UNSTABLE SLOPE CRITERIA PROJECT:

OBJECT-BASED LANDFORM MAPPING WITH HIGH RESOLUTION TOPOGRAPHY

Draft Final Report

~~April~~September 2025

Authors:

Elise Freeman (NWIFC), Dan Miller (M2 Environmental Services), Julie Dieu (Rayonier), Tiffany Justice (Weyerhaeuser), Jeff Keck (DNR), and Ted Turner (Weyerhaeuser)

Commented [HB1]: All my comments posted on 250524 should be considered yellow. Most are clarification suggestions. Generally, well written document.

Commented [HJ2]: RED – The apparent purpose of this report is to document the first in a series of sub-projects within the broader Unstable Slopes Criteria Project. There appears to be confusion about what was supposed to be done, what has been done, and what will be done in this sub-project and subsequent sub-projects. This theme recurs throughout the document. For a good example, one does not have to look any further than the Executive Summary. There appear to be five iterations of statements and/or outlines describing the sub-project's objectives. None of ...

Commented [HJ3]: RED – The document is unnecessarily long, probably two to three times as long as it needs to be. Again, the Executive Summary provides an example of this. There is a lot of material included in this report that is ...

Commented [HJ4]: Yellow – It appears that the project team learned a lot about methods for evaluating the models from Phase 1. The team tried to apply those findings to study area 1 b. It is unfortunate that the four main study ...

Commented [HJ5]: RED – The terms "landform", "high hazard landform", "rule-identified landform" and "RIL", and "inner gorge" and "bedrock hollow" are all used interchangeably throughout the report. Correct and ...

Commented [HJ6]: YELLOW – I'd like to provide some context on how the current process in forest practices works in response to some of the commentary in the report. Qualified experts use maps to screen operating areas for ...

Commented [HJ7]: RED – It does not appear that maps of contiguous landforms have been generated. Contiguous would generally imply that the entire landscape for a study area has been classified into landforms, not just bedrock ...

Commented [HJ8]: YELLOW – It is unclear how a stream and/or channel head network was generated and/or provided for the different mapping exercises. It is also unclear how and why this process was modified along the ...

Commented [HJ9]: YELLOW - The exercise of generating a "truth" map to compare the models against would greatly benefit from a field-based verification component.

Commented [HJ10]: YELLOW – Some of the numbers in the tables, for example Tables 4, 7, and 8, don't appear to add up. These tables should be double-checked. If the numbers are correct, then the tables need to be better ...

Commented [HJ11]: YELLOW – With regards to the appendices, Appendix A and Appendix B are useful. Appendix C should be removed. Appendix D through G should stick to presenting factual data. Any pertinent ...

Commented [HJ12]: YELLOW – For the reasons stated above, I recommend rewriting this report under the direction of a technical writer/editor and then resubmitting it to CMER for review. As it currently reads, I have low ...

EXECUTIVE SUMMARY

Forest practices in Washington State seek to ensure both a healthy ecosystem and a viable forest industry. Guidelines for doing so were established with the Forests & Fish Report in 1999, enumerated in the state Forest Practices Habitat Conservation Plan (HCP) in 2006, and specified as Rules in the Washington Administrative Code (WAC). Timber lands in Washington include extensive areas in mountainous terrain where landsliding is an integral part of the natural disturbance regime. Landslide activity can have a positive influence (e.g., introduction of coarse sediment and ~~introduction,~~ large woody debris), but can also adversely impact ecosystem functions and increase threat to public safety. It is well recognized that forest practices can ~~influence~~elevate landslide rates, so a key goal of the HCP is to ~~keep forest practices from altering~~minimize those negative effects. ~~those rates to an extent that would adversely impact ecosystem functions. Another important goal is to ensure that forest practices do not increase the threat to public safety posed by landslides~~.

To meet these goals, the WAC ~~specifies certain areas of the landscape~~defines landforms ~~where landslides are common~~ that are likely more susceptible to landslides. In these areas, timber harvest and associated activities are ~~precluded~~prohibited without review by ~~geotechnical specialists~~Qualified Experts and approval by Washington Department of Natural Resources. ~~The current paradigm~~method for identifying Llandslide-prone areas that are potentially sensitive to forest-practice activities, known as Rule-Identified Landforms (RILs), ~~is based on~~are identified ~~through~~by ~~the~~ narrative descriptions of ~~specific~~ landforms . ~~Landforms serve as the primary framework for both map screening based on remotely sensed data and the final field detection and interpretation of potential landslide hazards and risks. The regulated subset of landforms, known as Rule-Identified Landforms (RILs). , and c~~Criteria for their identification include specific slope and geomorphic characteristics which were established through extensive surveys and analyses by the Washington Watershed Analysis Program and the Landslide Hazard Zonation Project (LHZ). Practitioners interpret RILs by conducting thorough office reviews using a variety of maps and remotely sensed data prior to verifying the landforms in the field. ~~Both map screening based on remotely sensed data and field detection and interpretation play important roles in the identification and protection of the RILs.~~

The Unstable Slopes Criteria (USC) Project is a suite of sub-projects that were developed to evaluate and suggest potential modification to current RIL definitions, incorporating new data and analysis methods, in order to answer the following critical question (UPSAG 2017):

What modifications to the unstable slopes criteria and delivery-assessment methods would result in more accurate and consistent identification of:

i. unstable slopes and landforms,

Commented [GS13]: I recommend a large reduction in word count for the executive summary. No need to go into the full background of the USC project here. Quickly get into the summary of the current report. Structure it like a condensed version of a scientific paper: intro, methods, results, discussion, conclusions. Each section should be no more than a paragraph or so.

Commented [PJ14R13]: I agree, this summary should be substantially shortened and focused on this sub project.

Commented [TH15]: The first two paragraphs of the Exec. Sum. Have been edited with track changes. The remainder of the Exec Summary is replaced in its entirety with new text in red

Commented [PJ16]: It doesn't specify areas of the landscape, it defines landforms (and landowners are required to find them on the landscape)

Commented [PJ17]: This belongs right after the first sentence.

ii. unstable slopes and landforms sensitive to forest-practices-related changes in landslide process, and
iii. locations susceptible to impacts from upslope landslides such that an adverse impact to public resources or a threat to public safety is possible?

This document describes the Object-based Landform Mapping sub-project, which has the objectives of 1) identifying methods for consistent automated delineation of landforms using computer-based techniques and high-resolution LiDAR DEMs, and potentially other data sources, and 2) the automated landform model provides the baseline geomorphic context from which to evaluate landslide susceptibility and runout, and incorporates data from process-based models to train the automated classification of landforms.

Manually mapping RILs remotely and~~/or~~ in the field is both a subjective and time-intensive process that is difficult to consistently implement over large areas or iteratively adjust based on modified landform criteria. Therefore, this project explores automated mapping methods that replicate key features that can be identified using lidar ~~of manually mapped RILs~~ and can be repeatedly and consistently implemented over large areas. Here we report on the first in the sequence of sub-projects: Object-Based Landform Mapping with High-Resolution Topography. The goal of this sub-project is to develop automated methods to identify ~~the current set of shallow rapid RIL~~bedrock hollows and inner gorges (i.e., WAC 222-216-050(1)(d)(i)(A)). This foundational algorithm will be expanded on in subsequent sub-projects. This project did not directly attempt to identify other RILs such as convergent headwalls, toes of deep-seated landslides steeper than 33 degrees (65%), groundwater recharge areas for deep-seated landslides, outer edges of meander bends, or other areas of potential slope instability (WAC 222-216-050(1)(d)(i)), but some of these are captured by the model efforts.

Commented [PJ18]: RED: The original phrasing implies that the automated mapping was compared with complete manual mapping. It needs to be very clear throughout this report that the manual mapping was never complete because there was no field verification. There is nothing in the WAC or the board manual that implies that any of this mapping can be completed without field verification.

We developed two automated approaches, both of which use DTMs as the primary input data: Object-Based Image Analysis (OBIA) and a Virtual Watershed (VW) terrain analysis. These methods were developed by two of the Project Team members based on input provided by the entire Project Team. Both methods use similar topographic inputs from the DTM, though they differ in approach. OBIA applies image-analysis techniques to detect spatial patterns and classifies them into landform types based on defined rules. In contrast, the VW method begins with landform definitions and identifies zones that meet those definitions by tracing surface flow paths and spatial connectivity. Differences in outcomes between the methods can reveal ambiguities or refinements needed in model inputs.

To evaluate the adequacy of these two methods, the Project Team conducted several exercises over ~~2~~two phases. Phase 1 focused on spatial comparisons of computer-based map products and remotely-drawn map products ~~drawn by our~~created by the project team of 4-5 experienced

practitioners (EP). Phase 2 was a qualitative evaluation of computer-based map products by the practitioners. Choosing to use multiple experienced practitioners in Phase 1 required an additional analysis step to quantify the consistency of map products drawn by the practitioners prior to comparing them to the computer-based maps products. An acceptable consistency target was not achieved; therefore the products were not used in further model validation efforts. However, key findings from the exercise are reported.

Phase 1 revealed ~~unresolvable~~ discrepancies in how the experienced practitioners delineate bedrock hollows and inner gorges using remote data. Quantitative comparisons using confusion matrices and summary metrics such as balanced accuracy, recall, and precision confirmed these discrepancies. While mappers generally agreed on the locations where ~~these landforms~~bedrock hollows and inner gorges did not exist, their interpretations of where potential RILs were present and how to draw the discrete boundary were inconsistent. Spatial overlap between mapper's data was ~~minimal~~unsatisfactory, and efforts to increase the overlap did not produce useful results.

Commented [HJ19]: Context is needed if "minimal" is going to be retained in this sentence.

These findings underscored the need for a more objective and reproducible approach to evaluating model performance. We initiated Phase 2 to address this need by removing manual reference maps and instead assessing model output directly through experienced practitioner review of computer-generated polygons. This shift allowed the Project Team to focus on the relative strengths of each model, reduce bias introduced by individual mapping styles, and better identify opportunities for improving automated delineation of RIL landforms.

The Phase 2 evaluation focused on each model's ability to (1) reasonably screen for the presence of target landforms and (2) accurately delineate their boundaries based on lidar interpretation. Results indicated that both models generally performed well in identifying landform presence. Across all sites, expert mappers judged that 70% or more of the modeled landforms were reasonable screens (RS) for either bedrock hollows or inner gorges. However, delineation accuracy showed reduced model performance and greater variability, with a marked difference between models and landform types. Out of the bedrock hollow polygons designated reasonable screens, tThirty-nine percent of bedrock hollow polygons and 62% of inner gorge polygons were designated to be correctly drawn (CD) based on lidar interpretation. Inner gorge polygons consistently received higher CD ratings than bedrock hollows, and VW model polygons tended to perform slightly better than the OBIA polygons on average for both RS and CD criteria.

Commented [PJ20]: RED: Throughout, it needs to be clear the "correctly drawn" is still a screen. They have not been verified in the field.

Commented [HJ21]: Ensure that the "39%" and "62%" values listed are number of polygons correctly drawn out of total number of polygons, not out of number of reasonable screen polygons.

The variability in correctly drawn ratings highlights areas for model refinement, especially for bedrock hollow delineations. EP comments and site-specific patterns suggest that some modeled polygons failed to capture the necessary geomorphic boundaries or sufficient topographic convergence expected by practitioners. These outcomes suggest targeted improvements to input parameterization, particularly those controlling convergence and planform curvature thresholds,

are needed to improve the models' abilities to accurately delineate bedrock hollows and inner gorges.

Through the development and evaluation of the OBIA and VW models, we demonstrated viable foundational approaches for using lidar to delineat~~ing~~ inner gorges and bedrock hollows based on terrain attributes and rule-based classification. The models successfully generated interpretable and repeatable landform maps, satisfying the first project objective. Based on the results of Phase 2, the automated methods developed in this report delineate landforms in a way that adequately replicates manually mapped landforms for use as a screen but not for use as definitive landform boundaries. Section 1.3 describes the ways this project did and did not meet the objectives from the study design. These automated methods will serve as a basis for model development in ~~–~~the next sub-project. The results suggest that, with clear definitions and calibrated rulesets, the automated methods offer scalable and consistent methods for lidar-based landform delineation across diverse landscapes.

The Susceptibility & Runout sub-project, the next step, will focus on integrating landslide inventory data to evaluate and refine the automated landform models for bedrock hollows and inner gorges and for other landform types where landslides are found. That sub-project will assess OBIA and VW model outputs based on how well they align with observed landslide distributions. Specifically, landslide density and proportion will be calculated for each landform type produced by the models. Ranking landforms in terms of these metrics can identify which delineations are most strongly associated with landslide occurrence.

This approach will allow for:

- Statistical comparison between model versions;
- Evaluation of model sensitivity to rule parameters;
- Refinement of landform rules based on landslide data; and
- Identification of landforms with consistently high or low landslide susceptibility.

Ultimately, these steps aim to develop robust, data-driven criteria for ~~defining~~evaluating landforms used in landslide hazard assessments and forest practice regulations.

Commented [PJ22]: Was this done?

Commented [HJ23]: This is a good lead-in statement and good spot to include an abbreviated summary of the Objective Rewrite explaining that Objective #2 was not met and the project approach will need to be modified. The remainder of the Executive Summary will likely need to be modified to accommodate the Objective Rewrite summary.

Commented [PJ24]: Does the next sub-project use the same locations? Or does it expand the study areas to more diverse locations? The study sites will need to include much more diverse landscapes to refine landform rules, including areas in Eastern WA.

Commented [PJ25]: Which steps? The sub-projects? Are there any sub-projects after Susceptibility and Runout?

Commented [PJ26]: The definitions may not need to change

Table of Contents

Table of Figures

List of Tables

1.0 INTRODUCTION

The Unstable Slopes Criteria (USC) Project will provide information with which to evaluate the current Rule-Identified Landform (RIL) criteria. The project involves a sequence of sub-projects. ~~This report addresses the first in the sequence~~We address the first sub-project in this report, Object-Based Mapping with High-Resolution Topography (OBM).

1.1 THE UNSTABLE SLOPE CRITERIA PROJECT HISTORY

For the OBM sub-project we developed and evaluated new methods for automatically mapping potential unstable slopes defined by the Washington Forest Practice rules (described below). Before detailing this sub-project and results, some context is necessary to understand the need for this sub-project, its objectives, and the tasks undertaken. ~~Detailed background information for the USC Project can be found in~~ T~~t~~he USC Research Alternatives (USC TWIG 2017) and OBM study design (Dieu et al. 2018) documents provide detailed background information for the USC project, ~~and~~ summarized here.

The over-arching goals for forest management are specified in the state Forest Practices Habitat Conservation Plan (HCP) (WADNR 2006) as first enumerated in the Forests and Fish Report (USFWS 1999):

- Provide compliance with the Endangered Species Act for aquatic and riparian dependent species;
- Restore and maintain riparian habitat to support a harvestable supply of fish;
- Meet the requirements of the Clean Water Act for water quality; and
- Keep the timber industry economically viable in the state of Washington.

Within the Forests & Fish R~~r~~eport, Schedule L-1 sets performance targets for a variety of measures to meet resource objectives described therein. The performance target for the measure of mass wasting sediment delivered to streams is "*no increase over natural background rates from harvest on a landscape scale on high-risk sites*."

~~Forest practice rules in~~ Washington's Forest Practice rules ~~-~~specify the following requirements:

> WAC-222-16-050. Class-IV Special Forest Practice: "*…application to conduct forest practices involving the following circumstances requires an environmental checklist in compliance with the State Environmental Policy Act (SEPA), and SEPA guidelines, as they have been determined to have potential for a substantial impact on the environment… (d) Timber harvest, or construction of roads, landings, gravel pits, rock quarries, or spoil disposal areas, on potential unstable slopes or landforms described in (d)(i) of this subsection that have the potential to deliver sediment or debris to a public resource or that have the potential to threaten public safety...*"

Commented [PJ27]: Updating the text to have more active voice and concise sentences will help the readability. Consider subdividing sentences that are three or more lines into multiple sentences.

Commented [TH28]: As indicated in the matrix, the outline of the document has been changed, so whole sections were moved around...this was not done in track changes to help readability. Consult outline docs in folder to see changes.

Commented [TH29]: Figures have been replaced incorporating requested changes...not done in track changes to help with readability

Commented [HB30]: Since almost all FFR lands have been cut twice, how do you determine natural background rates let alone the effect of climate change changes in precipitation patterns?

To meet these objectives, sites need to be identified where:

1. landslides have the "potential for a substantial impact on the environment" via delivery of "sediment or debris to a public resource or that have the potential to threaten public safety;" and
2. forest practices can increase the potential for those impacts or threats to public safety.

The area precluded from timber production should be based on best available science, in accordance with the state's HCP goal to "keep the timber industry economically viable in the State of Washington." Therefore, potentially unstable slope areas need to be accurately identified and precisely delineated to avoid unnecessarily protecting areas from harvest that do not pose risks of substantial resource impacts or threats to public safety. Although landslides mapped within RILs often deliver material to stream channels (e.g., Dieu and Toth, 2009), not all do and some landslide-prone areas may be high hazard but not necessarily high-risk. To determine areas with a high risk, we must consider the likelihood of slope failure *and* the potential for delivery to streams or threats to public safety and resources.

The Forest Practice Rules in Washington, via WAC-222-16-050, identify the following five unstable slopes and landforms with associated geomorphic criteria that may require forest practices prescriptions if deemed high risk upon further assessment. These are referred to as Rule-Identified Landforms (RILs):

Table 1: Rule-Identified Landform (RIL) features that meet the specified criteria for unstable slopes and landforms via WAC-222-16-050.

(A)	Inner gorges, convergent headwalls, or bedrock hollows with slopes steeper than thirty-five degrees (seventy percent)
(B)	Toes of deep-seated landslides, with slopes steeper than thirty-three degrees (sixty-five percent)
(C)	Groundwater recharge areas for glacial deep-seated landslides
(D)	Outer edges of meander bends along valley walls or high terraces of an unconfined meandering stream
(E)	Any areas containing features indicating the presence of potential slope instability which cumulatively indicate the presence of unstable slopes.

~~While current RILs represent a subset of landslide prone landforms, there are additional landforms in which landslides occur but may not meet the criteria of an RIL.~~ RILs, as defined by Section 16 of the Forest Practices Board Manual, are those landforms that can be readily detected in the field by practitioners, have high relative hazard and risk when compared to other landform classes, and are inferred to be particularly sensitive to applied forest practices.

The Forest Practices Board Manual~~,~~ Section 16 (hereafter called out as Board Manual~~BM16~~)~~,~~ provides detailed descriptions of these features using quantitative slope and slope-break

Commented [AW31]: Yellow. This is a true statement. Please and clarify how to address this issue in the context of the goals of this project.

Commented [HJ32]: Check that the terms "Board Manual", "Board Manual Section 16", and "BM 16" are used appropriately and consistently from here on.

thresholds, qualitative convergence and curvature characteristics, and a geomorphic description based on topographic and field evidence of fluvial or landslide erosional processes. Of the listed landforms, bedrock hollows and inner gorges lend themselves to automated modeling techniques more readily available than other landforms that require geomorphic evaluation and field identification, such as deep-seated landslides or Category E landforms (a catch-all designation for features with clear field indication of instability that do not meet the other definitions – see Board Manual). Additionally, ~~bedrock hollow~~bedrock hollows and ~~inner gorge~~inner gorges are the most common RILs, with one potentially occurring more frequently than the other in specific watersheds. Their quantitative definitions often capture terrain of the other landforms (e.g., outside meander bends are really a type of inner gorge, and a convergent headwall includes a concentration of bedrock hollows), or they rely more strongly on subjective geomorphic interpretation (e.g., toes of deep-seated landslides eroded by streams may be susceptible to either shallow or deep-seated landslides and while the Board Manual cutoff is 65%, they often exceed 70% and ~~may~~ also ~~have~~meet the inner gorge landform~~s associated with them~~ definition). In summary, we have an established set of landform criteria and runout guidance for inner gorges and bedrock hollows because they are very prevalent on the landscape and mapping/modeling from their criteria covers other named RILs. As specifi~~ci~~ed in the Study Design, Deep-seated landslides and Category E features are not covered here.

Since the RIL definitions were originally derived, new data sources and analysis tools have become available. The Unstable Slopes Criteria Project uses these new data sources and tools to address~~es~~ the critical question:

What modifications to the unstable slopes criteria and delivery-assessment methods would result in more accurate and consistent identification of:

i. *unstable slopes and landforms,*
ii. *unstable slopes and landforms sensitive to forest-practices-related changes in landslide process, and*
iii. ~~*unstable slope and landform conditions where landslide runout would likely have an*~~ *locations susceptible to impacts from upslope landslides such that an adverse impact to public resources or a threat to public safety is possible?*

~~The~~ Ultimately, the objective of ~~the~~the USC ~~is sub~~ project is to evaluate ~~unstable slopes~~RIL criteria, ~~-~~and recommend modifications ~~so that unstable slopes with the potential to deliver sediment or debris to a public resource or that has the potential to threaten public safety can be identified more~~to improve RIL mapping accura~~cy~~tely and consisten~~tly~~cy. Specifically, the intent of this project is to recommend modifications to the criteria that increase the landslide "rate" of the RIL.

The concept of landslide "rate" stems from the Watershed Analysis Program (Washington Forest Practices Board, 2011) and the Landslide Hazard Zonation (LHZ) Project (UPSAG, 2006), which

Commented [HB33]: ??

Commented [HJ34]: Yellow - Was the decision to focus exclusively on bedrock hollows and inner gorges endorsed by CMER and/or the ISPR? If so, that should be clearly stated.

Commented [GS35]: Redundant

Commented [KJ36R35]: corrected

Commented [GS37]: See my earlier comment. This is not the same as in the study design and is nonsensical.

Commented [JM38]: Please explain. More accurate than what. Is there a baseline.

Commented [KJ39R38]: edited to reflect that more accurate and consistent is relative to the current RIL criteria

provided the basis for identifying specific features as landforms prone to landsliding. Landslide rate is simply the landslide density per unit time. Landslide density is defined either as the number of landslides per unit area or the landslide surface area per unit area, with the two versions of landslide density being differentiated by their units. The LHZ Project used landslide rate[1] to define susceptibility.

In 2017, the Forest Practices Board's Cooperative Monitoring, Evaluation, and Research Committee (CMER) approved a "Best Available Science and Research Alternatives" document. In agreement with the authors, the Timber, Fish, and Wildlife Policy committee (TFW Policy) chose the research alternative that outlined a sequence of five research projects to address these questions. The five projects of the selected alternatives were:

1. Compare/Contrast Landslide Hazard Zonation (LHZ) Mass Wasting Map Units with RILs;
2. Object-Based Landform Mapping with High-Resolution Topography;
3. Empirical Evaluation of Shallow Landslide Susceptibility and Frequency by Landform;
4. Empirical Evaluation of Shallow Landslide Runout; and
5. Models to Identify Landscapes/Landslides Most Susceptible to Management.

Since that time, there has been some evolution of the sequence. Comments received during the Independent Scientific Peer Review (ISPR) of the study design for this project (originally Project 2) encouraged the Project Team to fold Project 1 into the other projects. While designing Projects 3 & 4, the sensibility of combining the two projects became apparent because characterizing landslide initiation sites and quantifying runout both require identification of landslides. Within the Unstable Slopes Criteria Project, we therefore now identify three sub-projects that incorporate the objectives of the original five projects:

1. Object-Based Landform Mapping with High-Resolution Topography (this sub-project).
2. Empirical Evaluation of Shallow Landslide Susceptibility, Frequency and Runout by Landform (originally Projects 3 and 4; "Susceptibility and Runout" henceforth).
3. Physical Models to Identify Landforms and Shallow Landslides Most Susceptible to Management (originally Project 5).

This document reports the results of the first sub-project – Object-Based Landform Mapping with High-Resolution Topography (OBM) – which details automated methods for mapping the ~~bedrock hollows and inner gorges~~RILs specified in the study design as those described in WAC 222-16-050(1)(d)(i)(A)) (Category A, Table 1) that will be used in later sub-projects. In the next

[1] ~~Note that~~Again, "rate" is simply "density" per unit time, where "density" is defined as the number of landslides or surface area of landslides per unit basin area. Landslide density calculated from a landslide inventory collected over time can provide an estimate of rate; inventories based on landslides from single storm events can only provide measures of density.

sections, we provide background information necessary to understand the objectives and tasks of the first project.

