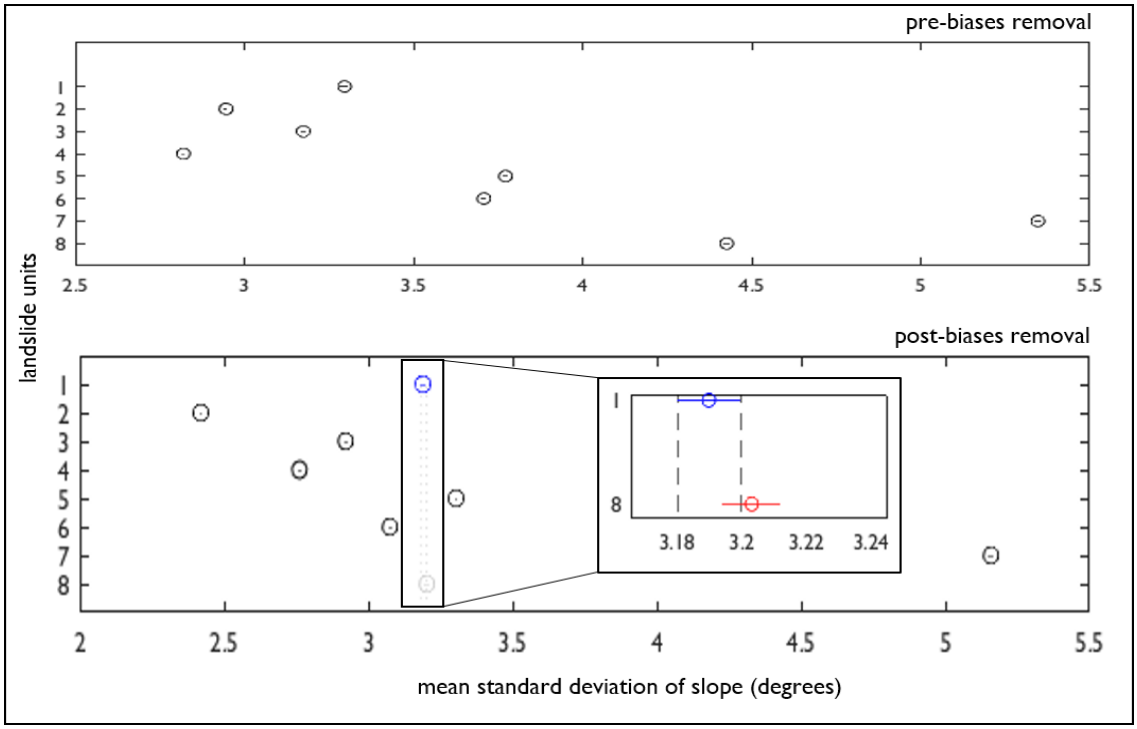

Figure 3B. Results of the Tukey-Kramer Honestly Significant Difference test. The top plot shows results before biases removal, none of the means intersect meaning all means are significantly different. The bottom plot (post-biases removal) shows an intersection between the means of landslide units 1 and 8, meaning their means are not significantly different.

One-way ANOVA – pre-biases removal

Source*	SS	df	MS	F	ρ-value
landslide ι	1945728.61	7	277961	54333.5	0
Error	24542292.8	4797318	5.1		
Total	26488021.5	4797325			

* source of variability

SS, sum of squares

df, degrees of freedom

MS, mean squares

F-statistic, ratio of the mean squares

Table 3B. ANOVA table captures the variability in the model by source, the F-statistic for testing the significance of the variability, and the ρ-value for deciding on the significance of this variability. Landslide row is the variability due to the differences among group means (variability between groups). Error is the variability due to the differences between the data in each group and the group mean (variability within groups). And Total is the total variability.

Tukey-Kramer HSD – pre-biases removal

landslide unit	mean	standard error	group
1	3.2960	0.0052	a
2	2.9459	0.0018	b
3	3.1742	0.0028	c
4	2.8199	0.0028	d
5	3.7722	0.0031	e
6	3.7080	0.0027	f
7	5.3470	0.0046	g
8	4.4261	0.0041	h

Table 4B. Groups to show the specific landslide means (compared with each other) that are different.