~~*The Unstable Slopes Criteria (USC) Project is focused on landscape features associated with shallow-rapid landslides, which can occur in the RILs described in (A), (B), (D), and potentially (E) in Table 1. Shallow-rapid landslides result from the sudden failure of relatively thin layers of soil, up to several meters deep. Shallow-rapid landslides within inner gorges, along over-steepened slopes above meander bends, and at the toes of deep-seated landslides adjacent to channels can deliver material directly to streams, which are regarded as an important public resource. Shallow-rapid landslides originating in bedrock hollows or bedrock hollows within convergent headwalls may evolve into debris flows. Debris flows can travel long distances downslope when confined to a steep, narrow channel and, thus, pose risks to streams and public safety potentially far removed from the initiation site.*~~

~~*Common landslide locations recorded during Washington's Watershed Analysis Program (Washington Forest Practices Board, 2011) and by the Landslide Hazard Zonation (LHZ) Project (UPSAG, 2006) provided the basis for identifying specific features as landforms prone to landsliding. The LHZ Project focused on variations in landslide rate across different landforms to determine susceptibility. Rate was measured in two ways: by the number of landslides per unit area per unit time and by the surface area of landslides per unit area per unit time. Relative (Table 1) were found to have high rates of shallow rapid landslides that delivered material to stream channels (Dieu and Toth, 2009). A key task in evaluating the effectiveness of current RIL definitions in identifying potentially unstable slope areas with high risk is to calculate delivering landslide rates for RILs and other landforms over a broad range of terrain types and geographic locations to see if that observation continues to hold true with new data. This is the focus of the next sub-project in the USC Project: Empirical Evaluation of Shallow Landslide Initiation and Runout Susceptibility and Frequency by Landform. More information regarding project history and development can be found in the Study Design.*~~

1.2 Background: Landform Mapping and OBIA

Landform mapping includes~~constitutes the process of~~ detecting, delineating, classifying, and recording land surface features in visual and/or tabular data layers that can be used in a variety of scientific and technical applications. Landform mapping traditionally has employed a suite of d~~ifferent techniques ranging from~~ manual ~~to~~and computational ~~in nature~~techniques. Manual methods typically yield hand-drawn or digitized, qualitative interpretations of landforms based on topographic, thematic (e.g., geology, soils, land use), contextual, and aerial imagery information. Computational techniques involve mathematical derivations of land-surface features based on their geomorphometric (i.e., quantitative geomorphological) characteristics. Computer models typically apply high-resolution~~, lidar-derived,~~ DTM products to delineate and classify landforms based on geomorphometric variables like surface elevation, gradient, curvature, and roughness or texture (e.g., Dikau, 1989; Dragut and Blaschke, 2006; Wood, 2009; Jenness, 2013). More advanced computational methods also incorporate different combinations

Commented [PJ40]: This contradicts the previous paragraph that says it is focused on BHs and IGs. I recommend deleting this sentence or combining it into the paragraph above.

Commented [KJ41R40]: Adjusted the previous paragraph to clarify project is focused on RILs, not just BHs and IGs

Commented [JM42]: Revise to say, potentially degrading an

Commented [KJ43R42]: deleted

Commented [HJ44]: Yellow - Shallow-rapid landslides associated with inner gorges and deep-seated landslides may also generate debris flows.

Commented [KJ45R44]: deleted

Commented [PJ46]: Is this paragraph needed here? This is in the "history" section, and there is already a description of all the RILs and a statement that this project is focused on BHs and IGs. If we want to keep this information about shallow landslide processes, it would fit better earlier in the section, within the RIL descriptions.

Commented [MM47]: Is this referring to Table 1 in the reference? I don't see a related table in this document.

Commented [KJ48R47]: Deleted Table 1 reference

Commented [GS49]: Clarify.

Commented [KJ50R49]: Corrected, "Relative" should have been "RILs"

Commented [HB51]: Something missing here

Commented [KJ52R51]: Yes, "Relative" should have been "RILs"

Commented [MM53]: This is a challenge because there is context for delivery that goes beyond the RIL definition here. For example, is there a topographic bench below a RIL that limits delivery that would no be evaluated as a part of the terrain features? I don't think we want to use sediment delivery as a screen for hazard, just for risk.

Commented [PJ54]: Many screening tools were made before lidar availability and use low-resolution data. I'd delete this modifier, especially since the sentence references studies that were written before lidar was widely available.

Commented [KJ55R54]: deleted lidar to just state "high resolution"

of geospatial, spectral, thematic, and contextual data to enhance accuracy and precision of landform delineation and classification.

Numerous researchers have developed computer algorithms during the last forty years to delineate and classify land surface features using some combination of digital terrain and spectral data at a range of geospatial scales, from individual landforms to terrain or landscape levels, and for a variety of academic and applied purposes (e.g., natural hazards, communal water supply issues, forest management). Recent examples include semi-automated computational delineations of:

1. landform types – drumlins (Wang et al., 2017; Saha and Van Landeghem, 2021), gullies (d'Oleire-Oltmanns et al., 2014), dunes (Putri et al., 2021), landslides (Van den Eeckhaut et al., 2012; Li, et al., 2015), channel networks (Sofia et al., 2011), and roads (Sherba et al., 2014);
2. individual landforms – e.g., glacial moraines (Robb et al., 2015), geologic lineaments (Yeomans et al., 2019), and Martian and lunar impact craters (Bue and Stepinski, 2007; Vamshi et al., 2016);
3. multiple, spatially contiguous landforms – e.g., ridges, valleys, and floodplains (Gercek and Ugur, 2010); and
4. terrains- e.g., karst (Zylshal et al., 2013), loess plateaus (Lin et al., 2022), and volcanic flows (Feizizadeh et al., 2021).

In the Pacific Northwest, semi-automated methods with forest management applications have been developed to detect and classify terrain elements susceptible to shallow landslide initiation and runout (e.g., Miller and Burnett, 2007, 2008), deep-seated landslide landforms (e.g., Booth et al., 2009; Justice, 2021), and spatially contiguous landforms at multiple geospatial scales (e.g., Shaw et al., 2017).

Increasingly, researchers have turned to computer-automated methods for landform detection and classification because they can be applied efficiently across broad landscapes that might otherwise be too challenging logistically and technically to map manually. In addition, computer scripts can provide more consistent and reproducible results than individual manual mappers if applied systematically across the landscape. Observer bias is inherent, to some degree, in any mapping method, whether the observer is field mapping landforms (e.g., Rollerson, 1997), hand-digitizing landform polygons on GIS data layers, or deciding how best to parameterize a computational model. Once a computer algorithm is generated, however, it can be applied objectively, routinely, and transparently across a geospatial area of interest.

One method of automated landform detection and classification employs object-based image analysis (OBIA), or geographic object-based image analysis (GEOBIA) if earth-surface imagery and data are applied. OBIA originated decades ago as a technique for automating the partitioning and classifying of homogeneous regions (i.e., objects) observed in X-rays and other medical imagery.

Commented [JM56]: Are we trading efficiently for accuracy?

Commented [AW57R56]: Valid question. Higher bar to assess if the completed computer algorithm is accurate as it is still a human generated product even if the algorithm can be *applied* objectively

Commented [AW58]: Sounds like a marketing brochure.

Commented [KJ59R58]: thanks

GEOBIA emerged in the early 2000's as a method for partitioning and classifying earth-surface data (e.g., Darwish et al., 2003; Dragut and Blaschke, 2006; Blaschke, 2010; Dragut and Eisank, 2012). OBIA is distinguished from raster-based methods for automating landform mapping by its use of data segmentation techniques that aggregate pixels into objects in an "image" (i.e., geospatial data product). These techniques facilitate segmenting data into objects at native data resolution (e.g., pixel scale) or as grouped pixels. Objects are created by identifying homogeneous regions within the image based on selected geomorphometric, spectral, and/or contextual attributes. Homogeneity can be defined by spatial metrics (e.g., object compactness and smoothness) and by spectral values (e.g., relative intensities of lidar DTM grey-scale images). Objects retain statistical, geometrical, and nearest-neighbor relational data that allow them to be directly delineated and classified according to the chosen landform classification system (Blaschke and Strobl, 2003; Gercek and Ugur, 2010).

Dragut and Blaschke (2006) and Gercek and Ugur (2010) pioneered the use of eCognition (Trimble Geospatial) software in automating geospatial feature extraction, using computer algorithms to identify and isolate patterns in geographic data and then classifying specific patterns as landforms. ~~and object-based classifications of landforms.~~ More recently, researchers have used eCognition o~~r~~ comparable programming software to apply OBIA methods on an expanding array of land surface features.

This report evaluates two landform mapping approaches: the ~~Terrainworks'~~ Virtual Watershed (VW) automated model for landform detection developed by ~~(~~Terrainworks, 2015~~?: -~~(described in section 3.~~1.~~3) and an automated model for landform detection and delineation by Shaw et al. (2017) (WY model) using OBIA methods powered by eCognition software (described in section 3.~~1.~~2). Both are based partially or wholly on OBIA techniques and are used to satisfy the objectives stated in the ~~Unstable Slopes Criteria Project: Study Design for Object-Based Mapping with High-Resolution Topograph~~Unstable Slopes Criteria Project: Study Design for Object-Based Mapping with High-Resolution Topography (Dieu et al. 2018):

- Identify methods for consistent automated delineation of landforms using Geographic Object-Based Image Analysis (GEOBIA or OBIA) techniques and high-resolution LiDAR DTMs, and potentially other data sources.
- The automated landform models will provide the baseline geomorphic context from which to evaluate landslide susceptibility and runout, and it will incorporate data from process-based models to train the automated classification of landforms.

eCognition software operates with a user-friendly interface (e.g., for the less experienced computer programmer) that facilitates data segmentation and classification using a large assortment of ~~on-board, sophisticated, and~~ documented tools for geospatial, spectral, and temporal (e.g., change-detection) analyses. eCognition users develop "rulesets" (i.e., computer scripts) that take advantage of software capabilities in analyzing data at multiple scales and

Commented [PJ60]: What is "geospatial feature extraction?" What is extracted?

Commented [JT61R60]: Reworded to explain that landforms are extracted from isolated patterns in geographic data.

Commented [MM62]: Is this the Weyerhaeuser model mentioned below?

Commented [PJ63]: These are not mentioned in the executive summary section but are a key part of the project. Please include these mapping approaches in the executive summary

Commented [PJ64]: If these are the main objectives of the project, it seems like they should be presented early in the document

Commented [PJ65]: I don't know what these mean in the context of this software. Needed?

hierarchical levels, processing multiple data types simultaneously~~,~~ and efficiently handling massive volumes of high-resolution data via a batch server application. The ~~Weyerhaeuser model (Shaw et al., 2017)~~ WY model evaluated in this report uses OBIA methods to segment, extract, and classify multiple lidar DTM derivative products (e.g., surface elevation, curvature, gradient) according to established geomorphometric criteria, as part of a comprehensive mapping package of contiguously mapped landforms (e.g., RIL and non-RIL landforms).

The goal of OBIA is to train the computer to emulate landform mapping conducted using conventional ~~field~~ interpretive ~~methods or~~ remote-sensing techniques but with the much greater standardization, repeatability, and geographic transportability that comes with an established computer script or ruleset. eCognition-powered OBIA also allows independent users to generate identical maps when using the same software package, input data, and transparent rulesets.

~~OBIA methods work relatively well~~best ~~when identifying more recently eroded landforms and activated or reactivated landslides because they tend to have crisp, discrete boundaries that can be delineated and classified readily from optical imagery and lidar DTM (e.g., Dai et al., 2020; Garajeh et al., 2022). A challenge for OBIA, or any other landform mapping technique, is identifying and delineating older, eroded landforms with indistinct boundaries and complex scale dependencies (e.g., dormant, deep-seated landslide complexes common in the Puget Lowlands). Etzelmuller and Sulebak (2000) have observed that geomorphic systems must be viewed in their complex hierarchical context. eCognition tools for analyzing objects hierarchically, applying edge-detection algorithms, and computing nearest-neighbor statistics can facilitate extracting complex, indistinct landform features and expressing geospatial uncertainties in landform boundaries.~~

Landform mapping critically depends on scale for extracting and delineating landform features based on parameters used in geomorphometric analyses (e.g., topographic curvature and gradient; Dragut and Eisank, 2012). Length scales or grid window sizes for surface calculations typically are determined heuristically (i.e., via a trial-and-error approach). Evans (2003) demonstrated convincingly that the size and spacing of individual landform types can be clustered around characteristic scales relative to process thresholds and/or space available for landform evolution. He also stated that specific landforms have characteristic dimensions and occur at different scales depending on a hierarchy of landform types and sizes. Consequently, length scales should be chosen to represent the landform size being extracted. ~~A method for automating length scale selection using sophisticated analysis techniques on high-resolution lidar DTM data would be advantageous; however, such a technique has not been developed yet and probably would require the use of a multi-scale or hierarchical approach.~~

Commented [PJ66]: This is the first mention of the Weyerhaeuser model. Is there a new eCognition model for this project? Or just the Weyerhaeuser model? If it uses the Weyco model or if it's a NEW model - that info should be in the executive summary

Commented [KJ67R66]: OK, add the reference for this model was moved to the paragraph above, where it is first mentioned.

Commented [PJ68]: This should be BH/IG and non-BH/IG. Please check the document.

Commented [PJ69]: Is it to replace field work? Or just to be used as a screening method? If it is to replace field work, then it should be trained with field-collected data.

Commented [JM70]: What does this mean, 50%, 70% of the time?

Commented [KJ71R70]: changed to "best"

Commented [PJ72]: I'm not sure what this sentence means.

Commented [JM73]: Trial and error?

Commented [KJ74R73]: yes, added (i.e., via a trial and error approach)

Commented [PJ75]: Is this planned? What's holding us back from developing this technique? Seems like it could be important when looking comparing landforms across the state in different terrain.

1.3 Objectives and Deliverables

~~As stated in~~ Section 1.2, the 2018 Study Design for this sub-project, Object-Based Mapping with High-Resolution Topography (OBM), stated two objectives:

- Identify methods for consistent automated delineation of landforms using Geographic Object-Based Image Analysis (GEOBIA or OBIA) techniques and high-resolution LiDAR DTMs, and potentially other data sources.
- The automated landform models will provide the baseline geomorphic context from which to evaluate landslide susceptibility and runout, and it will incorporate data from process-based models to train the automated classification of landforms.

Commented [JM76]: Is it assumed that consistent is the same as accurate?

As described in this report, we sought to meet these objectives through development of two automated modeling methods that are based on geomorphic characterization of terrain features. One using image analysis with classic feature extraction techniques and a second using the virtual watershed approach, both based entirely on lidar topographic data.

The study design specified the following list of deliverables expected from this sub-project, which are all included in this report:

1. Vector-based multi-scale landform maps as baseline GIS files for the pilot and three additional study areas;
2. The frequency distribution and the statistics of topographic attributes describing each landform;
3. Comparison of frequency distributions and statistics of topographic attributes between manual and automated landform maps;
4. Tools for producing all topographic indices;
5. eCognition rule-sets and codes; and
6. A report describing our experience using LiDAR and object-based models to identify specific categories of unstable landforms as found at the project scale, the transferability of such models, lessons learned, and recommendations for future research.

1.4 Changes from the Study Design

The Research Alternatives (USC TWIG 2017) and OBM Study Design (Dieu et al., 2018) both called for evaluation of the automated methods by comparing the computer-generated maps with existing Mass Wasting Map Unit (MWMU) maps from Landslide Hazard Zonation (LHZ) products. That would require that the computer models provide contiguous (wall-to-wall) landform mapping, as the example in Figure 1 in the OBM Study Design illustrated. As the Project Team embarked on the OBM project, however, concerns about the precision and provenance of the MWMU maps were raised. The MWMU mapping was hand-drawn on 1:24,000-scale topographic base maps, whereas we now have high-resolution (~1m) DTMs and can manually map terrain features at much larger map scales, e.g., 1:500. The computer models use the high-resolution lidar DTMs, so the computer-model results are based on a different data source than the MWMUs. Additionally, mappers included varying levels of detail in the MWMU maps, raising concerns about their consistency between watershed administrative units.

The Project Team included members with extensive training and experience with landform mapping. We decided, therefore, that a subset of the team would draw a new set of landform maps using information inferred from the lidar DTMs, the same data source used by the computer models. All team members had experience mapping RILs, both on the ground and remotely using topographic maps, but not all of them had experience mapping MWMUs (particularly as the term MWMU is used in LHZ, where all the landscape is mapped as a landform). Therefore, the comparison between manual mappers and the models was limited to bedrock hollows and inner gorges.

As detailed in Sections 4, 5, and 6 of this report, comparing the computer-generated maps to manually delineated maps proved to be an inadequate evaluation approach. To briefly summarize: the Project Team produced inconsistent maps among themselves, resulting in no reliable reference product (a “correct map”) to assess the model results against. In response, the team developed alternative evaluation methods, which are described in Section 4.2.

The objectives in the study design (listed in Section 1.3) were intentionally framed with flexibility to accommodate the limited information and emerging methodologies available at the time they were developed. This sub-project demonstrated some of the difficulties encountered when attempting to automate a process typically performed manually and in-person by practitioners. Manual delineation of landforms involves subjective interpretation of remote data and rule definitions. This subjectivity led to inconsistent landform mapping results and landform delineation agreement. These inconsistencies precluded the use of landform maps as a reliable reference for automated methods. This forced us to reconsider and narrow our objectives simply to the first: to “identify methods for consistent automated delineation of landforms.”

Establishing a “geomorphic context from which to evaluate landslide susceptibility and runout” requires maps that reliably depict the extent of geomorphic processes directly influencing

landslide occurrence. The results of this sub-project provide useful insights into estimating that extent, but are based on subjective judgements, not empirical landslide data. Without such data, we cannot reliably determine the geomorphic context from which to evaluate landslide susceptibility and runout.

With the next subproject (Susceptibility and Runout), we seek to obtain that data using landslide inventories based on lidar differencing. As described in detail in Sections 6.2-6.4, we intend to identify the range of terrain characteristics associated with landsliding using the locations of these precisely mapped landslides. Constraining the geomorphic conditions where landslides are observed will provide part of the geomorphic context with which to guide the automated delineation of landforms. Methods for delineation and categorization of landforms are described in Section 6.0. Recognizing that delineation of the boundary area directly influencing landslide potential may vary, this next sub-project will include a sensitivity analysis. As briefly described in Section 6.3 of this report, this analysis will evaluate how measures of landslide susceptibility (density and proportion) respond to changes in the landform boundaries mapped. The final subproject, Physical Modeling, will then aid in constraining the range of potential landform extents by estimating the area with direct influence on landslide potential and how that area may vary depending on storm events and land cover.

2.0 DATA

2.1 STUDY SITES

This study focused on ~~study~~ sites in ~~4~~four areas (Figure 1): the North Fork Calawah Watershed Administrative Unit (WAU) (Sites 1A and 1B), the Willapa Hills~~Post Mortem~~ study area (Site 2) in the Stillman Creek WAU, the H~~arry Osborne~~oward Creek study area (Site 3) in the Howard Creek WAU and the Wishkah study area (Site 4) – which spans portions of the Lower Wishkah, Wynoochee River South, and Chehalis Sloughs WAUs. The Project Team selected a smaller section from each WAU to focus on. We chose two sites for separate exercises in the North Fork Calawah WAU.

Commented [GS77]: Because this project is so dependent on lidar interpretation, I think you should include information on the lidar flights used, XYZ resolution, etc.

Commented [PJ78]: It isn't showing up on the figure.

Commented [MM79]: Yellow- I would refer to this area as the Willapa Hills and rename throughout the document. You should cite the "Post Mortem" report by it's actual name to reduce confusion- a Google search is probably not going to take the reader where you want them to go.

Figure 1: Study area locations ~~over map of~~in western Washington.

The project team chose t~~T~~he North Fork Calawah WAU (Sites 1A and 1B) ~~was chosen~~ as a pilot area because cooperators from the Rayonier Corporation, including a qualified expert with decades of experience in RIL identification, had previously field-mapped RILs in the area to avoid timber harvest on unstable slopes per WAC 222-16-050. The Project Team then chose three other study areas that:

1. Provided a range of topographic, geologic, and climactic characteristics;
2. One or more Project Team members had field experience;
3. Current lidar was available for (see Table 2 for Lidar specifications); and
4. Could serve as landslide inventory data collection sites for sub-project 2.

~~4.~~

Table 2: Lidar information for each site.

Site	Dataset Name	XYZ Resolution	Flight Dates / Info	Citation
1A/1B	Olympics North OPSW 2018	≥ 8 points/m²; 3 ft rasters	Multiple dates: 10/5/17 to 12/5/18 (Olympic Peninsula Area 3)	Washington Geological Survey, 2019, *Olympics North OPSW 2018* [lidar data]: Collected under contract by U.S. Geological Survey (USGS), at https://lidarportal.dnr.wa.gov
2	Southwest WA OPSW 2019	≥ 8 points/m² (0.35 m spacing); 3 ft rasters	March 17, 2016 – June 6, 2017 (South AOI from Western WA 3DEP project)	Washington Geological Survey, 2019, *Southwest Washington OPSW 2019* [lidar data]: Collected under contract by U.S. Geological Survey (USGS), at https://lidarportal.dnr.wa.gov
3	North Puget 2017	≥ 8 points/m²; 3 ft rasters	March 17, 2016 – Sept 30, 2016 (North AOI from Western WA 3DEP project)	Washington Geological Survey, 2017, *North Puget 2017* [lidar data]: Collected under contract by U.S. Geological Survey (USGS), at https://lidarportal.dnr.wa.gov
4	Olympics South OPSW 2019	≥ 8 points/m²; 3 ft rasters	Multiple dates: 2/19/18 – 4/25/19 (Olympic Peninsula Area 2)	Washington Geological Survey, 2019, *Olympics South OPSW 2019* [lidar data]: Collected under contract by U.S. Geological Survey (USGS), at https://lidarportal.dnr.wa.gov

Site 1A & 1B– North Fork Calawah and Fahnestock Creek

The Project Team selected two study sites from the North Fork Calawah WAU. The pilot area, Site 1A, ~~(Figure 2)~~ has an area of 4.2 square kilometers (Figure 2). Site 1B has an area of 1.34 square kilometers. These sites have a mean annual precipitation of 309 centimeters per year since 2013 (PRISM Climate Group, 2024).

The core of the Olympic Mountains is uplifting at approximately 1mm/yr (Pazzaglia and Brandon, 2001), creating a landscape of high erosion with steep slopes and thin soils. Correspondingly, channel density is high, and most of the low-order channels route into the larger tributaries (Dieu and Shelmerdine, 1997). The bedrock in the study area dips northeast and is composed of Oligocene-Eocene marine sedimentary rocks and quaternary alluvium along Fahnestock Creek. The marine sedimentary rocks are composed of lithofeldspathic and felspatholithic micaceous sandstone which is massive and well indurated; lesser siltstone, slate and semischist slate/phyllite occur in thin lenses (Dragovitch et al., 2002).

An anticline ~~cuts down~~extends through the middle of the project area and a syncline extends along~~runs down~~ the west side, both northwest-southeast trending (Schuster, 2005). The study area is bisected by Fahnestock Creek, which ~~runs~~flows east-to-west and ~~feeds~~ into the North Fork Calawah River. The hillslopes along gorges feeding into Fahnestock Creek are predominantly steep, with slopes greater than 70 percent in most areas. The hillslopes in the northeast portion of the study area (north of the creek) are gentler, between 40-60 percent. The highest elevations are found on the ridges on the east half of the site, and elevations decrease to the west toward the North Fork Calawah River.

Commented [PJ80]: Seems like this should be section 2.1.1

Commented [PJ81]: It would be helpful to have imperial units in parentheses (throughout)

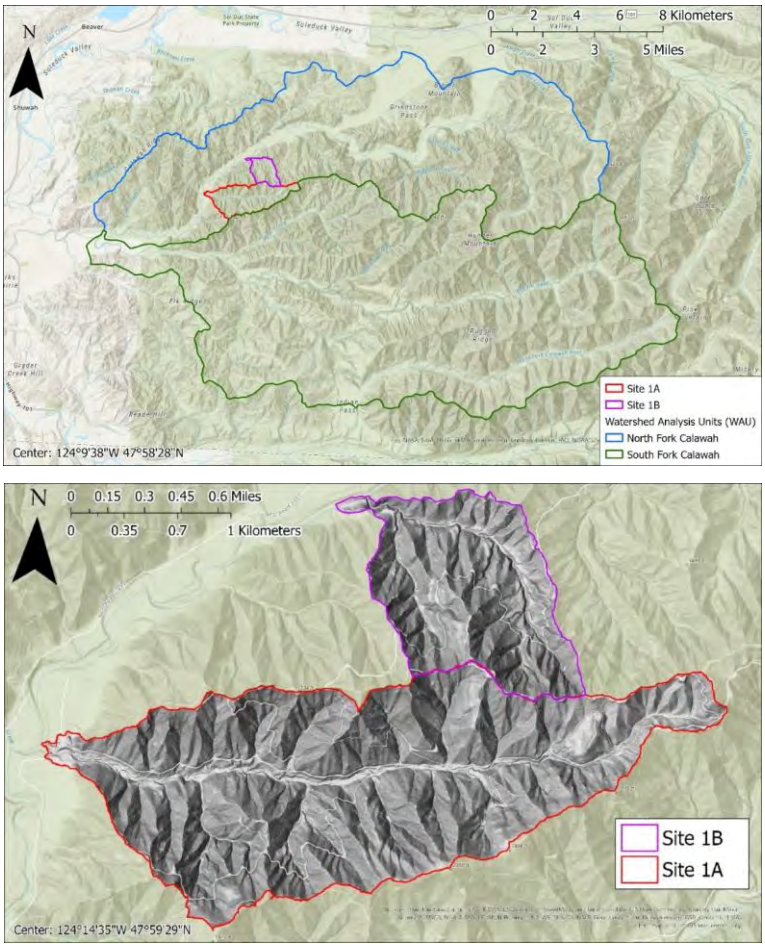


Figure 22: Location map of Calawah study area, with the North Fork Calawah and South Fork Calawah WAUs indicated in blue and green, Site 1A indicated in red and Site 1B in purple (top). Bottom figure shows bare earth hillshade layer over the study area.

~~Geology/Geomorphology~~

~~The core of the Olympic Mountains is uplifting at approximately 1mm/yr (Pazzaglia and Brandon, 2001), creating a landscape of high erosion with steep slopes and thin soils. Correspondingly, channel density is high, and most of the low-order channels route into the larger tributaries (Dieu~~

Commented [GS82]: For each of these location maps, I recommend having the hillshade image as the larger map and the location map as the inset. Would be much more valuable to see the topography of the watersheds themselves. The topo maps currently used as the main maps do not provide much beyond what is shown in figure 1.

Commented [PJ83R82]: I agree with Gus and I'd add that any text on figures should be large enough to read. Currently the legend, scale bar, and text are much too small. This comment applies to all location figures.

Commented [MM84R82]: Agree- it also took me a bit to figure out that the inset wasn't just a highlighted part of the other map.

Commented [JM85R82]: I agree.

Commented [PJ86]: The geology/geomorphology sections for each site should be integrated into the site description and not broken out as a separate sub-section.

~~and Shelmerdine, 1997). The bedrock in the study area dips northeast and is composed of Oligocene-Eocene marine sedimentary rocks and quaternary alluvium along Fahnestock Creek. The marine sedimentary rocks are composed of lithofeldspathic and felspatholithic micaceous sandstone which is massive and well indurated; lesser siltstone, slate and semischist slate/phyllite occur in thin lenses (Dragovitch et al., 2002).~~

~~An anticline cuts down the middle of the project area and a syncline runs down the west side, both northwest-southeast trending (Schuster, 2005). The study area is bisected by Fahnestock Creek, which runs east-to-west and feeds into the North Fork Calawah River. The hillslopes along gorges feeding into Fahnestock Creek are predominantly steep, with slopes greater than 70 percent in most areas. The hillslopes in the northeast portion of the study area (north of the creek) are gentler, between 40-60 percent. The highest elevations are found on the ridges on the east half of the site, and elevations decrease to the west toward the North Fork Calawah River.~~

Commented [PJ87]: This is duplicate

Site 2~~—~~ – Willapa Hills~~Post Mortem~~

The Willapa Hills~~Post Mortem~~ study area (Site 2, Figure 3) has an area of 13.5 square kilometers and a mean annual precipitation of 192 centimeters per year since 2013 (PRISM Climate Group, 2024). The project team ~~It was~~ chose~~n~~ this area due to the inventory of field-based landslide locations available (Stewart et al., 2013).

The eastern Willapa Hills are old, relatively stable terrain with numerous, broadly convex surfaces. Channel dissection is lower than at Site 1, primarily because deep soils in excess of 2m and regolith profiles of 5-10m are common (e.g., Stewart et al., 2013). The surface materials facilitate the infiltration of precipitation, much of which is retained as shallow groundwater until reaching larger channels ~~when~~where the water table intersects the surface. The site is composed of lower to middle Eocene Crescent Formation basalt breccia, middle to early Eocene intrusive gabbro and middle Eocene Cowlitz formation tuff.

The Crescent Formation flow basalt breccia consists of fine-grained basalt-lapilli tuff, basaltic sandstone and conglomerate, interbedded with submarine basalt flows, locally pillow lava (Walsh et al., 1987). The units are highly fractured due to faulting and are found on plateaus and benches in the western portion of the site (Laprade, 1994).

The intrusive gabbro unit consists of widespread fine- to very-coarse grained vesicular, flow-banded, gabbro and tuffaceous basaltic intrusions. The unit is found on moderately and steeply sloping hills throughout the site (Laprade, 1994)

The Pe Ell volcanics member of the Cowlitz formation consists of massive to well-bedded, palagonitic basaltic lapilli tuff and tuff breccia, basaltic sandstone, siltstone, and conglomerate (Henrikson, 1956).

Two normal faults are mapped in the site area, oriented northeast and northwest, dipping down to the north at 15-30 degrees (Laprade, 1994). Stillman Creek runs north-south through the east side of the study area and joins Little Mill Creek in the northeast corner. Elevation is highest in the west and decreases to the east and is lowest along Stillman Creek and Little Mill Creek. Slopes are steepest along tributary slopes feeding into the streams and landslide headscarps; few areas are greater than 70 percent.

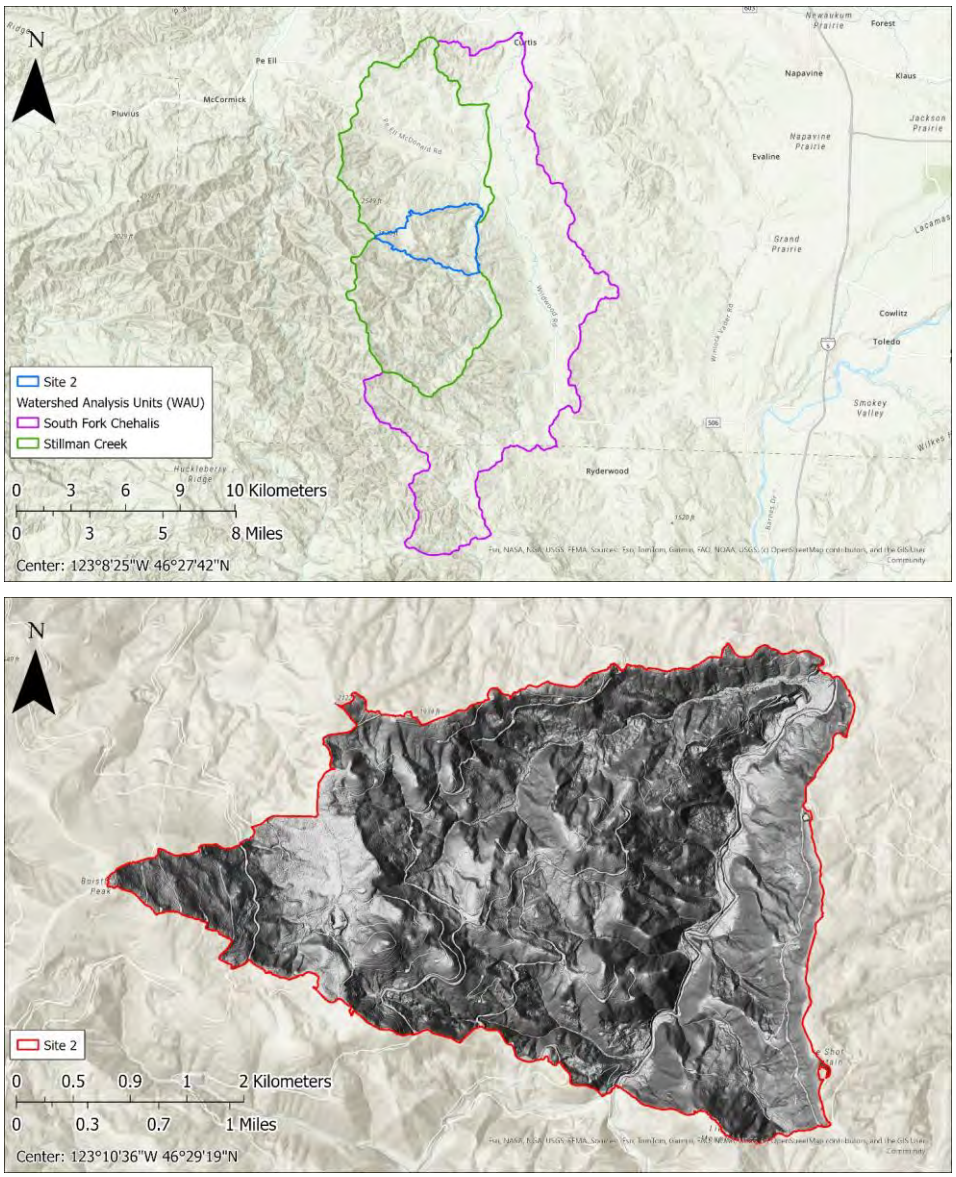


Figure 3: Location map of Willapa Hills study area, with the South Fork Chehalis and Stillman Creek WAUs indicated in purple and green and Site 2 indicated in blue (top). Bottom figure shows bare earth hillshade layer over the study area.

~~***Geology/Geomorphology***~~

~~The eastern Willapa Hills are old, relatively stable terrain with numerous, broadly convex surfaces. Channel dissection is lower than at Site 1, primarily because deep soils in excess of 2m and regolith profiles of 5-10m are common (e.g., Stewart et al., 2013). The surface materials facilitate the infiltration of precipitation, much of which is retained as shallow groundwater until reaching larger channels when the water table intersects the surface. The site is composed of lower to middle Eocene Crescent Formation basalt breccia, middle to early Eocene intrusive gabbro and middle Eocene Cowlitz formation tuff.~~

~~The Crescent Formation flow basalt breccia consists of fine-grained basalt-lapilli tuff, basaltic sandstone and conglomerate, interbedded with submarine basalt flows, locally pillow lava (Walsh et al., 1987). The units are highly fractured due to faulting and are found on plateaus and benches in the western portion of the site (Laprade, 1994).~~

~~The intrusive gabbro unit consists of widespread fine- to very-coarse grained vesicular, flow-banded, gabbro and tuffaceous basaltic intrusions. The unit is found on moderately and steeply sloping hills throughout the site (Laprade, 1994)~~

~~The Pe Ell volcanics member of the Cowlitz formation is mapped here, which consists of massive to well-bedded, palagonitic basaltic lapilli tuff and tuff breccia, basaltic sandstone, siltstone, and conglomerate (Henriksen, 1956).~~

~~Two normal faults are mapped in the site area, oriented northeast and northwest, dipping down to the north at 15-30 degrees (Laprade, 1994). Stillman Creek runs north-south through the east side of the study area and joins Little Mill Creek in the northeast corner. Elevation is highest in the west and decreases to the east and is lowest along Stillman Creek and Little Mill Creek. Slopes are steepest along tributary slopes feeding into the streams and landslide headscarps; few areas are greater than 70 percent.~~

Site 3 - Howard Creek~~arry Osborne~~

The Howard Creek~~arry Osborne~~ study area (Site 3, Figure 4) has an area of 9.9 square kilometers and a mean annual precipitation of 269 centimeters per year since 2013 (PRISM Climate Group, 2024).

Bedrock in the study area is composed of Mesozoic metavolcanic rock and phyllite of Mt Josephine. This bedrock is exposed at higher elevations where hillslopes of moderate to steep gradients and significant erosion potential exist. At lower elevations, the bedrock is overlain by Pleistocene continental glacial drift along the South Fork Nooksack River and the tributaries that feed into it (Dragovich et al., 2002). These slopes are of low and moderate gradients with narrow

Commented [PJ88]: RED: The Harry Osborne state forest is in the Skagit Valley, not the Nooksack. If it is in the Nooksack watershed, it should not be called the Harry Osborne site.

Commented [KJ89R88]: Thanks, we will consider renaming to "Harry Osborne" to "Howard Creek"

bands of steeper slopes created by both glacial (i.e., relict features) and modern erosional processes.

The phyllite is composed of foliated graphite, muscovite or sericite-quartz phyllite with abundant quartz veins or lenses. In some areas, it is interlayered with cataclastic sandstone, greenschist, and blueschist. The metavolcanic rocks are composed of Jurassic metabasaltic and basaltic meta-andesite flows, flow breccias, greenstone, dacite to andesite flows, tuffs, and breccia with argillite interbeds. The Pleistocene continental glacial drift is composed of glacial till and outwash sediments (clay, silt, sand, gravel, cobbles and boulders) (Dragovich et al., 2002).

Two faults have been mapped in the southwest portion of the study area. One fault has a northwest-southeast trend and the other is an east-west trending thrust fault with the upper block on the south side (Schuster, 2005). The study area drains towards the Skagit river. Some slopes along tributaries are steeper than 70 percent. The elevation is highest on the ridges to the north and south of the study area and decreases toward the river running through the center.

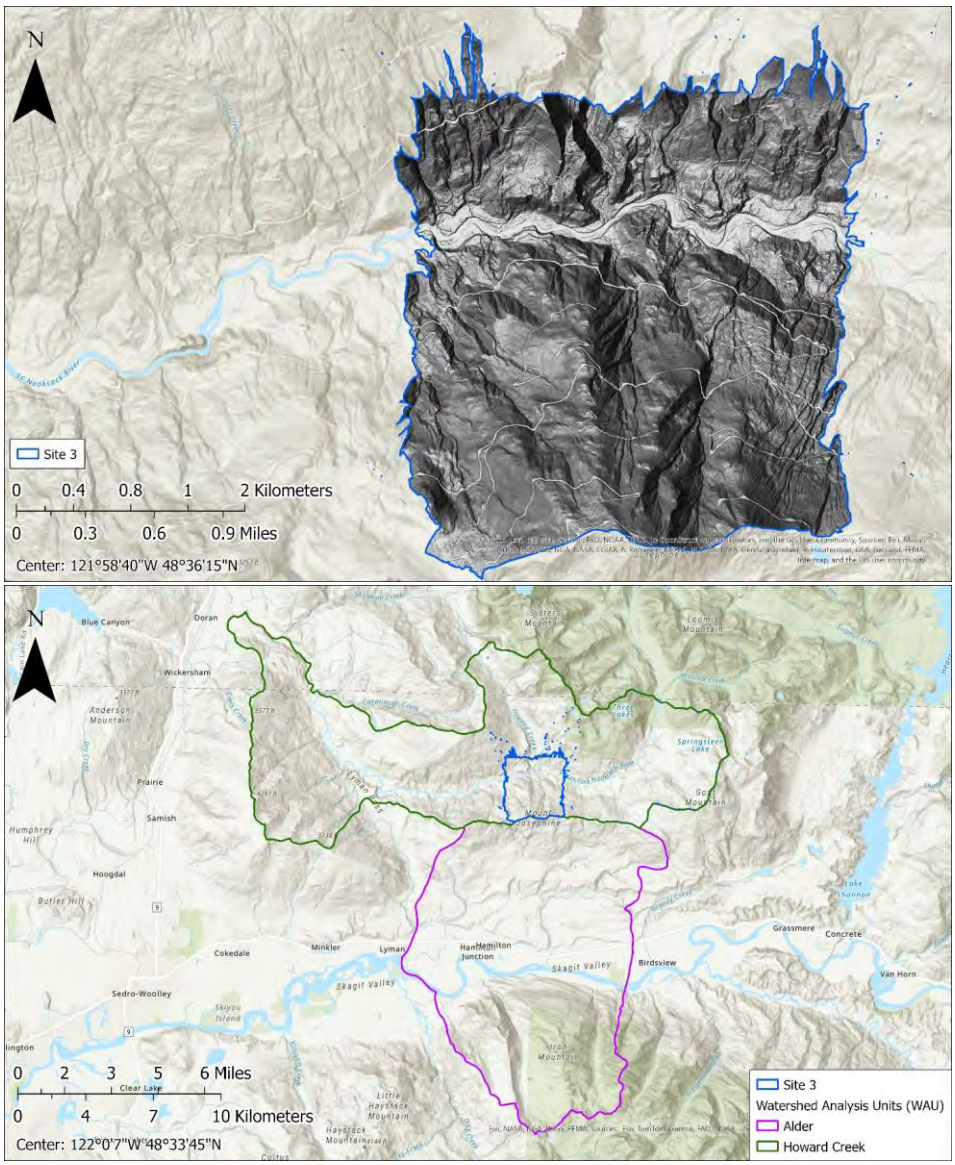


Figure 4: Location map of Howard Creek study area, with the Alder and Howard Creek WAUs indicated in purple and green and Site 3 indicated in blue (top). Bottom figure shows bare earth hillshade layer over the study area.

~~Bedrock in the study area is composed of Mesozoic metavolcanic rock and phyllite of Mt Josephine. This bedrock is exposed at higher elevations where hillslopes of moderate to steep gradients and significant erosion potential exist. At lower elevations, the bedrock is overlain by Pleistocene continental glacial drift along the South Fork Nooksack River and the tributaries that feed into it (Dragovich et al., 2002). These slopes are of low and moderate gradients with narrow bands of steeper slopes created by both glacial (i.e., relict features) and modern erosional processes.~~

~~The phyllite is composed of foliated graphite, muscovite or sericite-quartz phyllite with abundant quartz veins or lenses. In some areas, it is interlayered with cataclastic sandstone, greenschist, and blueschist. The metavolcanic rocks are composed of Jurassic metabasaltic and basaltic meta-andesite flows, flow breccias, greenstone, dacite to andesite flows, tuffs, and breccia with argillite interbeds. The Pleistocene continental glacial drift is composed of glacial till and outwash sediments (clay, silt, sand, gravel, cobbles and boulders) (Dragovich et al., 2002).~~

~~Two faults have been mapped in the southwest portion of the study area. One fault has a northwest-southeast trend and the other is an east-west trending thrust fault with the upper block on the south side (Schuster, 2005). The study area is bisected by the South Fork Nooksack River, with creeks feeding into the river from ridges to the north and south~~drains towards the ~~Skagit river~~. ~~Some slopes along tributaries are steeper than 70 percent. The elevation is highest on the ridges to the north and south of the study area and decreases toward the river running through the center.~~

Site 4 - Wishkah

The Wishkah study area (Site 4, Figure 5) has an area of 5.16 square kilometers and a mean annual precipitation of 220 centimeters per year since 2013.

The bedrock in the study area is composed of Tertiary marine sedimentary rocks. At the highest elevations, these are overlain by an iron-cemented alluvial material originally identified as Qoa, Quaternary old alluvium (Tabor and Cady, 1978) which forms an erosion-resistant cap; it has been more recently identified as alpine outwash of older pre-Wisconsinan (Pleistocene) origin (Logan, 2003). Quaternary alluvium, occurs along the along the lower tributaries such as Van Winkle and Bear creeks (Walsh et al., 1987). The study area encompasses the lower valley floors, intermediate hillslopes of moderate gradient and deep soil formation where the Tertiary marine sediments are exposed, and high ridges with over-steepened edges that are relict, deep-seated landslide scarps (Othus and Parks, 2009) created when catastrophic failure extends up into the $Qapwo_1$. Slopes greater than 70 percent can be found on hillslopes along the tributaries feeding the creeks and are gentler along Bear and Van Winkle creeks.

Commented [MM90]: Probably don't need unit abbreviations here- they don't provide clarity and aren't used in other study areas.

Commented [KJ91R90]: deleted abbreviations

The Tertiary marine sediments are composed of lithofeldspathic and feldspatholithic sandstone and siltstone; locally siltstone is the predominate rock type, and it is deeply weathered and very friable. In the southeast area of the site, the bedrock is overlain by Pleistocene alpine glacial drift composed of sandy gravel deposits with interbedded silt lenses (Logan, 2003).

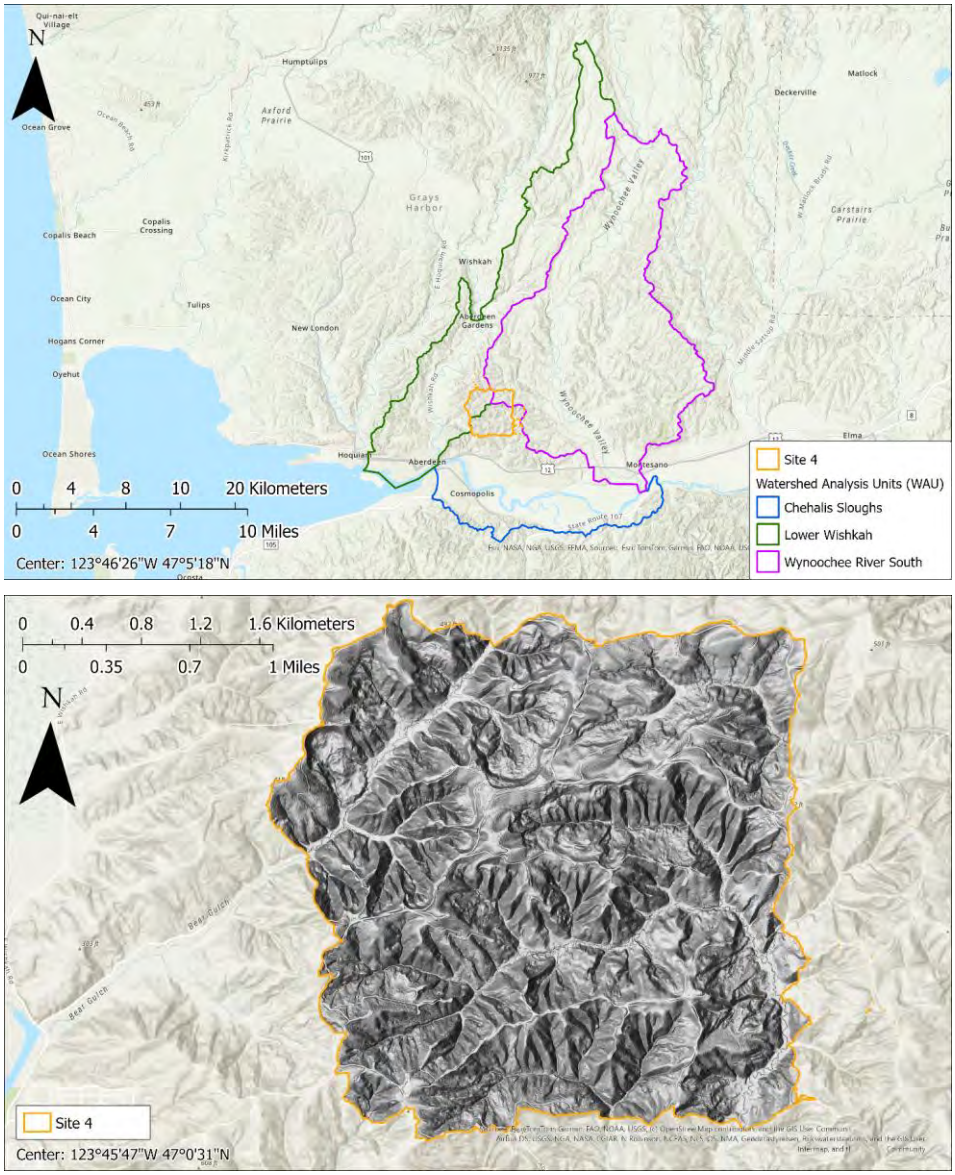


Figure 5: Location map of Wishkah area, with the Chehalis Sloughs, Lower Wishkah and Wynoochee River South WAUs indicated in blue, purple and green and Site 3 indicated in yellow (top). Bottom figure shows bare earth hillshade layer over the study area.

~~The Tertiary marine sediments are composed of lithofeldspathic and feldspatholithic sandstone and siltstone; locally siltstone is the predominate rock type, and it is deeply weathered and very friable. In the southeast area of the site, the bedrock is overlain by Pleistocene alpine glacial drift composed of sandy gravel deposits with interbedded silt lenses (Logan, 2003).~~

2.2 Eco-Regions

Sites 1A, 1B, 2 and 4 all reside within the "Northwest Coast" Level III Eco Region, which are areas home to Douglas fir-dominant coniferous forests with lesser Sitka spruce, red alder and big leaf maple, with salal and salmonberry in the understory (U.S. EPA, 2012). Sites 1A & 1B lie within the further-subdivided Level IV Eco-Region "Low Olympics," described as a "lush, epiphyte-rich rainforest" with western hemlock, western redcedar, and Douglas fir. Site 4 is in the Level IV Eco-Region "Outwash," which is described as glacial outwash materials in areas with high timber industry activity, that lie outside influence of marine processes and has lower stream flows (Pater et al., 1998).

Site 3 lies in the Level IV Eco-Region "North Cascades Highland Forests" within the Level III "North Cascades" region, described as steep, glaciated ridges with colder climate conditions, home to Pacific silver fir/mountain hemlock forests (Pater et al., 1998). Lower elevations in Site 3 are~~fall~~ in western hemlock forest zone, transitioning to silver fir and mountain hemlock in the higher elevations (Franklin & Dyrness, 1973).

Commented [GS92]: The highest elevations of Mount Josephine may be in the silver fir/mountain hemlock zone, but most of the area is solidly in the western hemlock forest zone.

2.3 Topography

How do topographic characteristics differ between the four study sites? The shaded relief images in Section 2.1 provide a visual comparison. We can also compare quantitative measures of ~~model-based topographic attribute~~ topographic attributes measured from the DEM~~s~~, four of which are shown by the bar charts in Figure 6. These reveal a rather striking difference between the Site 1A and the other three sites: Site 1A has the largest channel density (channel length per unit basin area), the steepest mean slope, and by far, the highest density of zero-order basins[2], with the corresponding smallest mean zero-order basin size. By these measures, the other three sites appear relatively similar, with Site 3 having a significantly higher channel density than the other two. Differences in the frequency distributions of gradient are shown in Figure 7. More than half of the area of Site 1A ~~area~~ is steeper than 70%, whereas at the other three sites, less than 20% of the area is steeper than 70%. These modeled differences in topography translate to differences in the number and size of IG and BH landforms between the sites. Steeper slopes together with greater channel length and number of zero-order basins per unit basin area suggest greater

Commented [GS93]: These would be much more effective if they were blown up to the size of the full map. You could also experiment with underlays of the bare earth DEM set to a fixed elevation range scale across all maps, emphasizing absolute elevation differences across the sites.

Commented [GS94]: 'Model-based topographic attributes' is confusing. I think what you mean is 'topographic attributes measured off the DEM'. Is the model you are referring to the DEM? If so, I suggest using the above language or similar. If you are talking about some other model or technique for deriving these attributes, provide clarifying language.

Commented [KJ95R94]: Changed to "topographic attributes measured from the DEM"

Commented [PJ96]: My understanding is that footnotes go on the outside of commas and periods. Search the document.

[2] A "zero-order" basin is the area that drains to the upslope extent of a channel. Bedrock hollows form within zero-order basins. See Appendix B.

potential for existence of ~~inner gorge~~inner gorge and ~~bedrock hollow~~bedrock hollow RILs. Detailed descriptions of the topographic indices are included in Appendix E.