LS pair		lower confidence interval	estimate	upper convidence interval	ρ-value
1	2	0.3335	0.3502	0.3669	5.99E-08
1	3	0.1039	0.1218	0.1397	5.99E-08
1	4	0.4583	0.4762	0.4940	5.99E-08
1	5	-0.4945	-0.4761	-0.4577	5.99E-08
1	6	-0.4297	-0.4120	-0.3942	5.99E-08
1	7	-2.0720	-2.0509	-2.0298	5.99E-08
1	8	-1.1501	-1.1301	-1.1101	5.99E-08
2	3	-0.2385	-0.2284	-0.2182	5.99E-08
2	4	0.1159	0.1260	0.1361	5.99E-08
2	5	-0.8373	-0.8263	-0.8153	5.99E-08
2	6	-0.7721	-0.7622	-0.7523	5.99E-08
2	7	-2.4162	-2.4011	-2.3860	5.99E-08
2	8	-1.4938	-1.4803	-1.4668	5.99E-08
3	4	0.3424	0.3544	0.3663	5.99E-08
3	5	-0.6106	-0.5979	-0.5852	5.99E-08
3	6	-0.5456	-0.5338	-0.5220	5.99E-08
3	7	-2.1891	-2.1727	-2.1564	5.99E-08
3	8	-1.2668	-1.2519	-1.2370	5.99E-08
4	5	-0.9649	-0.9523	-0.9396	5.99E-08
4	6	-0.8999	-0.8881	-0.8764	5.99E-08
4	7	-2.5434	-2.5271	-2.5107	5.99E-08
4	8	-1.6212	-1.6063	-1.5914	5.99E-08
5	6	0.0516	0.0641	0.0766	5.99E-08
5	7	-1.5917	-1.5748	-1.5579	5.99E-08
5	8	-0.6695	-0.6540	-0.6385	5.99E-08
6	7	-1.6551	-1.6389	-1.6227	5.99E-08
6	8	-0.7329	-0.7181	-0.7034	5.99E-08
7	8	0.9022	0.9208	0.9394	5.99E-08

Table 5B. Results from the one-way ANOVA Tukey Kramer Honest Significance Difference test pre-biases removal. First two columns show which group means are compared with each other. The last column showas the ρ-values for the tests.

One-way ANOVA – post-biases removal

Source*	SS	df	MS	F	ρ-value
landslide units	1322957.4	7	188993.9	73614.42	0
Error	10698803	4167257	2.6		
Total	12021760	4167264			

* source of variability

SS, sum of squares

df, degrees of freedom

MS, mean squares

F-statistic, ratio of the mean squares

Table 6B. ANOVA table captures the variability in the model by source, the F-statistic for testing the significance of the variability, and the ρ-value for deciding on the significance of this variability. Landslide row is the variability due to the differences among group means (variability between groups). Error is the variability due to the differences between the data in each group and the group mean (variability within groups). And Total is the total variability.

Tukey-Kramer HSD – post-biases removal

landslide unit	mean	standard error	group
1	3.1897	0.0038	a
2	2.4196	0.0014	b
3	2.9217	0.0021	c
4	2.7617	0.0020	d
5	3.3043	0.0023	e
6	3.0751	0.0021	f
7	5.1583	0.0039	g
8	3.2030	0.0036	a

Table 7B. Groups to show the specific landslide means (compared with each other) that are different.

LS pair		lower confidence interval	estimate	upper convidence interval	ρ-value
1	2	0.7578	0.7701	0.7823	5.99E-08
1	3	0.2549	0.2680	0.2811	5.99E-08
1	4	0.4149	0.4280	0.4410	5.99E-08
1	5	-0.1281	-0.1146	-0.1011	5.99E-08
1	6	0.1014	0.1146	0.1278	5.99E-08
1	7	-1.9851	-1.9687	-1.9522	5.99E-08
1	8	-0.0292	-0.0133	0.0025	0.1733
2	3	-0.5096	-0.5020	-0.4945	5.99E-08
2	4	-0.3495	-0.3421	-0.3347	5.99E-08
2	5	-0.8929	-0.8847	-0.8764	5.99E-08
2	6	-0.6631	-0.6555	-0.6478	5.99E-08
2	7	-2.7512	-2.7387	-2.7263	5.99E-08
2	8	-0.7951	-0.7834	-0.7717	5.99E-08
3	4	0.1512	0.1599	0.1687	5.99E-08
3	5	-0.3921	-0.3826	-0.3731	5.99E-08
3	6	-0.1624	-0.1534	-0.1444	5.99E-08
3	7	-2.2500	-2.2367	-2.2234	5.99E-08
3	8	-0.2939	-0.2814	-0.2688	5.99E-08
4	5	-0.5519	-0.5426	-0.5332	5.99E-08
4	6	-0.3222	-0.3134	-0.3045	5.99E-08
4	7	-2.4098	-2.3966	-2.3834	5.99E-08
4	8	-0.4538	-0.4413	-0.4288	5.99E-08
5	6	0.2196	0.2292	0.2388	5.99E-08
5	7	-1.8678	-1.8540	-1.8403	5.99E-08
5	8	0.0883	0.1013	0.1143	5.99E-08
6	7	-2.0966	-2.0832	-2.0699	5.99E-08
6	8	-0.1406	-0.1279	-0.1153	5.99E-08
7	8	1.9393	1.9553	1.9713	5.99E-08

Table 8B. Table 5B. Results from the one-way ANOVA Tukey Kramer Honest Significance Difference test pre-biases removal. First two columns show which group means are compared with each other. The last column showas the ρ-values for the tests.