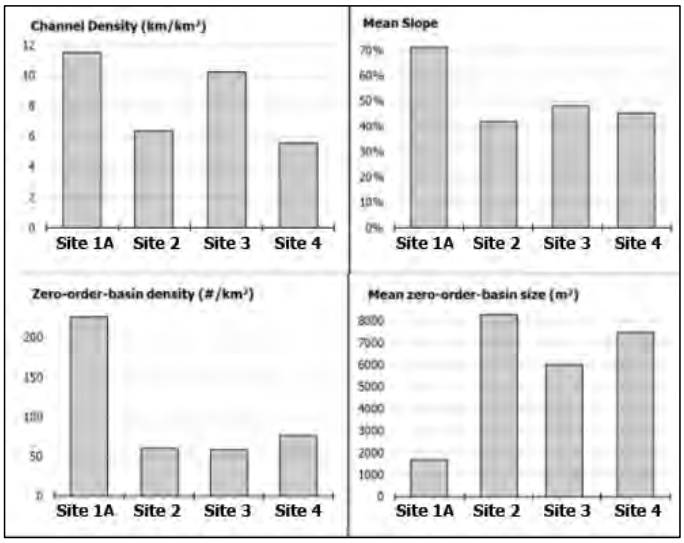


Figure 6: Topographic characteristics of the four study sites.

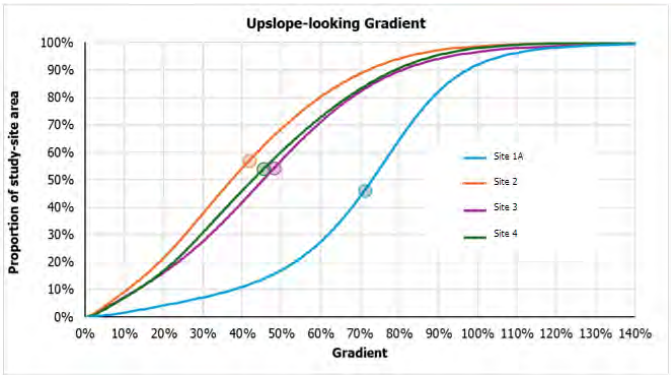


Figure 7: Frequency distributions of surface gradient for the four study sites. Circles indicate the mean.

3.0 MAPPING METHODS

~~*MAPPING METHODS*~~

We used manual and automated mapping methods to produce landform maps based on RIL definitions and Board Manual guidance for ~~inner-gorge~~inner gorge and ~~bedrock-hollow~~bedrock hollow landforms. ~~The manual mapping was executed by 4-5~~ Four~~-~~ to ~~-~~five Project Team members, all experienced practitioners (EP), independently ~~executed~~conducted the manual mapping. ~~and the automated mapping methods were developed by T~~two of the Project Team members developed the automated mapping methods based on input provided by the entire Project Team. The EPs ~~who participated in this sub-project~~ have a broad range of experience and backgrounds in modeling, field mapping, and remote mapping, which created sample diversity in these exercises. ~~The manual mapping was performed independently by each mapper. The mapping team then identified the best landform map for comparison with the automated mapping products by voting separately on individual polygons.~~ The manually-generated landform maps were used in Phase 1 of this study to compare with the computer-generated OBIA and VW landform maps.

The computer-generated landform maps were produced using two approaches: an object-based-image-analysis (OBIA) approach described in Section 3.1.2 and Appendix A and a virtual watershed (VW) approach described in Section 3.1.3 and Appendix B. Although these methods use some of the same input data for mapping landforms, they approach that task in different ways. OBIA uses image-analysis techniques to detect spatial patterns in topographic attributes derived from the DTM. Rules are specified to delineate specific types of patterns into distinct zones. These rules reflect expectations of how differences in topography differentiate landform types. The polygons encompassing these zones are called objects and these objects are then classified into landform types. The virtual watershed approach works with the same underlying DTM data and with the same topographic attributes derived from the DTM to identify patterns, but it starts with the landform definitions and then traces out the zones that meet those definitions. Key components of those definitions are specifications of how landforms are related spatially in terms of the movement of water and sediment through the landscape. For that, a virtual watershed first determines patterns of connectivity of all DTM grid points via surface flow paths and Euclidean distance. These two modeling approaches can produce different outcomes. Those differences can help us identify ambiguities or alternatives to use in our landform definitions. The computer-generated OBIA and VW landform maps were compared with manual mapping in Phase 1 and were evaluated with a voting exercise in Phase 2 of this project.

Commented [PJ97]: Please provide a summary of the methods to help the reader understand what occurred in Phase 1 vs Phase 2. It's confusing to read about the different mapping methods before having an introduction to what happened and when.

Commented [PJ98]: This is implied

Commented [PJ99]: This sentence should be moved to the manual mapping section below. Maybe as the last sentence of the first paragraph in section 3.1?

Commented [PJ100]: Was this for Phase 2? Was only one mapper's map chosen?

Commented [PJ101]: I added a few sentences into the methods section to tie the mapping to Phase 1 vs Phase 2. I don't care if you use the exact language, but the two phases aren't mentioned in the main text of this report until section 4. It would also help to include a brief overview of the phases into the introduction section of this report.

Commented [PJ102]: Is this Phase 2?

3.1 Manual Mapping Methods

Four mappers manually delineated potential ~~bedrock hollow~~bedrock hollow and ~~inner gorge~~inner gorge RILs based on WA Forest Practice definitions (WAC 222-16-050) in Sites 1A, 2, 3, and 4~~. based on the criteria described in Section 1.3.~~ An additional mapper, with extensive field experience in the area, participated in an additional mapping exercise at Site 1B. The manually-mapped landforms were used to compare with computer-generated landforms in Phase 1 of this project.

~~The automated mapping algorithms operate entirely on the DTM representation of the terrain, so t~~The manual mappers based their interpretation solely on topographic and geomorphic characteristics observable in the DTM. We did not use field evidence and other remote sensing products. We visualized the terrain using 3-foot (1-meter) resolution lidar derivatives, including a multi-directional hillshade overlayed with 5-foot contours and colored slope classes (% gradient) using the following slope increments: 60% to 70%, 70% to 80%, 80% to 90% and >90%. The manual mappers agreed that ~~A~~at least half of the landform polygon needed to contain slope pixels with values greater than 70% to be considered a hollow or inner gorge, in addition to the Board Manual RIL definition. ~~A common, pre-defined channel head dataset was used (see Appendix B), and all mapping was completed at a scale of 1:500 for use in the secondary mapping effort in Site 1B.~~ The team used a common, pre-defined channel head dataset (see Appendix B) and completed all mapping at a scale of 1:500 for the secondary mapping effort in Site 1B.

3.2 OBIA via Image Segmentation and Classification (WY model) Methods

As a proof of concept for applying OBIA methods to delineate landforms, we opted to use an available model developed by Weyerhaeuser technical staff (herein referred to as WY OBIA; ~~see~~ Shaw et al., 2017). This model employs an object-based approach to classify landforms using a variety of image analysis tools that allow us to mimic the interpretation process an expert uses to ~~"see" RILs (i.e.,~~ visually identify and delineate landforms ~~in the field or remotely~~ with digital imagery~~)~~. Visual recognition of a landform is subjective and at times difficult to communicate, and it is even more difficult to translate into rulesets or computer scripts (Van Den Eeckhaut, 2012). We know, however, that experts can conceptualize RILs using only derivatives of high-resolution, bare-earth lidar data (e.g., hillshades, hillslope curvature, and gradient). We also know that landforms contain several physical features with significantly different morphometric characteristics. Therefore, our model is designed to delineate and classify landforms based on mathematical derivations of key morphometric parameters (e.g., surface elevation, gradient, curvature; see Table A1, Appendix A) that can be derived from DTMs using image analysis tools. The WY OBIA model has been applied regionally (in Washington and Oregon) across landscapes with diverse terrain characteristics because it uses only one ruleset with one set of parameters. The WY OBIA approach enables extraction of features without using any additional data (e.g., landslide inventories or values extracted from manual mapping products) to calibrate the model. Consequently, the WY OBIA model generates an independently derived suite of landforms that

Commented [PJ103]: Is this Phase 1?

Commented [PJ104]: Should this say "objectives?"

Commented [PJ105]: Did the manual mappers just use the DTM because that was what was used for automated mapping? Or is it because that's the best data to map IGs and BHs? (Is the first half of the first sentence needed?)

Commented [MM106]: Yellow- This isn't part of the RIL definition, so it would be good to be explicit about what was being mapped.

subsequently can be compared geospatially with landslide distributions, mapped landforms, and other types of thematic data.

The WY OBIA model takes advantage of the hierarchical, statistically based analytical capabilities of eCognition software tools to extract and classify features. Basic model steps include:

1. creating data layers used in the analysis, including lidar derivative grids (e.g., gradient and curvature components based on definitions found in Florinsky, 2016);
2. segmenting an image (e.g., of a curvature grid) into multi-scale objects by grouping or aggregating pixels according to specified threshold values (e.g., convergent vs divergent slope forms), as shown in left panel of Figure 8;
3. classifying those objects using statistical data values, geometry, or their relationship (e.g., adjacency) to other objects using the decision tree process described below, as shown in right panel of Figure 8; and,
4. assigning landform names to extracted objects based on RIL definitions and criteria; our classification scheme also includes non-RIL features (e.g., ridges) that ~~are~~were extracted first in order to isolate the RIL landforms considered for this sub-project (bedrock hollows and inner gorges).

Commented [PJ107]: Can the text that is after this list be directly tied back to these steps?

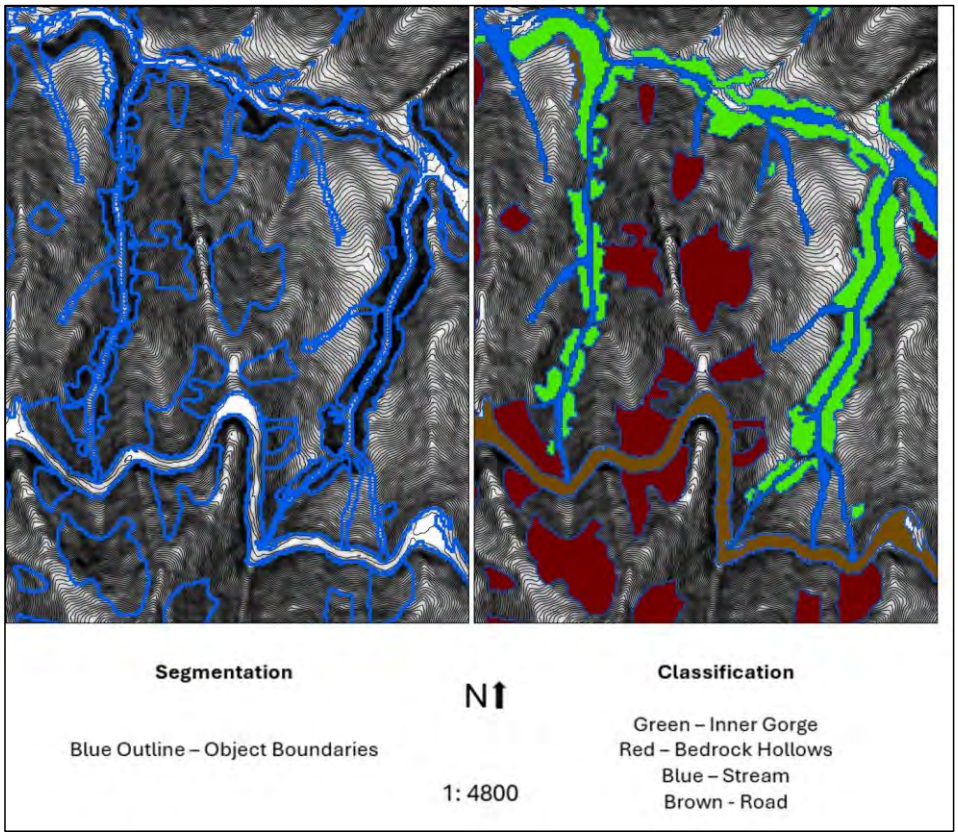


Figure 8: Fundamental image analysis processes; segmentation and classification at Site 1B. Hillshade shown as background.

We approached the RIL feature extraction process by building decision trees similar to what experts may use. For example, experts can determine when a channel-adjacent hillslope is being actively incised by using topographic criteria that include the presence of steep, very tightly confined hillslopes immediately adjacent to a channel, and evidence of a distinct gradient break that defines the upslope extent of the steep hillslope. However, the upslope gradient break is not always crisp or continuous, and channel-adjacent steep slopes are not always topographically uniform. An expert is forced to decide how and when to connect segments of steep slopes as one spatially continuous feature. Visual interpretation allows experts to demarcate these inner gorges by conceptually drawing a continuous line along the hillslope, even though the pattern of slope steepness and the slope break can vary significantly (Van Den Eeckhaut et al., 2012). The human brain is further able to distinguish features that are not inner gorges such as bands of bedrock outcrops that run perpendicular to the stream channel. Image processing techniques in

Commented [PJ108]: Can the following paragraphs in this section be simplified?

Commented [PJ109]: The following paragraphs still use present tense. Is this explaining what Weyerhauser does (does "we" refer to Weyco?), or is describing the methods that were employed for this project? If for this project, then it should be in past tense.

this workflow such as pixel segmentation, object classification, class relationship to neighbors, and use of multi-scale data to help identify discrete boundaries allow us to replicate the iterative process experts use when delineating landforms.

The main objective of the WY OBIA process is to design a classification scheme that identifies and integrates the geomorphic signatures of the landforms. Classification is accomplished by utilizing multiple image analysis tools including neighborhood relationships of pixels and objects. For example, classification of objects as 'Inner Gorge' is based on their steepness and immediate proximity to a stream. We also integrate processes where we classify seed clusters, or groups of pixels with homogeneous properties, that are significantly different from their neighbors (e.g., long, flat, and narrow objects like roads), and then grow them out into areas that aren't statistically different or are best described using data at a different scale (e.g., landings). We analyze data on hierarchical levels to evaluate membership to features extracted at different scales; for example, we are able to ~~evaluate~~model whether ~~if~~ streams are contained within tightly convergent hillslopes. We also consider geometric shapes of extracted features, most notably length and width relationships, to determine if an object is compact, as is the case with bedrock hollows, or if it is long and narrow like a tightly confined and steep headwater channel. We build the model using a sequential classification routine that begins by extracting features on the landscape that could either interfere with or aid in the classification of our target features, bedrock hollows and inner gorges. This approach is similar to how an expert would interpret, for instance, where the stream channel initiates, in order to define the lowermost extent of a bedrock hollow. Non-RIL features extracted in this model include large terraces, valley bottoms, landslide benches, streams, ridges, and roads. Once extracted, these features are not considered as candidates for future classifications.

The ruleset developed for this sub-project contains over 150 segmentation and classification algorithms to allow for analyses at multiple spatial scales. The numerous mathematical and statistical methods employed by eCognition are described in peer-reviewed literature and additional documentation maintained by Trimble. User input required to run the model is limited to parameter threshold values (e.g., 70% gradient thresholds for bedrock hollows) and length scales, or grid window sizes, over which a particular parameter is analyzed. These values, along with further description of the method, are provided in Appendix A.

3.3 Virtual Watershed Methods

In addition to the WY OBIA model, we also implemented a geomorphically oriented approach based on the concept of a "virtual watershed" (Barquin et al., 2015; Benda et al., 2016) implemented using NetMap tools (Benda et al., 2007). A virtual watershed enables geospatial simulation of watershed processes and human interactions (Figure 9) ~~provides a digital data structure that~~ by enumeratinges watershed landforms linked explicitly to a DTM-traced channel network (Miller et al., 2015). ~~to enable geospatial simulation of watershed processes and human interactions (Figure 9).~~ A virtual watershed is characterized by five analytical capabilities:

Commented [PJ110]: Confusing sentence

Commented [PJ111R110]: I still don't know what this sentence means.

1. Landform characterization: Every cell in a DTM is characterized by its geomorphometry (e.g., gradient, curvatures) and its topography (e.g., landscape position, landform type, and location relative to other landform types).
2. Discretization: Channels, floodplains, and hillslopes are discretized into facets of size appropriate for the analyses performed.
3. Attribution: All facets and channel segments are assigned the suite of attributes required for the analyses performed.
4. Connectivity: Flow paths are determined from every DTM cell to the basin outlet or a nondraining closed depression.
5. Routing: Ability to transfer information up and downstream over all channel and hillslope flow paths.

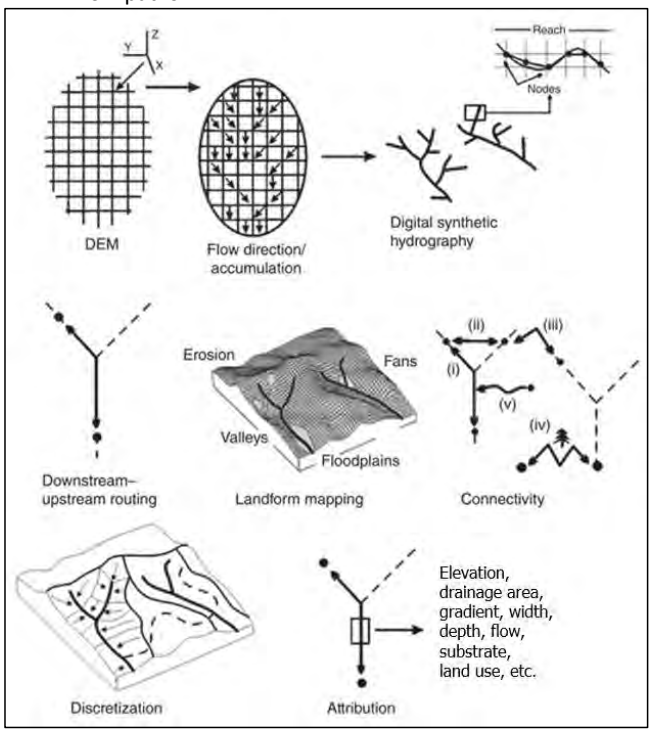


Figure 9: Steps in construction of a dataset for a virtual watershed. These data structures utilize multiple connectivity pathways (discussed in Appendix B), including: (i) river connected, (ii) Euclidean distance, (iii) slope distance, (iv) gravity-driven flow paths, and (v) modified slope distance.

Commented [PJ112]: If these are important, they should also be mentioned in the main text

We used the NetMap suite of tools (Benda et al., 2007) for creating and using a virtual watershed, with which we can delineate bedrock hollow and ~~inner-gorge~~inner gorge landforms using the RIL definitions in WAC 220-16-050 and the descriptions in the ~~Forest Practices~~ Board Manual~~,~~ Section 16). There are two approaches for delineating these landforms. One is to specify the landform

definitions *a priori* based on the Board-Manual descriptions. The other is to calibrate the definitions so that the delineated landform polygons best match those drawn by experienced field practitioners. We used both.

Commented [PJ113]: Were they in different phases?

Delineation of landform boundaries relies heavily on variations in the ground-surface gradient. Field measures of gradient are typically made with a clinometer sighting to a target looking either upslope or downslope. In vegetated terrain, it can be difficult to see a target much more than about 15 meters away. For the virtual watershed analysis, we measured gradient on the DEM in a similar fashion. Gradient at each DEM grid point was measured looking upslope 15 meters and again downslope 15 meters (slope distance), as described in Appendix E. Each DEM grid point thus had both an upslope-looking gradient value and a downslope-looking gradient value. Differences in these values were used to identify slope breaks.

The scheme for delineating hollow and ~~inner gorge~~inner gorge landforms described above and in Appendix B relies on a channel network traced from a DTM. Methods used are described in Miller et al. (2015). A key factor determining characteristics of the resulting network are the criteria used to set the upslope extent of the traced channels. For this case, because delineated hollows are limited to within those areas (the drainage wings, Appendix B Figure B~~3~~2) flowing directly into unchannelized flow paths, the traced network needs to extend beyond likely channel-head locations and into unchannelized swales.~~,~~ These unchannelized swales are referred to as~~the~~ zero-order basins.[3]~~-~~ A number of studies have looked for topographic indicators to differentiate erosional "process domains" (e.g., McCleary et al., 2011; Montgomery, 1999). These transitions are identified using inflections in slope-area plots thought to indicate the transition of planar and divergent hillslopes to convergent swales to debris-flow corridors to fluvial channels (e.g., Ijjasz-Vasquez and Bras, 1995; Stock and Dietrich, 2006) and can be used therefore to identify slope-area thresholds for tracing the upslope extent of a DTM-derived channel network.

Commented [JM114]: A map would be helpful.

Clarke et al. (2008) describe an approach for calibrating a contributing-area threshold, with optional slope dependence, using a log-log plot of channel density as a function of the threshold value. An inflection in the plot (Figure 3 in Clarke et al., 2008), where the rate of change of channel density starts to increase with decreasing threshold value, indicates the threshold below which "channels" are traced in profusion onto planar hillslopes, referred to as "feathering" of the traced network (Montgomery and Foufoula-Georgiou, 1993). Within the resolution of the DTM, this inflection indicates the contributing area at which planar and convex hillslopes transition into concave (topographically convergent) landforms (Montgomery and Foufoula-Georgiou, 1993).

[3] Channels may extend upslope into a zero-order basin after a debris flow scours soil down to bedrock. This channelized condition persists until accumulated soil and woody debris refill the channel. We assume ~~that,~~that most of the time, zero-order basins are unchannelized because the recurrence interval for debris flows is greater than the time for the scoured channel to refill.

Thus, the threshold identified with the inflection in the area versus channel-density plot provides a lower limit for tracing channels that extend through zero-order basins.

Using thresholds near or below that inflection point~~value~~ can, however, result in "feathering" of the traced channel network. To preclude feathering, we first identify all potential channel-initiation points for a specified threshold value, sort these from highest to lowest elevation (or optionally, from highest to lowest contributing area), and then, working from the highest to lowest points, remove all neighboring points within a specified radius. We set that radius to what we estimated as the lower cross-slope width limit for a hollow landform, which may vary with terrain (~~Add appendix citation~~Table B1, Appendix B1).

With this scheme, hollows are limited to those areas draining to unchannelized flow paths and inner gorges to those areas draining to channelized flow paths, so we still need to identify the transition from an unchannelized to channelized flow path. This transition occurs at the channel head, a key geomorphic feature for deciphering landscape development (Dietrich and Dunne, 1993; Wohl, 2018). Much effort has been expended in finding ways to identify channel-head locations using remotely sensed data (e.g., Anderson, 2019; Avcioglu et al., 2017; Clubb et al., 2014; Garrett and Wohl, 2017; Li et al., 2020; Metes et al., 2022; Płaczkowska et al., 2021; Shavers and Stanislawski, 2018; Shavers and Stanislawski, 2020; Wu et al., 2021; Zhang et al., 2021). ~~Such efforts are challenged by the transient nature of channel heads, which can migrate up and down slope depending on recent storm, fire, and debris flow activity (BM16, Dunne, 1991; Hattanji et al., 2021; Hyde et al., 2014; Wohl, 2013), Were channel head locations static over time, we may find that they are associated with specific topographic thresholds. so that~~Given their transient nature, channel-head locations are appropriately characterized using distributions of topographic attributes (Zhou et al., 2019).

We lack field surveys of channel-head locations with which to define such distributions, but we can specify that modeled channel heads need to occur at or downslope of the manually mapped bedrock hollows. Thus, the distribution of slope-area values along the traced flow paths where they intersect the downstream end of the manually drawn hollow polygons provides a basis for choosing a slope-area threshold value for channel heads.

With this scheme, RIL polygons are delineated using concepts familiar to field practitioners and consistent with Board-Manual descriptions. Hollows lie within steep swales draining to 1st-order channels and inner gorges on steep channel- or floodplain-adjacent slopes and extend upslope to a slope break. We can define the channel-head and slope thresholds *a priori*, however, in this study, we ~~or~~ adjusted their values to best match manually drawn ~~RIL~~ landform polygons delineated based on the criteria derived from RIL definitions. We ~~can also~~ adjusted the length scale over which gradient and curvatures are measured. Calibration of these thresholds, based on the frequency distribution of topographic attributes, is discussed in Section 4.2.2. For both

Commented [PJ115]: What value?

Commented [PJ116]: If the thresholds and radius values (or range of values) are important to landform delineation, then are the values reported somewhere in this document? If so, refer the reader to their location. If not, should they be added?

Commented [GS117]: This sentence needs additional clarification. First, I don't believe channel heads commonly do migrate up and downstream on timescales related to recent storm, fire or debris flow activity. And second, the second clause of the sentence (after the citations) does not follow from the first clause. It seems like the first clause is trying to demonstrate that channel head locations are *not* appropriately characterized using distributions of topographic attributes.

Commented [GS118]: This seems like circular reasoning. In the previous paragraph, you said that 'we still need to identify the transition from an unchannelized to channelized flow path' to distinguish between inner gorge and bedrock hollow landforms. If you are going to use human-mapped hollow polygons to define that transition, how is this an automated process? Yellow.

Commented [JM119R118]: Good question.

Commented [MM120]: Yellow- You developed a model that tries to replicate what a group of humans voted was a good looking representation of a landform. Neither of these are the actual RILs, so it is important to be explicit about what is being mapped, and how it relates to our definitions of RILs.

Commented [PJ121]: Is this what we CAN do or what was done? Please adjust language to make it clear about what was done vs what is recommended for future landform mapping.

automated methods, the ability to calibrate model outputs to better match observations will be used in the next sub-project, Susceptibility and Runout, to adjust model parameters to find the optimal landform criteria for distinguishing variations in landslide density.

4.0 Evaluation of Modeled Map Products~~:~~

~~***Evaluation Process***~~

For this project, the modeled landform maps aim to reasonably represent what an experienced practitioner (EP) would produce, but without field verification. The Project Team initially intent~~t~~ded to produce manual maps of landforms, based on RIL criteria, that could act as a reference map for the modeled landform maps to assess their ~~consistency~~accuracy. ~~To use the manually drawn landform maps as a reference for what a practitioner would produce, they need to be consistent across mappers.~~ The Project Team evaluated the landform maps produced by automated mapping methods in two phases. In Phase 1, we compared ~~them~~ the automated maps to landform maps manually produced by experienced practitioners (EP) and also to the maps of one EP ~~who was~~ selected by the Project Team through a consensus exercise. We also evaluated the variability among the EP's maps. We characterized the variability between mappers in a way that allowed comparison with the automated mapping products. We developed a sequence of consensus exercises for this sub-project to compare the EP's landform maps to each other and to the computer.

In Phase 2, the Project Team evaluated the landform maps produced by automated methods in a consensus exercise, referred to in this report as the Voting Exercise. The methods and results for each of these phases are described below.

4.1 Phase 1 – Model Evaluation via Manual Delineation Comparison

4.1.1 Phase 1 Methods

Phase 1 of this project aimed to assess the consistency between manually mapped landform maps and landform maps produced by the automated methods. The project team also evaluated the variability among experienced practitioner (EP) delineations of bedrock hollows and inner gorges by comparing independently mapped landform polygons across four study areas. To evaluate mapping consistency, the Project Team employed a multi-tiered approach that included pairwise comparisons of mapped polygons, voting-based consensus exercises, and spatial agreement analyses using confusion matrices. We implemented these methods to quantify the extent of agreement among mappers, identify sources of divergence in landform interpretation, and assess the alignment between EP delineations and computer-generated landform predictions. By examining overlap in mapped areas, topographic attribute distributions, and EP voting patterns, the team sought to isolate procedural differences, evaluate the effectiveness of defined mapping criteria, and improve consistency in future mapping exercises.

Commented [HJ122]: This Phase 1 objective statement does not line up with the sub-project objectives. The aim was not just to assess variability in experienced practitioners. That should have been secondary or tertiary. The primary aim was to develop models that can produce landform maps that reasonably represent what an experienced practitioner would produce. This opening objective statement should be rewritten to line up with the previous objective statements.

4.1.1.1 Comparison of EP mapped landforms

All four mappers produced ~~bedrock hollow~~bedrock hollow and ~~inner gorge~~inner gorge landform maps for each of the four study sites. We compared these maps using pair-wise overlays of the mapped landform polygons and image analysis techniques to measure the degree of agreement between each pair of mappers. These results motivated a closer look at the mapping techniques used by each EP.

At Site 1A, the Project Team~~we~~ used other methods in addition to the pairwise overlays to evaluate agreement among mappers. Instead of looking at the area of overlap between mapper's landform polygons, we looked at how often manual mappers did or did not delineate a ~~bedrock hollow~~bedrock hollow polygon in the same area. ~~in 47 randomly selected sites.~~ Because we wanted sites that could be delineated as bedrock hollows by the mappers, we randomly selected ~~These sites were identified using~~ 23 potential ~~bedrock hollow~~bedrock hollow polygons generated by the virtual watershed model and 24 potential ~~bedrock hollow~~bedrock hollow polygons generated by the OBIA model. We counted the number of times a mapper drew a polygon at a site that another mapper drew one, and how often a mapper did not draw a polygon at a site that another mapper also did not. We compared each mapper's polygons against each other's individually and counted the number of times they agreed.

As another way to identify the sources of variability amongst the EP-mappers' polygon sets, we conducted a spread-sheet comparison of manually mapped polygons within the same 47 randomly selected ~~bedrock hollow~~bedrock hollow areas in Site 1A. In the exercise, each mapper voted "yes" or "no" for each of the manually drawn ~~bedrock hollow~~bedrock hollow RIL polygons. A "yes" vote indicated that the mapper felt the polygon *could* accurately delineate and represent the bedrock hollow in that area. We used the results of this exercise to evaluate agreement among mappers in terms of "yes" or "no" votes at each of the areas. For example, if two mappers agreed that the site represented a ~~bedrock hollow~~bedrock hollow RIL, that would count as an agreement between those two at that site. We did these comparisons individually against all other mappers.

In an effort align the EP's mapping criteria, the Project Team filled out a spreadsheet shown in Appendix D. This exercise ensured the mappers were all using the same parameters (scale, planform curvature, inner gorge slope breaks etc.) specified within the rule definitions. The mappers also selected which mapper's polygon from the set that they felt *most* accurately delineated and represented the ~~bedrock hollow~~bedrock hollow RIL for each site. ~~The purpose of this exercise was for the Project Team to come to an agreement on which EP's landform map was most consistent with what they believed to be an adequate RIL delineation.~~ The Project Team conducted this exercise to agree on which EP's landform map most closely matched their expectations for an adequate RIL delineation. Team discussions enabled them to identify in what

Commented [GS123]: Why only at one site?

Commented [PJ124]: I'm a bit confused - From the first half of the paragraph I understand that Site 1A was unique because you randomly chose 47 computer-generated RIL sites to see if mappers also ID'd those as RILs. But the second half of the paragraph describes a comparison between mappers, not between the mappers and the computer.

Commented [PJ125]: Is this different from what's described in the first few sentences of the previous paragraph? Are these the same 47 sites?

Commented [HJ126]: Yellow - What is the statistical basis for this approach? Were these methods reviewed and/or advised by a statistician?

way the selected EP's tools and methods varied from the rest of the team. Using ~~this~~ information from these exercises, the team then re-calibrated their mapping methods and repeated the mapping exercise in a new area (Site 1B) in an effort to increase consistency among their landform maps.

We extended the visual inspection and voting exercise, using the same selected locations at Site 1A, to computer-generated ~~bedrock hollow~~bedrock hollow landform polygons. This provided a measure of agreement by the mapping team with the computer-model results and showed that the computer models were drawing potential bedrock hollows and inner gorges ~~high-hazard landforms~~ where none had been identified by any of the mappers. Each member of the mapping team then visually examined a set of such computer-generated "outlier" polygons, with no matching mapper polygons, and voted on whether they thought they represented an actual potential bedrock hollow or inner gorge~~high-hazard landform~~ or not.

Commented [PJ127]: Change to BH + IG throughout document

Commented [HJ128]: Bedrock hollows and inner gorges

Commented [HJ129]: Bedrock hollow or inner gorge?

Commented [JM130]: Was this field checked? If not will it be checked?

4.1.1.2 ~~Comparison between EP-mapped polygons~~

~~As described in Section 3.1.1, all manual mappers used the same approach to delineate RILs, utilizing the same DTM derivatives (a multi-directional hillshade of a 1-meter lidar DTM, 5-foot contour interval layer, slope class with colors based on percent ranges 60 to 70, 70 to 80, 80 to 90, and >90). Maps showing number of overlaps for each site are in Appendix C. In an effort to verify that we were using the same mapping criteria, the Project Team filled out a spreadsheet shown in Appendix D. Additionally, the Project Team conducted a consensus exercise to determine which EP's map most accurately represented what a practitioner would produce. The goal of this was for the team to identify any tools and methods used by the EP that differed from those used by the rest of the team so that they could re-calibrate their mapping methods and, ideally, increase consistency among the teams maps.~~

Commented [HJ131]: Depending on how you define "practitioner", most if not all of your EP maps were in fact "what a practitioner would produce". This statement should be clarified.

Commented [HJ132]: I believe all of this was already stated (?). Seems redundant.

Confusion matrices

Confusion matrices are tables used to visually represent the performance of a classification algorithm or model. Here we use confusion matrices not as a measure of model performance, but as a means of comparing one set of landform polygons to another; specifically, we can compare one EP-mapper's polygons to another EP-mapper's polygons or to the polygons generated by the computer models. To generate a confusion matrix, we overlay the two sets of polygons and measure four types of overlap,[4] described in Table ~~2~~3 below. We refer to one

[4] For the examples here, we measured the area of these four types of overlap as follows: 1. Rasterize the study-area polygon using the DTM as a template. 2. Rasterize both the reference and comparison RIL polygons, again using the DTM as a template. 3. Overlay the three rasters and go through pixel by pixel. Where there is a study-area pixel but no reference or comparison pixel, the pixel is a True Negative. Where there is a study-area pixel, a reference pixel, but no comparison pixel, the pixel is classified as a False Negative. Where there is a study-area pixel, a comparison pixel, but no reference pixel, the pixel is classified as a False Positive. Where there is a study-area pixel and both a

mapper as the "reference" and the other mapper, or the computer-model outputs, as the comparison.

Table 3. Confusion matrix term definitions as used in this project.

Type of overlap	Confusion-matrix term
The reference and comparison polygons overlap.	True Positive (TP)
The comparison has a polygon, but the reference does not.	False Positive (FP)
Neither the reference nor the comparison has a polygon.	True Negative (TN)
The comparison does not have a polygon, but the reference does.	False Negative (FN)

Note the following relationships:

- TP + FN = Total area of reference RIL polygons
- FP + TN = Total area of reference non-RIL zone
- TP + FP = Total area of comparison RIL polygons
- FN + TN = Total area of comparison non-RIL zone
- TP + FP + TN + FN = Total study area

A confusion matrix was~~is~~ constructed using th~~os~~e four values in ~~(~~Table 4~~3~~~~)~~. Summing across rows gives the area mapped in each class (RIL or Non-RIL) in the reference map; summing down columns gives the area mapped in each class in the comparison map. The tables were produced using only one landform type, either bedrock hollows or inner gorges.

Table 4. Confusion matrix terminology definitions as used in this study.

		Comparison Mapper (EP or Computer Model)	
		RIL	**Non-RIL**
Reference Mapper	**RIL**	True Positive (TP)	False Negative (FN)
	Non-RIL	False Positive (FP)	True Negative (TN)

A variety of single-valued metrics have been defined for summarizing information in a confusion matrix, each focusing on different aspects of the relationships between the four types of overlap. We use four metrics to summarize relationships between the reference and comparison sets of polygons: accuracy, balanced accuracy, recall, and precision.

Accuracy tells us what proportion of the total study area the reference and comparison mappers agreed on. It is defined as:

Commented [GS133]: These equations are not showing up for me, despite having tried in multiple browsers.

Commented [TH134R133]: Sharepoint can mess up formatting, will make sure final word and pdf versions are formatted properly

reference and comparison pixel, the pixel is classified as a True Positive. Each pixel has a horizontal surface area of ~1 square meter. The number of pixels in each of the four classes determines the area in each class.

where "Matched RIL" indicates the area of overlap between the reference and comparison RILs and "Matched NonRIL" indicates the area of overlap between the reference and comparison NonRILs. For our study sites, all mappers interpret most of the area mapped as NonRIL, in some cases, by more than 90%. This large difference in the area mapped as RIL and NonRIL renders this measure of accuracy less informative. Even if the comparison map included no RILs, it would still match the majority of the area mapped as a NonRIL in the reference map and the accuracy would be high.

A simple alternative measure of accuracy, referred to as "balanced accuracy" (Tharwat, 2020), provides a better measure of the agreement between the two classes. Balanced accuracy separates the RIL and NonRIL areas and treats agreement between the reference and comparison maps for each equally. It takes the proportion of the reference RIL area matched by the comparison map (TP/(TP+FN)) and the proportion of the reference NonRIL area matched by the comparison map (TN/(TN~~FP~~+FP~~TN~~)), adds them together, and divides by two:

$$Balanced\ Accuracy = \frac{\frac{TP}{TP+FN} + \frac{TN}{TN+FP}}{2} = \frac{\frac{Matched\ RIL}{Reference\ RIL} + \frac{Matched\ NonRIL}{Reference\ NonRIL}}{2}$$

For the large differences in RIL and NonRIL area found at our study sites, if the comparison map had zero RIL area, the balanced accuracy could still approach 50%. This provides a more informative measure than the nearly 100% value that accuracy indicates. We ~~will~~ report values of both accuracy and balanced accuracy in our comparisons among mappers and between mapper and computer.

Recall tells us the proportion of the reference RIL area matched by the comparison map. It is defined as:

$$Recall = \frac{TP}{TP + FN} = \frac{Overlap\ of\ Reference\ and\ Comparison\ RILs}{Total\ Reference\ RIL\ area}$$

In the context of the reference map, higher False Negative values will produce smaller recall values. Measures of recall prioritize the false negatives, which show up in the denominator.

Precision tells us the proportion of the comparison RIL area that matches the reference RILs. It is defined as:

In the context of the reference map, higher False Positive values will produce smaller precision values. Measures of precision prioritize the false positives.

Commented [GS135]: Are the confusion matrices also based on area of overlap, or just whether or not any overlap occurred across the entire polygon? How do you calculate area of overlap?

Commented [HJ136]: It appears that the Balanced Accuracy equation figure has an error and the first TN should be an FN.

Commented [HJ137]: It appears that the Precision equation figure has an error and "total computer" should read "total comparison".

For all these measures, values of 1.0 indicate a perfect match and values of zero indicate no overlap of the RIL area between the reference and comparison maps.

4.1.1.3 Elevation derivatives

We sought to identify sources ~~clues~~ for the poor spatial alignment by examining how the frequency distributions of topographic attributes, within the mapped RIL polygons, varied among mappers and field sites. ~~We also compared maps in terms of the ranges of topographic attributes that mappers used to delineate landform polygons.~~ These topographic attributes were calculated as derivatives of the elevation values in the DTMs, specifically gradient, curvature, and contributing area. We compared the maps, in terms of the topographic attribute ranges used by mappers, using frequency distributions. Similarities and differences in these frequency distributions show how consistent mappers (and computers) are among themselves and across study sites in applying the topography-based criteria for delineating ~~high-hazard landforms~~RILs. Detailed descriptions of the elevation derivatives can be found in Appendix E.

4.1.1.4 Voting: Consensus exercise among manually mapped polygons

We then focused specifically on Site 1A to identify differences in the mapping protocols used by individual mappers that could be sources of inconsistency. The mapping team visually inspected and voted on sets of randomly chosen ~~bedrock hollow~~bedrock hollow RIL polygons to identify the polygon features and mapping protocols that produced the set of polygons agreed by the team to be the best. With a more standardized set of protocols, the team then produced ~~bedrock hollow~~bedrock hollow and ~~inner gorge~~inner gorge RIL maps for an additional study area, Site 1B.

~~We used the voting exercise to identify potential sources of inconsistency among manual mappers.~~ In the voting exercise, mappers voted on the accuracy of each other's (and their own) ~~bedrock hollow~~bedrock hollow polygons for a randomly selected set of 47 potential ~~bedrock hollow~~bedrock hollow sites. These sites were identified by randomly selecting modeled potential bedrock hollow polygons, from each computer model, at Site 1A (Figure 108~~4.19~~). In these polygons, there could be one bedrock hollow, many bedrock hollows, or none at all. At each of the selected sites, each mapper examined the manually drawn polygons drawn by all mappers for that site as well as the computer-generated polygons. They voted "yes" or "no" on if they thought the computer-generated mapped polygon represented an actual bedrock hollow at the site and, if it did, then voted on which mapper's polygon (either a different mapper's or their own) provided the best indication of likely RIL shape and extent.

Commented [GS138]: This language belongs in a methods section, not results. Limit results section to pure presentation of results.

Commented [HB139]: Maybe add a reminder note here that "potential" may have no, one or multiple bedrock hollows

Commented [HB140]: Could not find. Is this Fig 18?

Commented [GS141]: Methods

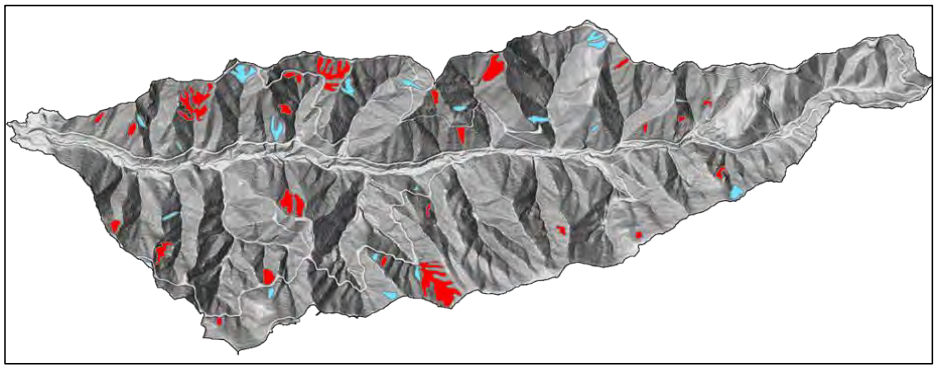


Figure 10: Location of randomly selected bedrock hollow areas mapped by OBIA (red) and VW (blue). See Figure 2.1A in Appendix A for site location (Site 1A).

~~Figure 18: Location of randomly selected bedrock hollow areas mapped by OBIA (red) and VW (blue). See Figure 2.1A in Appendix A for site location (Site 1A).~~

4.1.2 Phase 1 Results ~~– Comparison Among Mappers~~

In Phase 1, we implemented several methods to compare manual map products and evaluate consistency. These evaluations are summarized here and described in detail in Appendix C.

Using confusion-matrix derivatives of accuracy, balanced accuracy, recall and precision, we identified how closely the mapped polygons aligned spatially. We used maps showing the overlap of the ~~bedrock hollow~~bedrock hollow and ~~inner gorge~~inner gorge polygons drawn by all four mappers, at each of the four study sites, to construct a confusion matrix for each of the 12 possible pairs of mappers. From these confusion matrices, we calculated values for accuracy, balanced accuracy, recall, and precision for each pair of mappers (Appendix C). We found more variability in the manual maps than we had expected, which motivated us to look for sources of variability.

4.1.2.1 Manual Map Consistency at Site 1A

~~We focused on Site 1A with a visual inspection of individual bedrock hollow~~bedrock hollow ~~polygons in the hope of identifying polygon characteristics and mapping protocols used by the mappers (e.g., the mapping scale they use) that might help identify sources of variability among mappers. We chose Site 1A because one of the Project Team members, Julie Dieu, has extensive experience mapping landforms at this site, both remotely using methods similar to this Phase and with on-the-ground delineation of unstable terrain.~~

Commented [MM142]: It would be good to revisit the rule definitions of hollows and gorges here, so we can see the starting place for the mappers. Hopefully, it will be obvious to the reader that the reason for the lack of overlap is related to the vagueness of the definition. The results are exactly what one would expect- we often can't agree when we are on the ground.

Commented [GS143]: Why are the confusion matrix metrics (accuracy, etc.) not presented here?

Commented [GS144]: Delete. These types of sentences add clutter and do not push the narrative forward. Look for other similar opportunities to reduce word count by removing unnecessary language.

Commented [GS145]: This is redundant with the previous paragraph and should be moved to methods or deleted.

Shaded-relief images, showing the potential ~~RIL landform~~ bedrock hollow polygons drawn by each mapper of the four who completed the exercise (a fifth mapper completed part of the area and participated in the voting exercise described below), are shown for ~~bedrock hollow~~bedrock hollows in Figure 1~~0~~. Comparing the sets, the difference in size, shape and number of polygons is displayed between mappers. Figure 1~~1~~ overlays all four sets of polygons to show the number of mapper-polygon overlaps in an area. This shows the difference in delineation area between mappers. While the area of overlap among mappers is small, there are many areas in which a portion of their polygons overlap with each other – suggesting that they may agree that a bedrock hollow exists in that area, but don't agree on the delineation boundary. Figure 1~~2~~ quantifies these observations and shows differences in the number, sizes, and cumulative area of ~~bedrock hollow~~bedrock hollow RIL polygons drawn by each of the four mappers at Site 1A. Everyone agreed that a large portion of the area is NOT the ~~bedrock hollow~~bedrock hollow RIL. However, there was some disagreement as to where bedrock hollows do exist (Figures 1~~3~~ through 1~~7~~). ~~Oddly, the mapper with the largest cumulative area in polygons had the second lowest number of polygons.~~

Commented [PJ146]: It would be more useful if these statements told the reader what to take away from each figure (what's the point) not just describe what they are. Currently they read like a list of caption statements. (Same comment for each site)

Commented [PJ147]: This would be more informative if it said why it is odd. To me it just means they were a lumper, not a splitter. Also, which mapper?

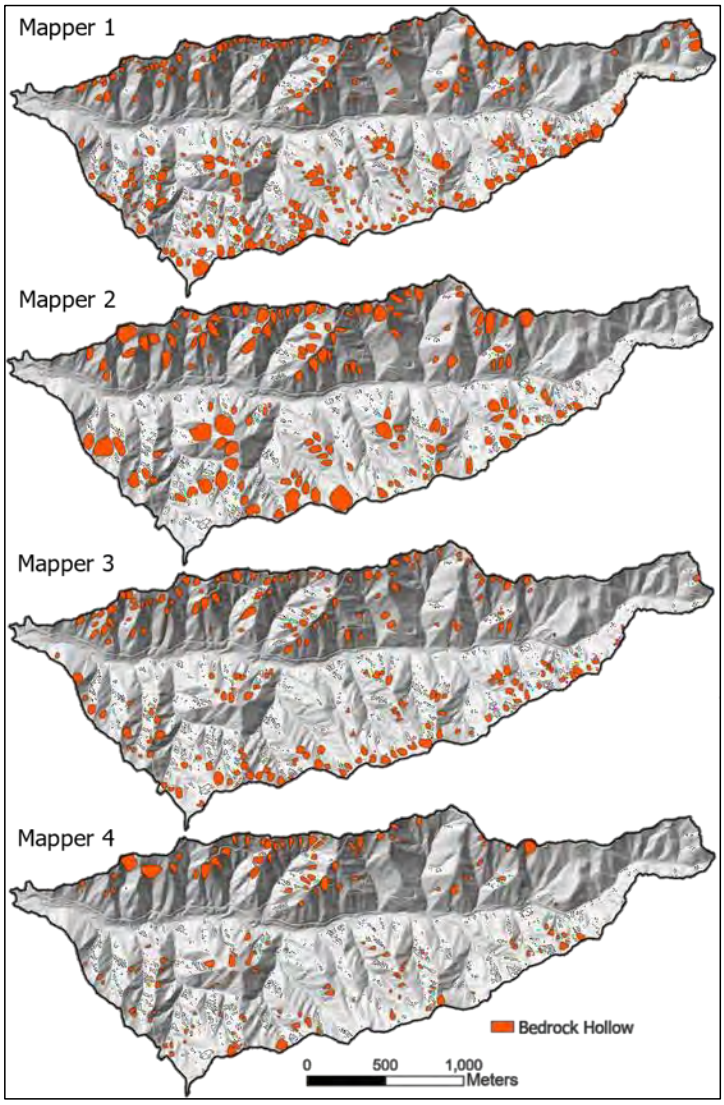


Figure 11: Each mapper's bedrock hollow RIL polygons overlain on a shaded relief image.

~~*Figure 10: Each mapper's bedrock hollow*~~*bedrock hollow RIL polygons overlain on a shaded relief image.*

Commented [JM148]: It would be helpful to see one bedrock hollow as mapped by the different mappers and compared to the models.

Commented [JM149]: It would be helpful to see one bedrock hollow as mapped by the different mappers and compared to the models.

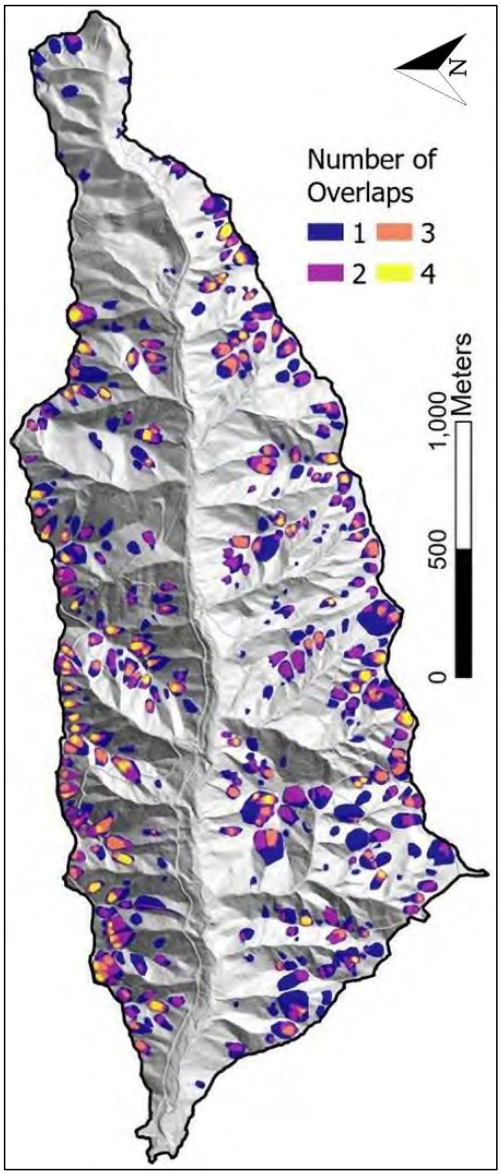


Figure 12: Number of overlaps among manual mappers.

Figure 1311: ~~Colors indicate the number of mapper's polygons that overlap in an area (top).~~ Left figure shows a bedrock hollow area with the polygons delineated by each of the models; the right figure shows the polygons delineated by the four mappers and the number of overlaps (bottom).