Roughness Biases	Landslide unit	Lidar coverage landslide area (m^2)	biases area (m^2)	% area removed	Total % area removed*
Slide Lake	7	216326	3512	2	0.08
Gullies	1	172986	33865	20	0.74
	2	1459007	259103	18	5.69
	3	621624	114504	18	2.52
	4	634803	119322	19	2.62
	5	492009	82864	17	1.82
	6	665284	112263	17	2.47
	7	216326	39616	18	0.87
	8	290413	24909	9	0.55
FIAT	2	1459007	84592	6	1.86
	5	492009	25777	5	0.57
	6	665284	50051	8	1.10
	7	216326	19455	9	0.43
	8	290413	76936	26	1.69
Infrastructure	2	1459007	8099	1	0.18
	3	621624	29062	5	0.64
Roads	2	1459007	38104	3	0.84
	3	621624	12565	2	0.28
	5	492009	16438	3	0.36
	6	665284	33854	5	0.74
	7	216326	1378	1	0.03

*percent area of biases in the total LiDAR coverage,4,552,452 m^2

Table 9B. Biases categories removed from data to show the percentage of available LiDAR that was removed/affected by post-depositional modification.

Figure 10B. Archived photo taken from Open Parks Network of a reactivation that occurred in 1959 on the toe of landslide 5.

Figure 11B. Same area of photo 10B showing a smoothed surface where the reactivation occurred in landslide five that was modified by the park service. This photo was taken in 2013 and was downloaded from Google Earth.

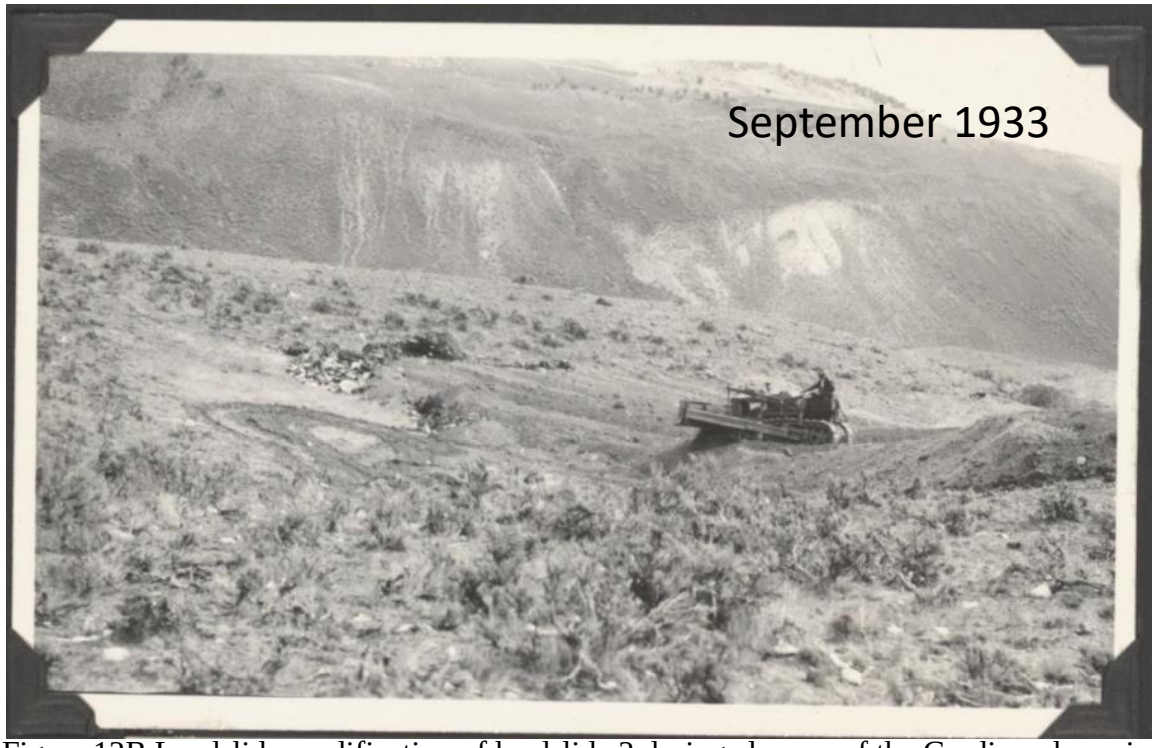

Figure 12B Landslide modification of landslide 2 during cleanup of the Gardiner dump in 1933. Picture downloaded from Open Parks Network.