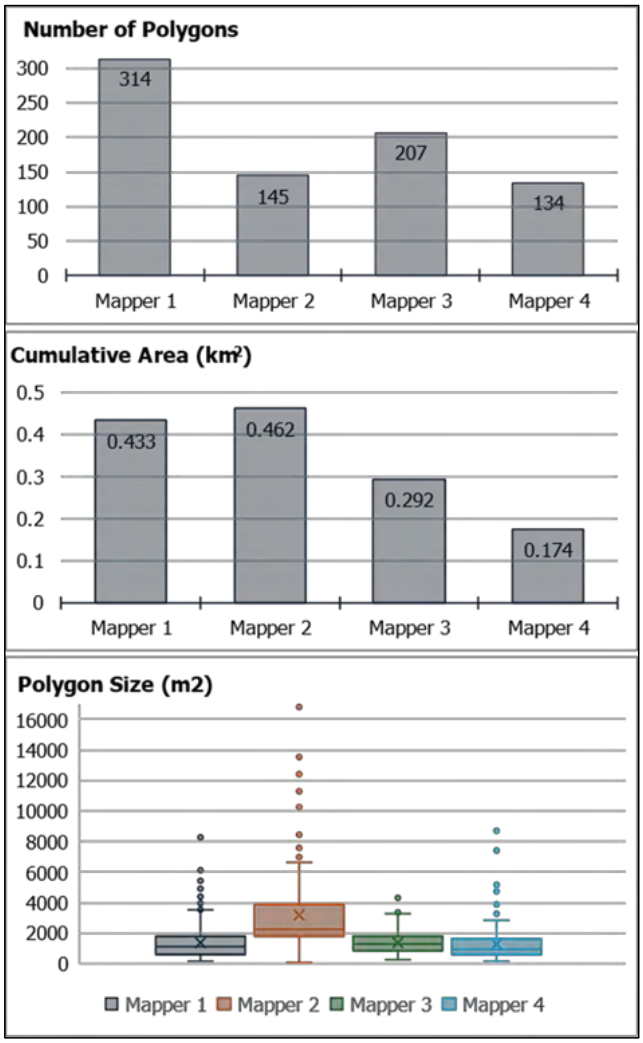


Figure 1412: The number, cumulative area, and size distributions of ~~bedrock hollow~~bedrock hollow RIL polygons drawn by four experienced practitioners for Site 1A.

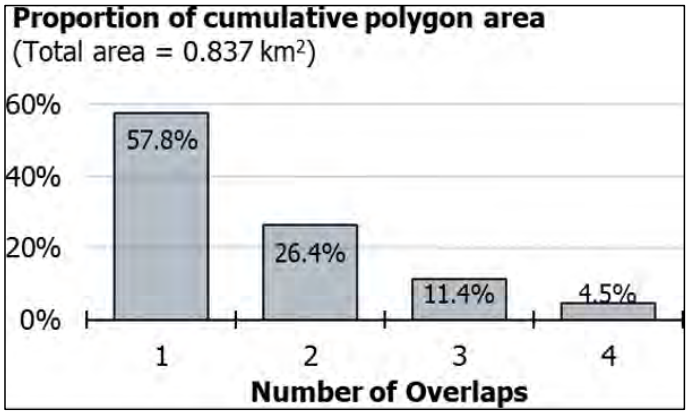


Figure 15~~13~~: The agreement across ~~all~~the four mappers, based on the number of overlaps across all four mappers' ~~bedrock hollow~~bedrock hollow RIL polygons in Site 1A. More than half (57.8%) of the total area encompassed by everyone's polygons included only one mapper; there was total agreement on less than 5% (4.5%) of the total area.

Error! Reference source not found.0Figure 16~~4~~ shows shaded relief images with ~~inner gorge~~inner gorge polygons mapped by each mapper. Figure 17~~5~~ overlays the four sets of polygons to show the number of overlaps in an area. Figure 18~~6~~ shows bar charts for the length of channel with adjacent mapped ~~inner gorge~~inner gorge RILs and the total polygon area for each mapper. For bedrock hollows, each polygon represented one hollow. Inner gorges are mapped adjacent to channels, so channel length is a better measure of how much inner gorge is mapped, rather than number of polygons. Figure 19~~7~~ shows the proportion of total cumulative area mapped as inner gorge RIL, combined over all four mappers, in terms of the number of mapper's polygons that overlapped. As with bedrock hollows, everyone agreed that a large portion of the basin is NOT the inner gorge landform. Generally, inner gorges are drawn adjacent to the streams. But there was unanimous agreement on less than 3% of the cumulative mapped area.

Commented [HJ150]: I'm used to differences in opinion among geologists, which is normal and healthy, but the low percentages of general consensus here, especially on inner gorges, indicates some kind of problem. I can see why the team tried to refine their approach for study area 1 b and then went to Phase 2.

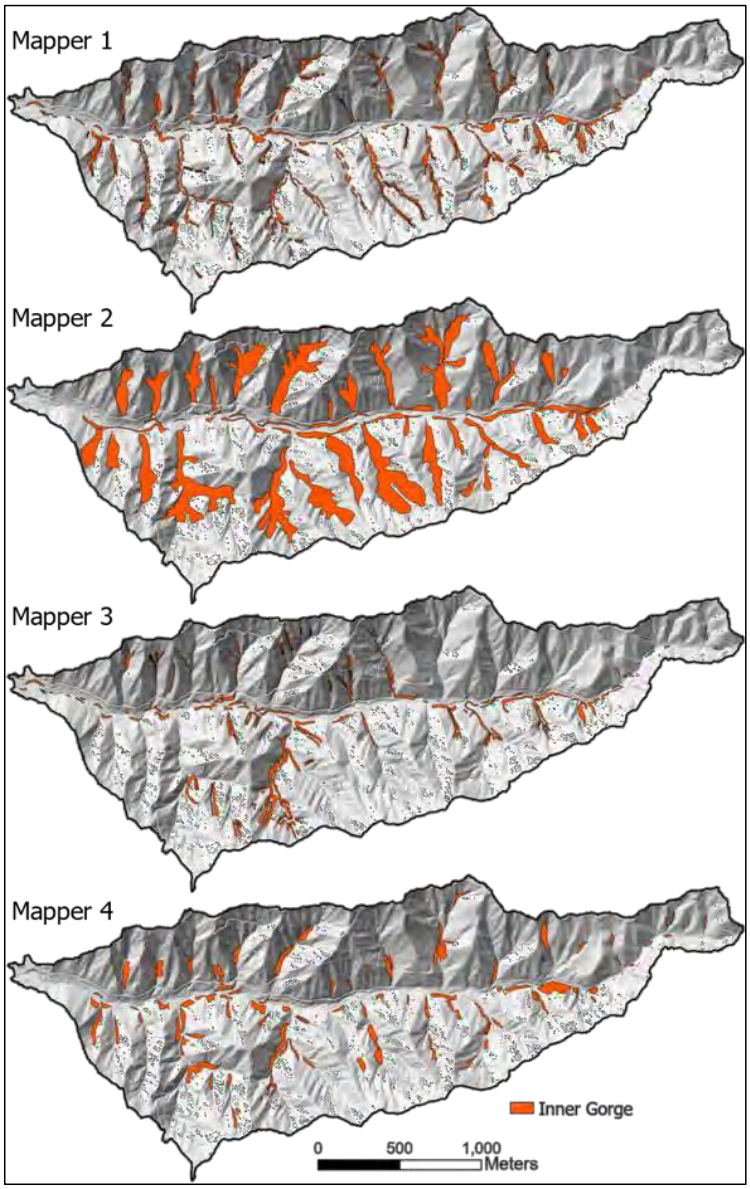


Figure 1614: ~~Inner gorge~~Inner gorge RIL polygons drawn by each of the four practitioners for the Site 1A.

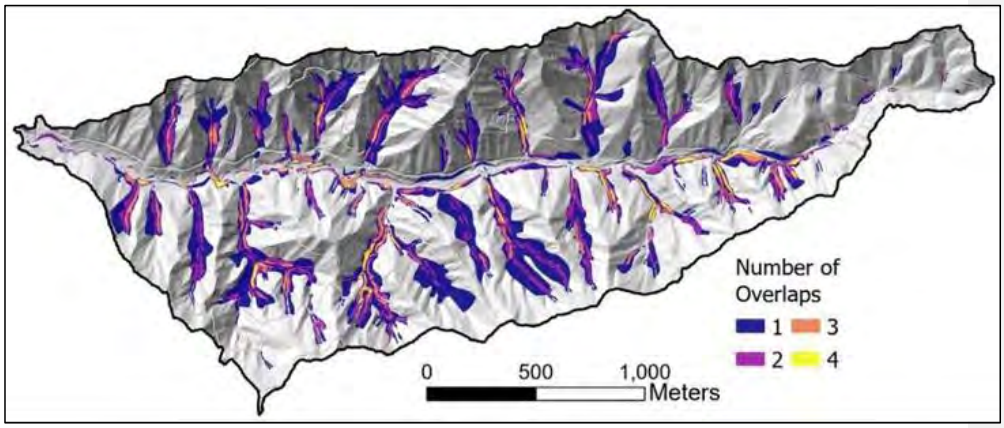


Figure 1715: ~~Inner gorge~~Inner gorge RIL polygons for all four mappers overlain. Colors indicate number of mappers' polygons that overlap in any location.

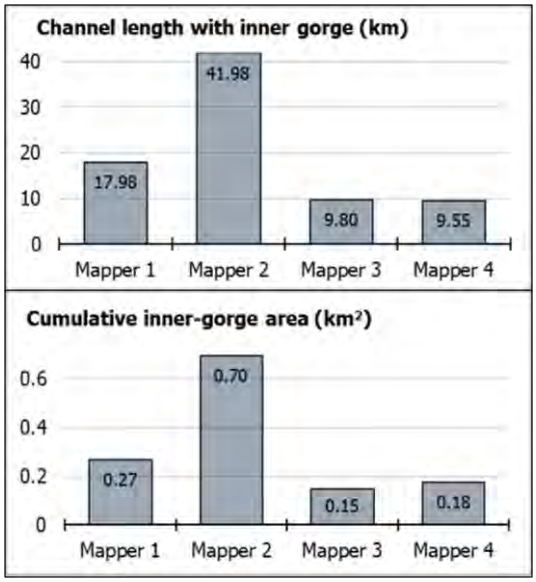


Figure 1816: Cumulative channel length with mapped inner gorge and cumulative mapped ~~inner gorge~~inner gorge area for each mapper in Site 1A. Channel length was measured for each side of a channel, so where inner gorges were mapped on both sides, that channel length counted twice.

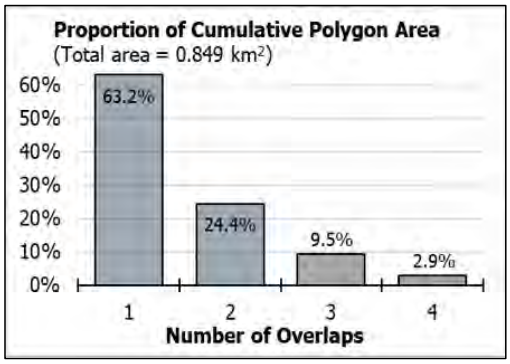


Figure 1917: Consistency across all mappers, based on the number of overlaps across all mapper's ~~inner gorge~~inner gorge RIL polygons in Site 1A.

4.1.2.2 Voting: Agreement among Mappers

The number of manually mapped polygon overlaps ~~with corresponding polygons~~ varied site by site. Five of the ~~potential~~ modeled bedrock hollow polygon sites had no coincident polygons drawn by any of the mappers. Some of the selected sites contained multiple ~~bedrock hollow~~bedrock hollow landforms delineated by the mappers. ~~If a mapper had multiple polygons for a single site, that was counted as a single site match~~ In cases where a mapper delineated multiple polygons within a single site area, we counted it as one site match~~.~~ The number of site matches for each mapper, along with the number of "best polygon" votes their polygons received, are listed in Table 45.

Table 5. The first column gives the total number of sites in which a mapper drew a bedrock hollow polygon. The second column gives the number of times a mapper received the majority "best polygon" votes for an area by all mappers. The third column gives the proportion of a mapper's bedrock hollow polygons voted "best."

	Number of matching polygons (out of the 47 sites)	Number of matching polygons voted "Best Polygon"	Proportion of matching polygons voted "Best Polygon"
Mapper 1	35	23	66%
Mapper 2	27	7	26%
Mapper 3	26	6	23%
Mapper 4	18	5	28%
Mapper 5	16*	5	31%

Mapper 5 provided mapped RIL polygons for only one third of the study site.

Mapper 1 had the most matching polygons (a "match" meaning they drew an RIL polygon at the selected ~~bedrock hollow~~bedrock hollow site), suggesting they had the most complete map in terms of not missing potential RILs, and the largest proportion of their polygons selected as the

Commented [PJ151]: I don't understand this sentence. I believe it is supposed to be the number of manually mapped polygons within the computer-generated polygons varied by site. Yes?

Commented [HB152]: OBIA or VW?
Does this indicate either are mapping areas too large relative to field experience of the mappers?

Commented [HJ153]: Does Table 4 refer to Study Area 1 A? I thought mapper 5 only mapped at Study Area 1 B.

Commented [HJ154]: Yellow - Shouldn't the second column add to 42? Or at least 47? Why 46? And shouldn't the third column sum to 100%? If not, this table needs to be better explained.

best. ~~We were also interested in the degree of agreement among the mappers as to whose map was the best.~~ We also wanted to assess how much the mappers agreed on whose map was the best. Out of all the sampled sites, 42 had matching mapper polygons. Of these, only 4 received a unanimous vote for the best polygon: three associated with Mapper 1 and one associated with Mapper 2. For the remaining 38 sites, more than one mapper's polygon received the "best polygon" vote and the choice of whose polygon was best depended on who was voting. This is illustrated in Figure 20~~19~~, which shows how the proportion of each mappers' matching polygons (which varied among mappers, see Table 5) receiving "best polygon" votes varied depending on who was voting.

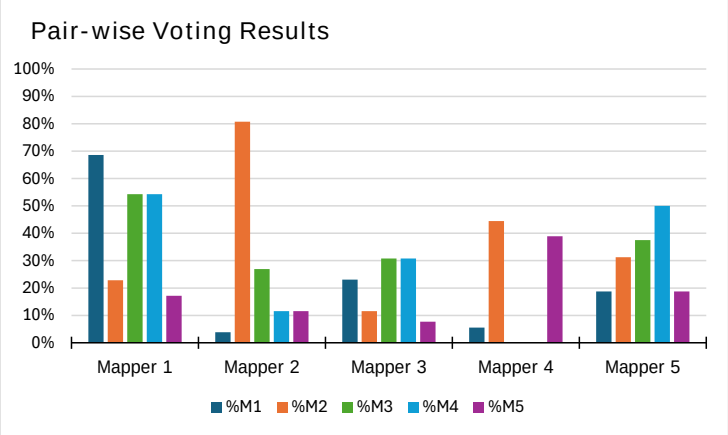


Figure 20~~19~~: Pair-wise voting results for which mapper had the most accurate ~~bedrock hollow~~bedrock hollow RIL polygons. The x axis shows the mapper voting and the y axis shows what proportion of each mapper's matching polygons they gave a "best" vote. Ex. 23% of the 35 matching bedrock hollow polygons drawn by Mapper 2 were voted "best" by Mapper 1.

Comparing the mappers' polygon sets against each other further elucidated differences between mappers and helped in identifying how those differences arose. Mappers identified d~~D~~ifferences in the mapping scales used and channel-head placements ~~were identified~~ as primary sources of variability. The results of the voting exercise determined that Mapper 1 had the most accurate set of landform polygons, based on the Project Team's opinion. Team discussion determined map scale to be the primary difference in method between the landform maps produced by Mapper 1 and the rest of the Project Team ~~was the scale used~~. Mapper 1 used a scale of 1:500 while the rest of the team used scales closer to 1:700-1:1000. To determine if map consistency could be improved with additional data and criteria, we decided to use a single channel-head dataset and a scale of 1:500 for mapping at an additional site, Site 1B.

4.1.2.3 C0omparison of Site 1B after mapper alignment

Commented [HJ155]: Yellow - Shouldn't the y-axis values of all five columns for each mapper on the x-axis add up to 100%? If not, this needs to be better explained.

Commented [HB156]: Looking at Figures 15 & 16, it looks like the polygon widths (perpendicular to the channel aspect), is quite variable also.

~~After identifying inconsistencies in mapping protocol, the manual mapping exercise was~~We repeated the manual mapping exercise in another study site (Site 1B), after identifying inconsistencies in mapping protocol, to see if ~~greater consistency between mappers resulted if everyone~~ mapping~~ed~~ at the same scale using the same GIS data files and the same estimates of channel-head location resulted in greater consistency between mappers. Site 1B is the small basin just north of Fahnestock Creek Basin (Site 1A) shown in Figure 2.1A in Appendix A. Site 1B has a similar geology and topography as Site 1A. Each mapper used the same DTM derivatives (a multi-directional hillshade of a 1-meter lidar DTM, 5-foot contour interval layer, slope class with colors based on percent ranges 60 to 70, 70 to 80, 80 to 90, and >90) as the previous exercises, but now utilized a common, pre-defined channel-head dataset, produced by the virtual watershed model, and a scale of 1:500 in addition to the DTM derivatives.

~~We had a~~ The fifth mapper who completed part of the exercise for Site 1A mapped ~~for~~ bedrock hollows at this site; they ~~, who~~ had extensive experience mapping landforms for this area, including on-the-ground delineation of RIL boundaries and many observations of landslides in this terrain. Their interpretation of the information available from the DTM was influenced by their experience at this site. Their ~~bedrock hollow~~bedrock hollow RIL polygons tended to be larger than everyone else's, reflecting what was actually deleted from a harvest proposal (highlighting the differences between remote mapping and actual field verification).~~.~~

Figure 2~~0~~1 shows each mapper's ~~bedrock hollow~~bedrock hollow and ~~inner gorge~~inner gorge RIL polygons overlain on a shaded relief image, similar to those shown in Figure 10. Figure 2~~1~~2 shows the number of polygon overlaps, similar to those shown in Figure 11.

Figure 2~~2~~3 shows balanced accuracy, recall, and precision for a one-to-one comparison of each mapper's ~~bedrock hollow~~bedrock hollow RILs for both Site 1A and Site 1B. The results of averaging these measures over all values for combinations of Mappers 1 through 4 at each site are shown in Table 6~~5~~.

Table 6. Averaged balanced accuracy and recall and precision for mapped bedrock hollows at Site 1A and Site 1B.

Bedrock Hollows	Site 1A Mean Values (± 1 standard deviation)	Site 1B Mean Values (± 1 standard deviation)	Difference (± 95% confidence using t-distribution)
Balanced Accuracy	0.666 (± 0.062)	0.663 (± 0.083)	0.003 (± 0.114)
Recall and Precision	0.389 (± 0.135)	0.373 (± 0.097)	0.016 (± 0.183)

Even with great effort to improve consistency among mappers, these measures of spatial agreement did not appear to have improved; in fact, they are very slightly smaller at Site 1B. However, our sample was small and variable, so these data do not preclude the possibility of either a plus or minus difference between the sites. A larger set of mappers could tighten the

Commented [HB157]: Will this translate to the OBIA model?

Commented [HJ158]: Yellow - Explain what the fifth mapper was supposed to add to the study. It seems to have confounded the study with partial mapping of the area and/or features using a different approach based on laying out buffers, not mapping landforms.

Commented [HJ159]: It sounds like mapper 5 was mapping the leave areas they prescribed and not the actual landforms. This seems to confound the study in area 1 b.

Commented [HB160]: Did these include the evaluation or confirmation of a DNR Forest Practices geologist or consultant. Regardless. Was that person or consultant being overly cautious because of the need for and approved FPA.

confidence intervals. Likewise, a different set of four mappers may have yielded different averages.

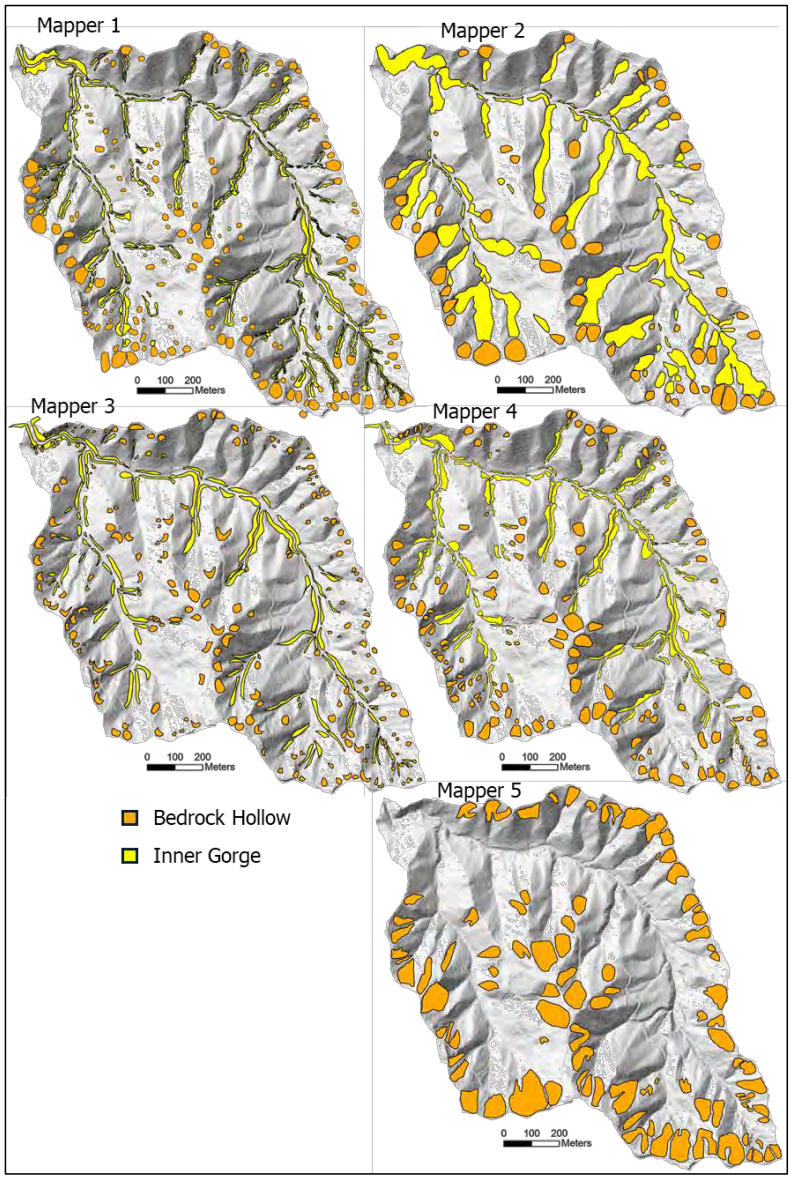


Figure 2120: Each mappers' ~~bedrock hollow~~bedrock hollow and ~~inner gorge~~inner gorge RIL polygons for Site 1B. Mapper 5 did not map inner gorges.

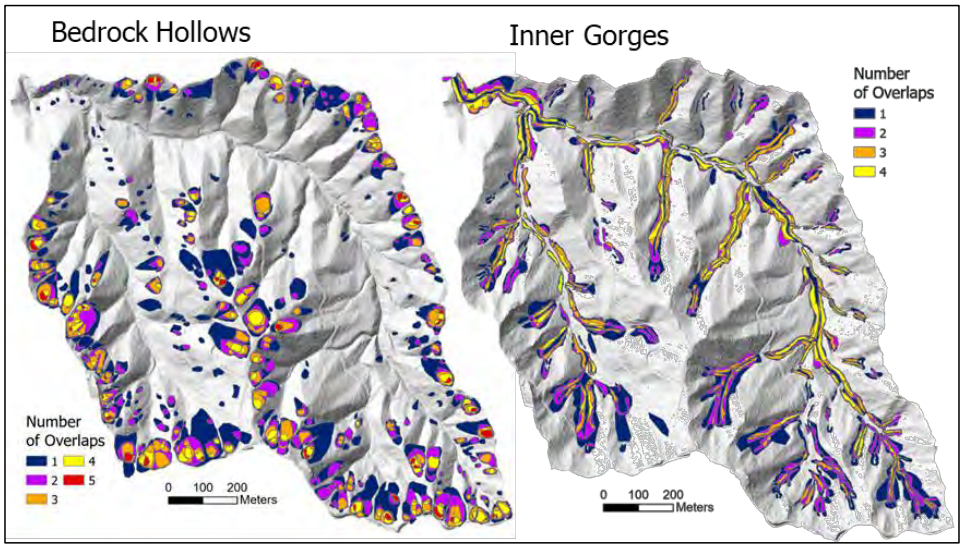


Figure 2221: Overlap count of manually mapped polygons in Site 1B. Each color represents the number of polygons that overlap in that area.

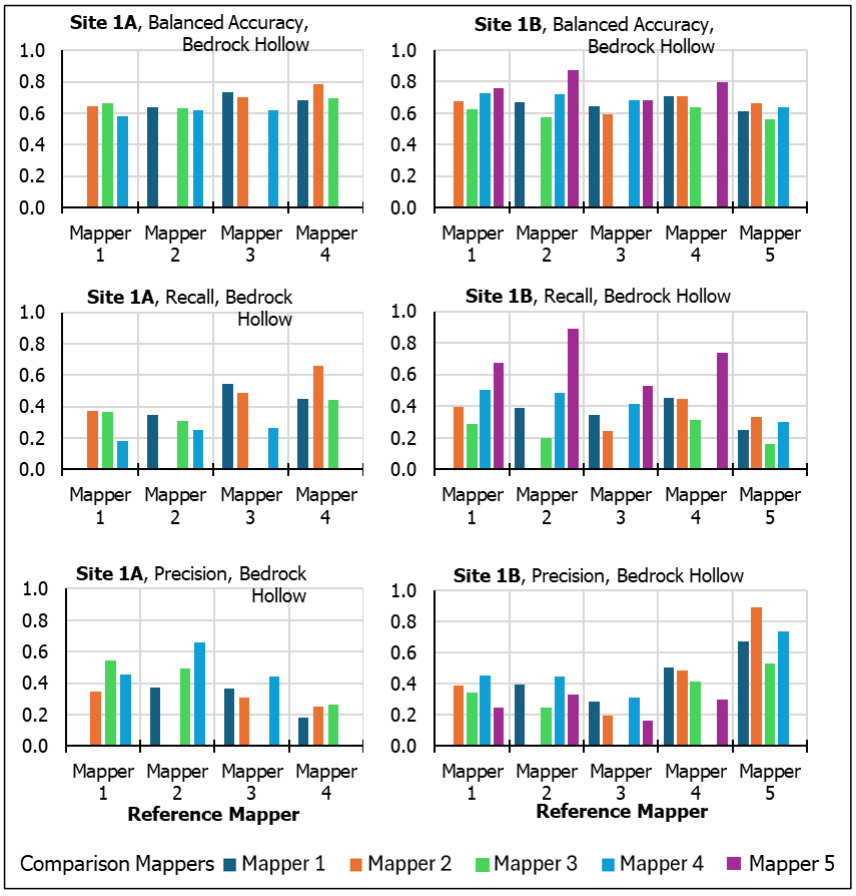


Figure 2322: Balanced accuracy, recall, and precision for manually mapped bedrock hollows at Site 1A and 1B.

Bar charts of balanced accuracy, recall, and precision for mapped ~~inner gorge~~inner gorge RILs at Sites 1A and 1B are shown in Figure 24. The averages over all combinations of mappers are shown in Table 76.

Table 7. Averaged balanced accuracy and recall and precision for mapped ~~inner gorge~~inner gorges at Sites 1A and 1B.

Inner Gorges	Site 1A Mean Values (± 1 standard deviation)	Site 1B Mean Values (± 1 standard deviation)	Difference (± 95% confidence using t-distribution)
Balanced Accuracy	0.665 (± 0.074)	0.745 (± 0.065)	-0.081 (± 0.108)
Recall and Precision	0.386 (± 0.192)	0.549 (± 0.164)	-0.163 (± 0.278)

There was an apparent improvement at Site 1B, although again, because of the small and variable sample, we cannot preclude the possibility (with 95% confidence) of an opposite trend. Interestingly, at Site 1A, the average values were almost the same between the bedrock hollow maps and the inner gorge maps. At Site 1B, the values for bedrock hollows remained about the same, but the values for inner gorges increased, indicating a greater degree of spatial alignment among mappers.

Commented [HJ161]: Is it right to compare 1A and 1B when you had 4 mappers at 1A and five at 1B?

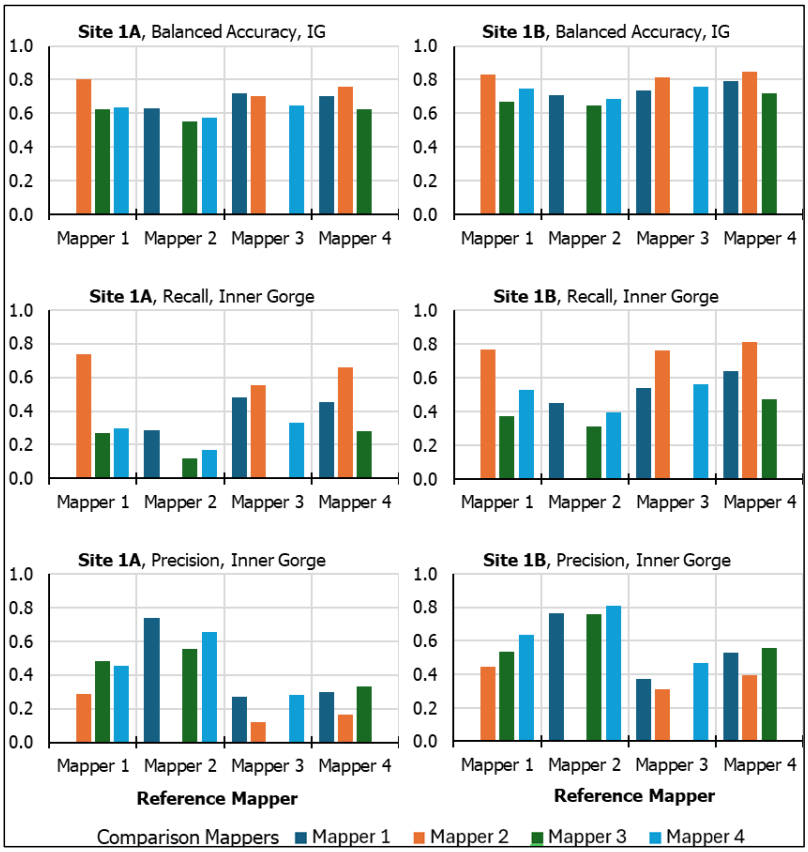


***Figure 2423*: Balanced accuracy, recall, and precision for manually mapped inner gorges at Sites 1A and 1B.**

Sites 1A and 1B are adjacent to each other and have essentially the same terrain features. ~~What does it tell us that ensuring a consistent mapping protocol among mappers had essentially no impact on the consistency of the mapped RIL polygons? Perhaps t~~The little improvememt in

Commented [MM162]: There is no right answer here for comparison- all of the mappers were following the definitions they were given and were all essentially correct in their remote mapping.

Commented [HJ163]: Yellow - This is a poorly worded and unsubstantiated comment as acknowledged in the previous paragraph.

consistency after revising our mapping protocol~~is simply reinforcess~~would appear to support the idea that, even for the same terrain using the same data and mapping protocols, each mapper interprets the information available slightly differently based on their experience and understanding of the processes involved with landsliding. We cannot enforce consistency for those factors, although a certain level of training and supervised experience of practitioners is required. We are limited to remote methods for this sub-project, but it is possible that field verification would improve consistency. It is also possible, that another iteration of map might have led to improved results.

Commented [PJ164]: It is also possible that field verification would improve agreement.

4.1.2.4 Uncertainties

In Section 4.1.1, we evaluated mapper agreement by area, looking at the amount of overlap the mapper's landform polygons had with each other. ~~Another way that W~~we also evaluated agreement among mappers by ~~was to~~ seeing how often mappers both did or did not delineate a bedrock hollow polygon in that site. Figure 2~~5~~4 shows the proportion of each mapper's polygon set that matched each other mapper's polygons. For example, of the 47 bedrock hollow sites, Mapper 1 and Mapper 2 both mapped ~~bedrock hollow~~bedrock hollow landforms at 47% of those sites.

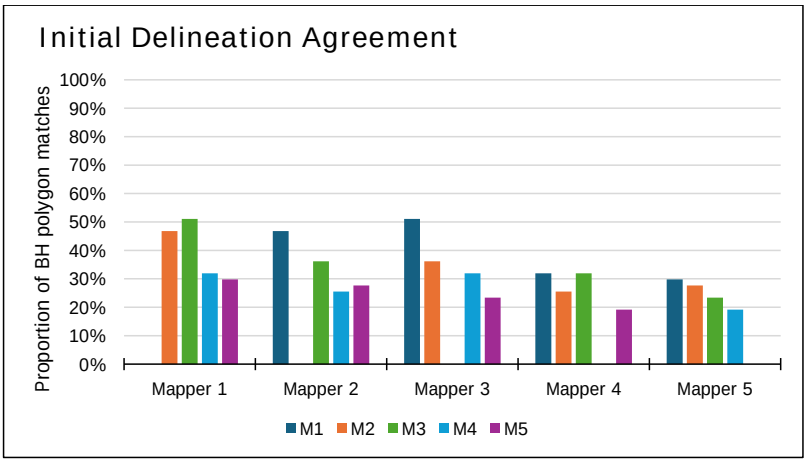


Figure 2~~5~~4: Proportion of agreement between mappers' initial bedrock hollow delineations at Site 1A. The height of the bars shows the percentage of ~~bedrock hollow~~bedrock hollow polygons drawn by each of the other mappers that matched the polygons drawn by the mapper on the x axis.

As discussed in Section 3.2.1, some rule definitions and Board Manual criteria for defining RILs lack quantifiable criteria and leave room for individual interpretations among mappers, which led to uncertainties in their initial delineations. To describe some of these uncertainties, we will use "Mapper A" and "Mapper B" to represent any two EP mappers on the Project Team. In some

cases, Mapper A delineated a ~~bedrock-hollow~~bedrock hollow polygon at a site that Mapper B did not – however, after Mapper A explained their interpretation of the RIL criteria, Mapper B agreed with the delineation and determined they should have drawn a polygon at that site (NonRIL-to-RIL in Table ~~7~~8). This case also occurred vice versa, where Mapper B explained their interpretation of the criteria and Mapper A, then, determined that they should not have drawn a polygon there (RIL-to-NonRIL in Table 8~~7~~). We wanted to look at proportion of uncertainty that the manual mappers had in their initial delineation of bedrock hollow landforms. We did this by looking at the number of both NonRIL-to-RIL and RIL-to-NonRIL cases. These values are shown in Table 7, as well as the "proportion of RIL uncertainty," calculated by dividing the total number of areas (47) in which a mapper did or did not delineate a bedrock hollow by the number of those sites in which they changed their mind.

~~It is clear from the results of the voting exercise, where most mappers agree that most delineated polygons are probable bedrock hollows, that each individual mapper is missing some landforms as they map. This is not a surprising result –~~ field layout~~engineering~~ of harvest units found many more unstable landforms than were originally mapped by the watershed analysis practitioners. We might hope that modern tools, such as lidar derivatives, improve mapping (and ~~perhaps they d~~they likely do). We acknowledge that mapping ALL the RILs for a large area is a lengthy and mind-numbing exercise, and we suspect that there may be subjective interpretation of the criteria for mapping RILs in the details of the percentage of area exceeding 70% and in the degree of curvature which is not part of narrative description.

Table 8: Values demonstrating the number and proportion of sites (n=47) in which mappers switched from RIL to non-RIL and vise-versa. Described in detail in text.

Mapper	M1	M2	M3	M4	M5
RIL-to-NonRIL	2	2	0	2	2
NonRIL-to-RIL	4	11	16	16	18
# of sites in which a mapper ~~did not~~ initially dr~~e~~aw a polygon	35	26	26	18	16
# of sites in which the mapper did not initially dr~~a~~ew a polygon	12	21	21	29	31
Proportion of sites initially mapped as RIL switched to NonRIL (#switched/#mapped)	6%	8%	0%	11%	13%
Proportion of sites initially mapped as NonRIL switched to RIL (#switched/#mapped)	33%	52%	81%	55%	58%

Table 9~~8~~ shows both the number of sites identified by each mapper as an RIL and the number of their sites that other mappers also identified as an RIL. When looking at the proportion of

Commented [JM165]: Interesting.

Commented [MM166]: There clearly is, based on this work, but that is because the the definitions are subjective, as is noted. I suspect that this is intentional to allow professions to make those calls in the field based on the site conditions.

Commented [HJ167]: Yellow - Double check that the values in Table 7 line up with the values in Table 4.

matches with another mapper, Mapper 1 appears to have the lowest, only 89% – but they also had the highest number of sites in which they drew a polygon. When looking at the proportion of total sites in which mappers drew a polygon that matched another mapper's, Mapper 1 has the highest at 66% (and was voted best mapper, so this is unsurprising).

Table 9: Bedrock hollow polygon matches at Site 1A.

	# Sites with BH polygon drawn	# Sites with BH Polygon Matches by other mappers	Proportion of polygons matched (#sites with BH poly/#sites with BH matches)	Proportion of Sites with BH poly matches (#sites with BH matches/#total sites)
M1	35	31	89%	66%
M2	26	24	92%	51%
M3	26	26	100%	55%
M4	18	18	100%	38%
M5	16	15	94%	32%

Commented [HJ168]: Yellow - Double check that the values in Table 8 line up with the values in Table 7.

From these series of exercises, we noted several observations:

1. Although one mapper received the majority of "best" votes, each mapper was voted the "best" for some proportion of the polygons (Table 545).
2. Mappers generally chose their own polygon as "best," but not always.
3. Only 4 out of the 42 sites with matching mapper polygons got unanimous votes as to which mapper's ~~was~~polygons ~~–~~best represented a typical EP delineation. There was disagreement for the remaining 38.
4. In this project, ~~V~~variability in individual interpretation of RIL criteria leads to cases of uncertainty in initial delineation of landform polygons (RIL-to-NonRIL and NonRIL-to-RIL cases).

Commented [HB169]: Is this the correct table number?

From these we draw three conclusions:

1. The mappers in this project had~~ve~~ inconsistent opinions regarding what a ~~bedrock hollow~~bedrock hollow RIL looks like using only remote data, as indicated by the variable bar heights in Figure 245.

2. Uncertain initial delineation of boundaries lead to mappers potentially including some polygons that are not RILs and missing some polygons that are.
3. To reduce mapping inconsistencies, the mapping criteria would need to be defined with greater clarity and minimized ambiguity to limit subjective interpretation. We did not evaluate other landform classes, or any landform classes with more explicit GIS-only-based geomorphic criteria, nor did we include field observations of bedrock hollows and inner gorges using field indicators as described in the Board Manual. Future projects will help inform if and what changes need to be made to the rule criteria.

4.1.2.5 Agreement between Mappers and Models

Agreement between manually mapped and computer mapped landforms is evaluated by comparing distributions of specific topographic characteristics within the landforms and via a voting exercise similar to that described in Section 4.1.2. Topographic metrics used to compare manually mapped and computer mapped landforms include the area of the landform, gradient, gradient ratio (upslope-looking gradient / downslope-looking gradient), tangential curvature and the logarithm (base 10) of the contributing area (m^2). Distributions of both manually mapped and computer mapped metrics are illustrated via box and whisker plots in Figures 2~~5~~6 through Figure 30~~29~~ and as cumulative distribution function (CDF) plots in Appendix F and G.

~~Comparison of the topographic characteristics illustrate: (1) the previous finding that practitionerEP-mapped landforms can be highly variable and (2) that, although the model mapped landform polygons may not exactly match any experienced practitioner's polygons in terms of area, they are within the variability of the manual mapped polygons in terms of topography.~~

Commented [HB170]: Possible? Suggestions? Will the later study phases indicate ways to improve mapping?

Commented [JM171]: Would field observation have improved mapping?

Commented [PJ172]: What is the difference between the gradient upslope vs downslope? Is this above and below the polygon? Or within the polygon?

Commented [HJ173]: Yellow - Is this a statistically valid statement with only 4 or 5 mappers out of dozens of practitioners?

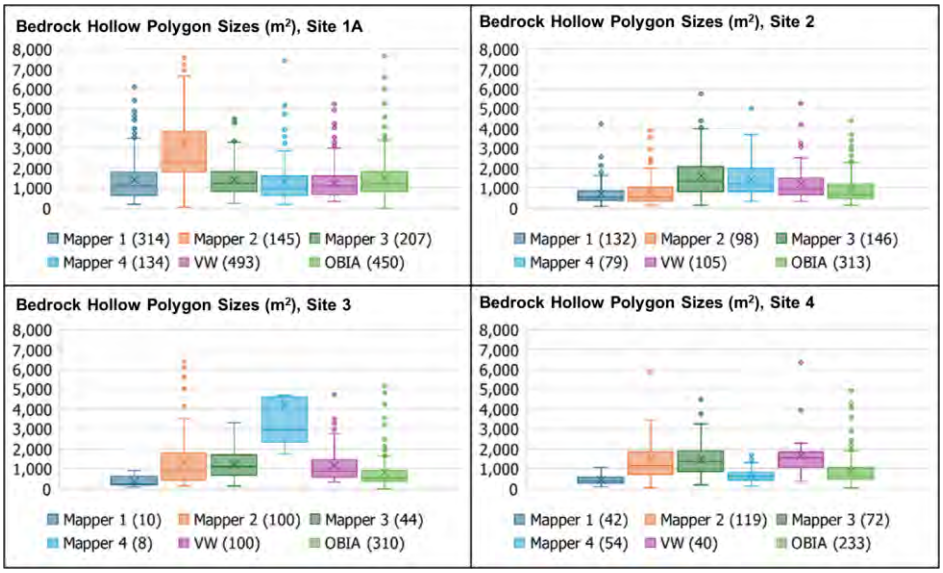


Figure 26~~25~~: Box and whisker plots of ~~bedrock hollow~~bedrock hollow RIL polygon size in square meters. The boxes include 50% of all polygon values, extending from the 25th to the 75th percentiles. The horizontal line inside the box indicates the median; the X indicates the mean. The size of the box indicates the "inter-quartile range." The whiskers extend either to the maximum or minimum value or to 1.5 times the inter-quartile range above and below the 75th and 25th percentiles, whichever is greater (or less). Polygons with values greater than or less than that range are indicated by circles. Number in parentheses indicate the total number polygons mapped.

Commented [PJ174]: Add a label on the plots for the y axis (Bedrock Hollow Polygon Size (m2)) - for all of the box and whisker plots

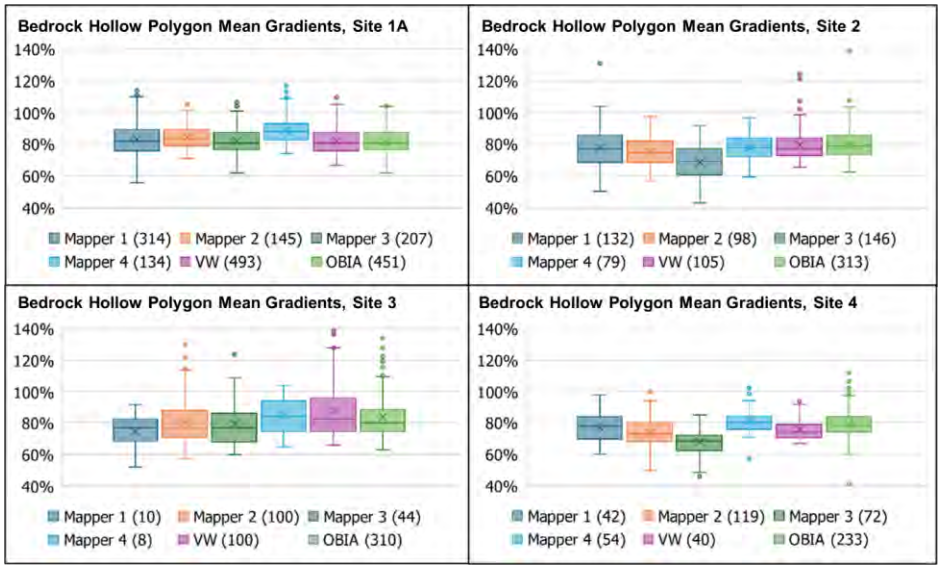


Figure 2726: Box-and-whisker plots of mean gradient for ~~bedrock hollow~~bedrock hollow RIL polygons for each of the mappers and the two computer models for each of the four study sites.

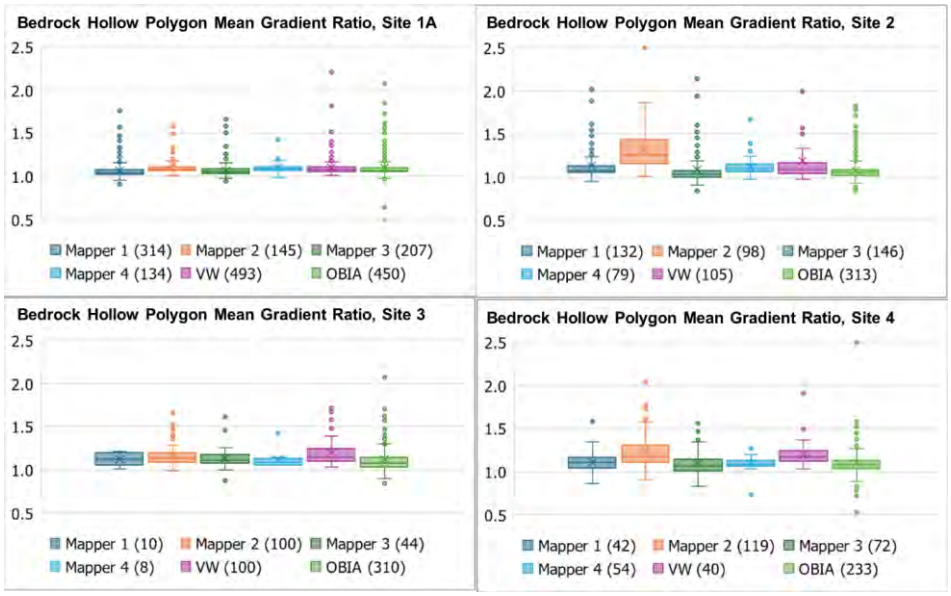


Figure 2827: Box and whisker plots of the mean gradient ratio (upslope-looking / downslope-looking) for ~~bedrock hollow~~bedrock hollow RIL polygons.

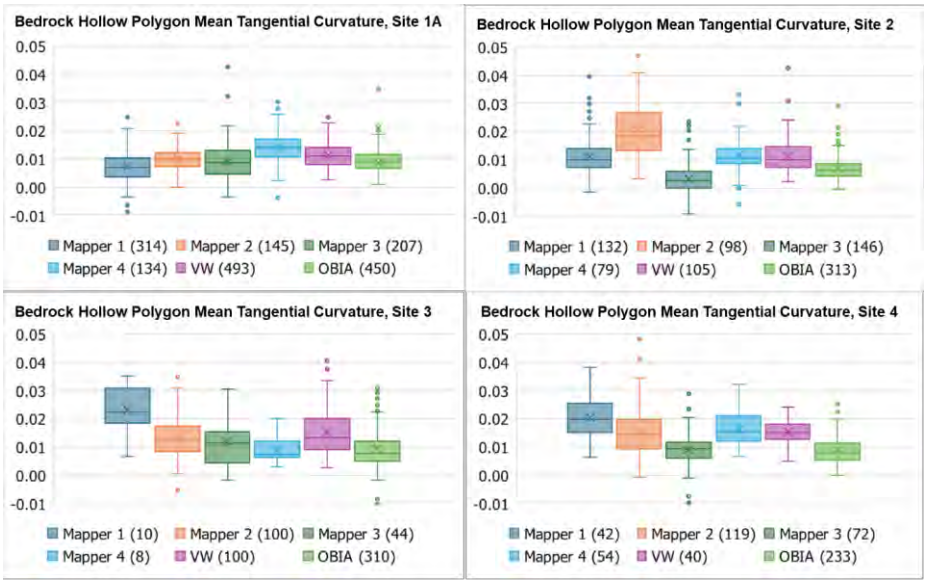


Figure 2928: Box and whisker plots of the mean tangential curvature for ~~bedrock hollow~~bedrock hollow RIL polygons for all mappers, including the two computer models, at all four study sites.

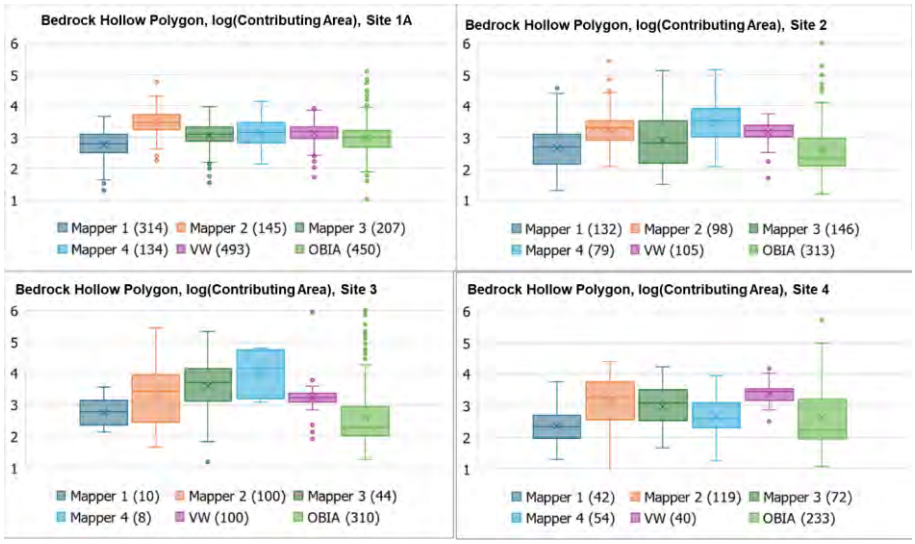


Figure 3029: Box and whisker plots of the logarithm (base 10) of the contributing area (m^2) at the outlet of the ~~bedrock hollow~~bedrock hollow RIL polygons for each mapper, including the two computer models, at each of the four study sites.

4.1.2.6 Voting: Agreement between Mappers and Models at Site 1A

Evaluation of the computer mapped RILs was completed using a voting exercise described as follows: The 47 randomly selected ~~bedrock-hollow~~bedrock hollow sites used for the voting exercise consisted of 23 VW-generated polygons and 24 OBIA-generated polygons. The mapping team voted on these modeled polygons as well, indicating "yes" if they agreed the polygon ~~represented~~adequately identified the presence of an actual RIL or "no" if it did not. Results of these votes are shown in Table ~~9~~10 and Figure 3~~0~~1.

Table 10: Number and proportion of "yes" votes given to both sets of model polygons by the EPs.

	#OBIA YES votes	#VW YES votes	% OBIA (#YES/24)	% VW (#YES/23)	% of all sites (#OBIA + #VW)/47
M1	17	19	71%	83%	77%
M2	19	16	79%	70%	74%
M3	22	21	92%	91%	91%
M4	17	14	71%	61%	66%
M5	15	17	63%	74%	68%
Average			75%	76%	75%

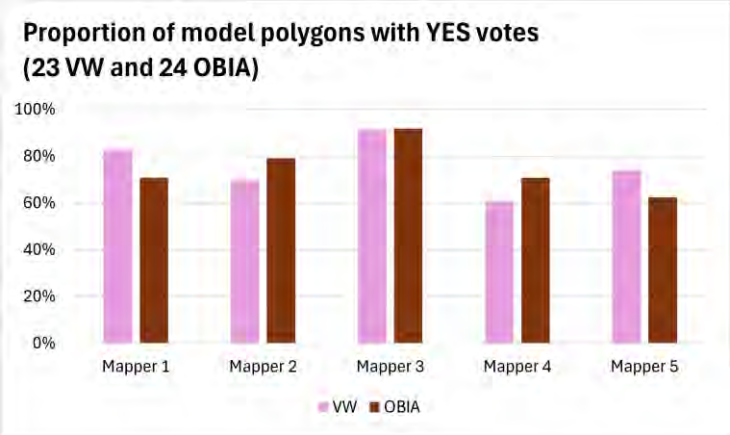


Figure 3~~1~~10: Mappers' agreement with computer models identification of potential ~~bedrock-hollow~~bedrock hollow sites.

Table 1~~1~~0 shows the number of matching sites each mapper had with the models and the number of yes-votes given to each model's polygons. A "match" indicates that, at the ~~bedrock-hollow~~bedrock hollow polygon site delineated by one of the models, that mapper had also drawn a polygon. Because Mapper 5 only mappeda portion of the study area, they are excluded from these results. The number of initial matches each mapper had with the models varied

Commented [PJ175]: This is a different topic and should be under a different heading than the box and whisker plot data

Commented [MM176]: This figure is a little confusing- if we are comparing agreement between each mapper and each model, then splitting the red and pink column would be clearer than stacking them. It would also more clearly show that some mappers agreed more strongly with one model and others with the other model, i.e. no preferred model.

Commented [HB177]: Does this indicate that the two models together or individually indicate that 25% of the sites were false positives? Could the mappers be wrong if sites wee evaluated in the field?

Commented [HB178]: Some places you discuss hollows and others bedrock-hollows. Are they different? Can they be differentiated using the remote imagery?

substantially, ranging from Mapper 1, who had 35 matches out of the 47 (74%) modeled polygons, to Mapper 45, who matched only 186 out of the 47 (3884%) model polygons (Table 10). This is consistent with the relative number of polygons each of them mapped: Mapper 1 identified more ~~bedrock hollow~~bedrock hollow RIL polygons (314) at Site 1A than Mapper 4 (134). All of the mappers voted "yes" for more of the sites than they initially included in their maps as RILs (Table 10). Figure 30 shows that all mappers agreed that 66% or more of the model polygons adequately ~~represent~~identify the presence of a ~~bedrock hollow~~bedrock hollow RIL, averaging 75% yes-votes for the OBIA polygons and 76% for the VW polygons.

Commented [HJ179]: Did mapper 5 even map the whole study area?

Table 11: The number and proportion of matching bedrock hollow polygons drawn by mappers in the 47 randomly selected areas is given in the second column (Matching Sites). The number and proportion of computer-generated bedrock hollow polygons for which a mapper voted "yes," out of the 47 randomly selected sites, is given in the third column.

	Matching sites (proportion)	Number of model polygons given YES votes (proportion)
Mapper 1	35 (74%)	36 (77%)
Mapper 2	26 (55%)	35 (74%)
Mapper 3	26 (55%)	43 (91%)
Mapper 4	18 (38%)	31 (66%)
~~Mapper 5~~	~~16 (34%)~~	~~32 (68%)~~

Outliers in computer-mapped landforms

Of the 188 bedrock hollow polygons delineated by OBIA at Site 1A, 29 of them had no overlap with the manually mapped polygons and 12 had no overlap with the virtual watershed generated polygons. Of the 581 bedrock hollow polygons delineated by the virtual watershed model at Site 1A, 103 of them had no overlap with the manually mapped polygons and 174 had no overlap with the OBIA polygons (Figure 321).

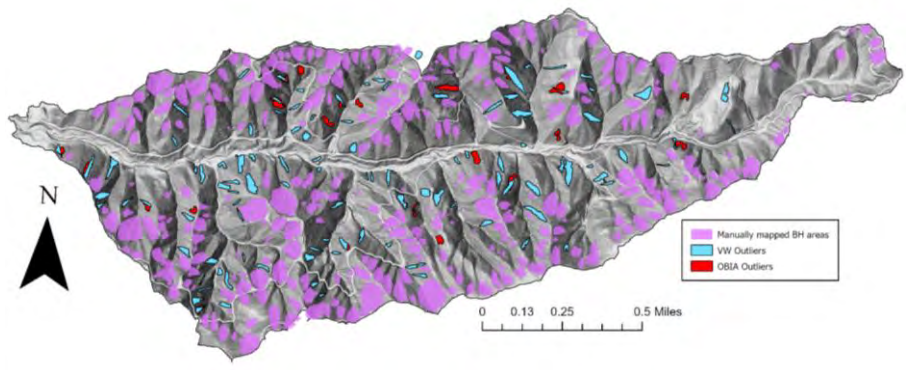


Figure 3231: Site 1A with manually mapped bedrock hollow areas highlighted in purple, OBIA outliers highlighted in red and VW outliers highlighted in blue.

Commented [PJ180]: It would be better to have a legend with the colors labeled

Of these "outlier" model polygons (those with no matching manually mapped polygons) we randomly chose 28 from each model set (Figure 323). The mapping team voted on these 56 outlier polygons, indicating "YES" if they agreed the polygon accurately identified a potential RIL or "NO" if it did not. The proportion of the 56 model outlier polygons with YES votes is shown in Figure 343 for each mapper.

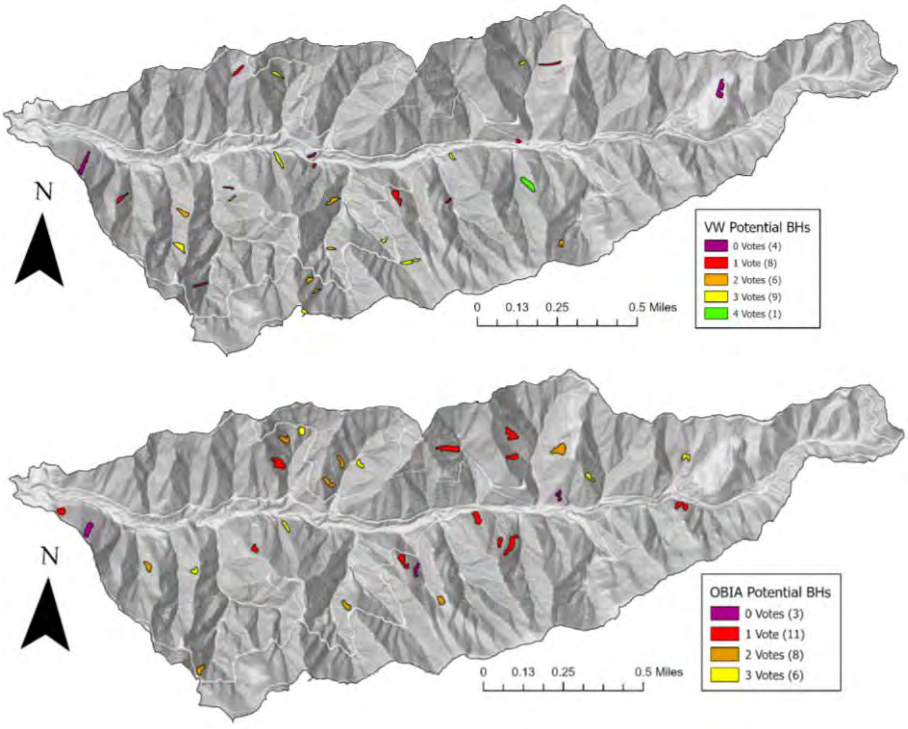


Figure 3332: Modeled polygon outliers with colors representing the number of "yes" votes.

Commented [PJ181]: It is hard to see these on the map because the map is so small. You could try removing the gray outlines or adding bright outlines. They might pop out more.

Commented [JM182R181]: Good idea.

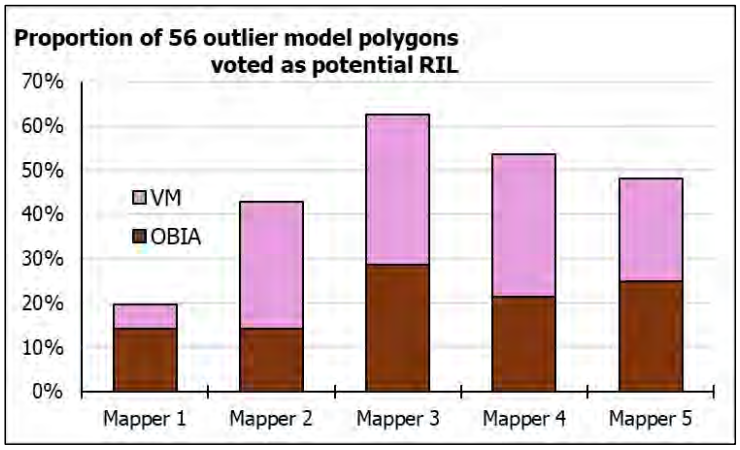


Figure 3433: Proportion of "yes" votes awarded by each mapper for the outliers in each modeled polygon sets in Site 1A.

Key observations from this voting exercise include:

1. Across all mappers, the yes-vote percentages for the computer-drawn polygons ranged from 63-92% for OBIA, with an average of 75%, and 61-91% for VW, with an average of 76%.
2. All of the mappers voted "yes" for more sites than they initially included in their maps as RILs.~~,~~
3. Each mapper indicated that some of the outlier polygons did represent bedrock hollows, even though none of the mappers had drawn one there. This, along with observation 2 above, suggests that all mappers in this project are missing some bedrock hollow polygons in their initial delineations.

Comparison with Site 1B and Calibration of Automated Methods

For the voting exercise described above, the OBIA model was based on the rule set described in Shaw et al. (2017), where landforms are delineated by lidar-derived geomorphometric parameters (e.g., slope shape and steepness, flow accumulation).~~;~~ ~~also note that~~ Landform classes were modified from this rule set to align with the Board Manual ~~, Section 16~~ definitions. The VW model used parameter values based on the RIL descriptions in the Board Manual, ~~Section 16~~. Following the visual inspection of the modeled polygons, we modified the OBIA ruleset and the VW threshold parameters to produce better visual alignment of the modeled polygons with what the mapping team viewed as appropriate-looking RIL polygons. We used a~~A~~ heuristic approach ~~was used~~ to modify the OBIA rule set with a focus on the geometric shape of extracted polygon and spatial relationship to the stream. We implemented changes to consider ~~compaction~~

Commented [HJ183]: Yellow - Explain why.

Commented [HJ184]: Yellow - This should also be better explained if it is going to be retained.

Commented [PJ185]: Are there more than these three? If so, can they all be listed here or are there too many?

Commented [PJ186]: Simplify/make into two sentences

Commented [HB187]: ??

the issue of multiple bedrock hollow polygons being mapped as a single feature~~of bedrock hollow polygons~~ and to refine edge extraction methods used for inner gorges. We m~~M~~odifi~~ed~~~~cation of~~ the VW parameter threshold values ~~was~~ based on visual inspection of the frequency distributions of the topographic attributes associated with the mapped RIL polygons at Site 1B. ~~These updated models~~ W~~w~~e~~re~~ then used these updated models to produce modeled ~~bedrock hollow~~bedrock hollow and ~~inner gorge~~inner gorge RIL polygons for Sites 1A, 2, 3, and 4.

Commented [MM188]: Is this the maximum size of a hollow?

As described in Section 3, a virtual watershed data structure provides a means of delineating landforms based on geomorphic relationships. We initially defined those relationships based on interpretation of the RIL landforms in the Board Manual, but after seeing the frequency distributions of gradient, curvature, and other topographic attributes encompassed by the RIL polygons delineated by our team of mappers for Site 1B, we decided to set parameter values based on those distributions to calibrate the model and then extrapolate that calibrated model to the other four sites.

Commented [MM189]: It would be good to expand on this outside of the Appendix, because it sounds like these are going to be used for field validation in later studies and could become the basis for refining RIL definitions.

The cumulative frequency distributions and description of variability between mapped and modeled polygons for these topographic attributes can be found in Appendix G.

4.1.3 PHASE 1 CONCLUSIONS

Phase 1 revealed ~~unresolvable~~ discrepancies in how the EPs delineate bedrock hollows and inner gorges using remote data. Quantitative comparisons using confusion matrices and summary metrics such as balanced accuracy, recall, and precision confirmed these discrepancies. While mappers generally agreed on the locations where these landforms did not exist, their interpretations of where potential RILs were present and how to draw the discrete boundary were inconsistent. Spatial overlap between mappers data was ~~minimal~~unsatisfactory, and efforts to align the overlap did not produce useful results.

Commented [HJ190]: Context is needed if "minimal" is going to be retained in this sentence.

A key finding was that uncertainty stemmed not just from different interpretations of Board Manual RIL criteria, but that each mapper brings their own degree of meticulousness, patience, and GIS (digitizing) skills to this task, which can also contribute to differences in RIL polygons drawn by different mappers, or even by the same mapper at different times.

Furthermore, voting exercises showed that mappers ~~frequently~~ recognized bedrock hollows in model-generated polygons that they had not initially delineated themselves, ~~suggesting that manual mapping alone likely omits some true RIL features.~~ Despite the inconsistency, mappers broadly agreed that the models identified many plausible bedrock hollows and inner gorges.

These findings underscored the need for a more objective and reproducible approach to evaluating model performance. We initiated Phase 2 to address this need by removing manual reference maps and instead assessing model output directly through experienced practitioner

review of computer-generated polygons. This shift allowed the Project Team to focus on the relative strengths of each model, reduce bias introduced by individual mapping styles, and better identify opportunities for improving automated delineation of RIL landforms.

4.2

PHASE 2 – MODEL EVALUATION VIA EXPERIENCED PRACTITIONER FEEDBACK

The insights gathered from Phase 1 of this project motivated the Project Team to design additional exercises that did not require the use of manual landform maps to evaluate the computer-based mapping methods. Team discussions determined that the voting exercises conducted during Phase 1 of the project provided valuable insights and could serve as a foundational basis for the exercises in Phase 2.

4.2.1 PHASE 2 METHODS

Phase 2 of this project focused solely on EP evaluations of the computer-drawn landform polygons with no reliance on manually drawn maps. For each of the four study areas, we had two sets of computer-generated bedrock hollow and inner gorge landform polygons, one produced by the OBIA model and another by the VW model. These polygons delineated every bedrock hollow and inner gorge recognized by the models. We randomly selected a subset of the computer-generated polygons for closer examination by the EPs. Maps showing the selected polygons for each study area are provided in Section 4.2.2. The EP's examined each of these selected polygons using the spreadsheet-method devised for the voting exercises in Phase 1, entering a "yes" or "no" vote across several prompts for each polygon. These prompts, listed below, were designed to help determine ways to improve the models' landform delineations.

For potential bedrock hollow polygons, the EPs provided yes-or-no votes for 5 prompts, with a column to write out comments as needed.

1. Does the polygon represent a reasonable screen (RS) for a bedrock hollow or inner gorge? A yes-vote indicated that the mapper believed that there is a bedrock hollow or inner gorge present where drawn by the model. If the model consistently identifies landforms in areas where a practitioner would reasonably expect to find them, based on the model's user-defined input parameters, this indicates that the parameters are suitable for guiding an initial screening of landform presence.
2. Is the polygon correctly drawn (CD)? A yes-vote indicated that the EP believed that the model did an adequate job delineating the boundary of the landform polygon. The answers in this category helped determine potential ways to improve the model's delineations of defined landform class boundaries.
3. Was deliverability considered for this polygon? If so, could the landform deliver sediment? A yes-vote indicated that the model considered deliverability ~~when voting yes~~

Commented [HJ191]: Yellow - This section could use some more detailed site-specific maps showing how bedrock hollows and inner gorges were delineated.

Commented [HJ192]: Why did the team consider both bedrock hollows and inner gorges in this question about whether a bedrock hollow polygon is a reasonable screen? I suspect the team was asking this question for a good reason, but it should be explained.

Commented [HB193]: I didn't see a summary of deliverability calls. Also, what criteria did the models use to characterize deliverability?

~~or no~~, and that the EP agreed the landform could deliver sediment. Answers in this category helped determine if the model is including landforms that meet the rule criteria.

4. Was convergence considered for this polygon (did the EP consider convergence when voting yes-or-no)? While the B~~b~~oard M~~m~~anual criteria for convergence is qualitative and based on practitioner observation, the models use tangential curvature to set specific convergence parameters. The answers in this category helped determine if the parameters set by the models align with the observations made by the EPs.
5. Is the landform a secondary feature (i.e., inside a DSL)?

For potential inner gorge polygons, the EPs only provided yes-or-no votes for prompts 1 & 2.

We evaluated the results of the voting exercises by calculating the average percentage of "yes" votes for prompts 1 and 2, looking at the prompt 3, 4 and 5 votes to identify trends (for bedrock hollows), and observing repeated similar comments among mappers.

~~Four EPs participated in the voting exercises in Phase 2. For Site 1A, 2, 3 & 4, each EP voted on a set of potential bedrock hollow and potential inner gorge~~inner gorge ~~landform polygons produced by each of the computer-based methods.~~ We determined the number of polygons selected (sample size) at the individual sites, from each modeled set, using Cochran's formula (1977). The formula calculates a minimum sample size needed for a specific margin of error (E) and confidence level:

$$n_0 = \frac{Z^2 \times p \times (1-p)}{E^2}$$

For this project, we chose a desired margin of error (E) of 10%. The z-value is determined by the area of the normal curve to the left of the desired confidence level (for this project, we chose a confidence level of 95%). The p-value represents the estimated proportion of an attribute that is present in the population – in this case, the proportion of "yes" or "no" votes out of all votes, which is unknown. Assuming maximum variability, we chose to use a p-value of 50%.

Cochran's formula is used for large (or infinite) population sizes. Due to our smaller sample size (N), we needed to use the modified formula for finite populations to get our final sample size (n) (Divakar, 2021):

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

Commented [HB194]: What if the model did not consider deliverability and the EP did and decide deliverability was possible?

Commented [GS195]: This description is difficult to follow and I still don't know what you did. What do you mean 'voted on'? What were the criteria used for voting? What was different between each set of potential bedrock hollow and inner gorge landform polygons? Why did you need to select a certain number of polygons? Clear, careful language is needed here to describe your process. Yellow.

Commented [GS196]: These image based equations show up fine.

The EPs used the same spreadsheet method used for the voting exercises in Phase 1, entering a "yes" or "no" vote across several prompts for each randomly selected polygon in the two model's sets. These are listed in Section 4.2.

Phase 2 of this project consists of additional automated model evaluation exercises that did not require the use of manual landform maps. Team discussions, following the results of Phase 1, determined that the voting exercises conducted during Phase 1 provided valuable insights and could serve as a foundational basis for the exercises in Phase 2.

Four EPs participated in this voting exercise. For Site 1A, 2, 3 and 4, each EP voted on a randomly selected set of potential bedrock hollow and inner gorgeinner gorge landform polygons produced by each of the computer based methods. The Project Team agreed upon voting categories that would help determine potential ways to improve the models' landform delineations.

For potential bedrock hollow polygons, the EPs provided yes or no votes for 5 prompts, with a column to write out comments as needed.

1. Does the polygon represent a reasonable screen (RS) for a bedrock hollow or inner gorgeinner gorge? A yes vote indicated that the mapper believed that there is a bedrock hollow or inner gorgeinner gorge present where drawn by the model.

2. Is the polygon correctly drawn (CD)? A yes vote indicated that the EP believed that the model did an adequate job delineating the boundary of the landform polygon. The answers in this category helped determine potential ways to improve the model's delineations of defined landform class boundaries.

3. Was deliverability considered for this polygon? If so, could the landform deliver sediment? A yes vote indicated that the model considered deliverability when voting yes or no, and that the EP agreed the landform could deliver sediment. Answers in this category helped determine if the model is including landforms that meet the rule criteria.

4. Was convergence considered for this polygon (did the EP consider convergence when voting yes or no)? While the board manual criteria for convergence is qualitative and based on practitioner observation, the models use tangential curvature to set specific convergence

Commented [GS197]: I think you need to list the prompts here. Otherwise it is not clear what you mean.

Commented [PJ198]: Green - The following few paragraphs and list are methods, not results. The project team should consider reorganizing the report to something like: 1) Introduction, 2) Data, 3) Phase 1 methods + results, 4) Phase 2 methods + results, 4) Discussion, 5) Recommendations (for next sub-projects). It could reduce redundancy, improve readability, and help clarify what, when, and how different phases of the project were conducted.

Commented [HJ199]: Why did the team consider both bedrock hollows and inner gorges in this question about whether a bedrock hollow polygon is a reasonable screen? I suspect the team was asking this question for a good reason, but it should be explained.

Commented [HB200]: I didn't see a summary of deliverability calls. Also, what criteria did the models use to characterize deliverability?

Commented [HB201]: What if the model did not consider deliverability and the EP did and decide deliverability was possible?

~~parameters. The answers in this category helped determine if the parameters set by the models align with the observations made by the EPs.~~

~~5. Is the landform a secondary feature (i.e. inside a DSL)?~~

~~For potential inner-gorgeinner gorge polygons, the EPs only provided yes-or-no votes for prompts 1 & 2.~~

~~The results of the voting exercises were evaluated by calculating the average percentage of "yes" votes for prompts 1 and 2, looking at the prompt 3, 4 and 5 votes to identify trends (for bedrock hollows), and observing repeated similar comments among mappers.~~

4.2.2 Phase 2 Results

Bedrock Hollows

To evaluate model performance in delineating bedrock hollows, we summarized experienced practitioner (EP) voting results for both the VW and OBIA models across all study sites. Figure 345 shows the percentage of potential bedrock hollow polygons to receive yes-votes for prompts 1 & 2 across all sites for both models. The "RS yes-vote percentage" represents the proportion of the total polygon set that received a yes-vote for prompt 1. The "CD yes-vote percentage" represents the proportion of those polygons given yes-votes for prompt 2. Of the VW polygons, Site 2 had the highest average RS-percentage at 91%, but the third ~~second~~-lowest CD-percentage at 37%, indicating that the model had correctly identified a bedrock hollow but the delineated boundary of the feature could use improvement. The average yes-vote percentage across all sites for VW was 82% for RS and 41% for CD. Of the OBIA polygons, Sites 1A, 2 & 43 had average RS yes-vote percentages within 1% of each other and an average of 71%, though the average CD yes-vote percentages varied more than RS and had an average of 46%.

Commented [PJ202]: Please add an intro sentence to this paragraph and the inner gorge paragraph.

Commented [HJ203]: Third?

Commented [HJ204]: How do the values displayed in Figure 35 generate an average of "41%"?

Commented [HJ205]: 4?

Commented [HJ206]: How do the values displayed in Figure 35 generate an average of "46%"?

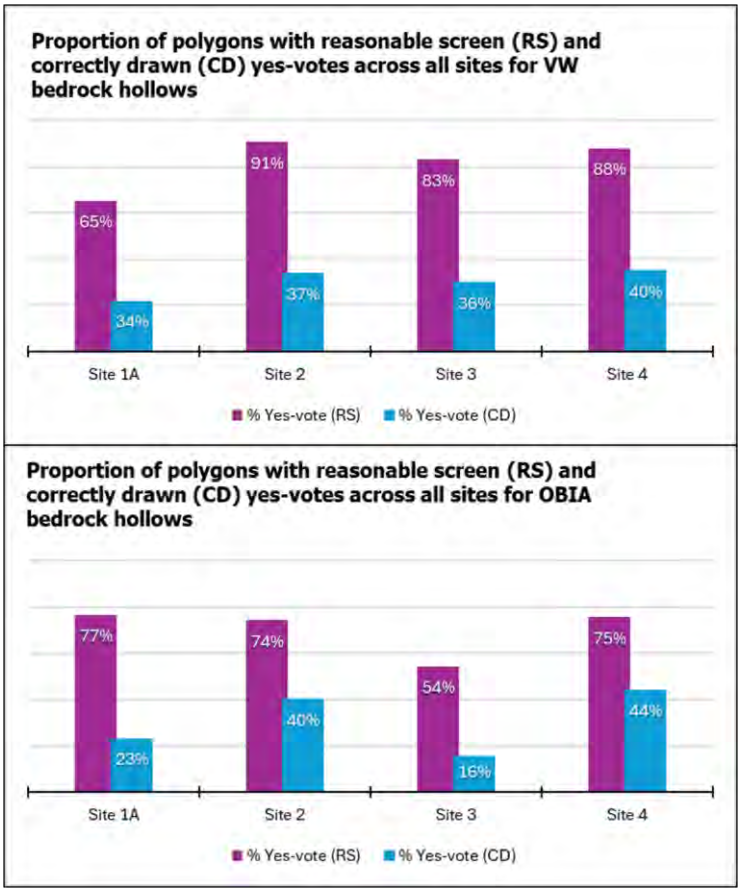


Figure 3534: Bedrock hollow reasonable screen (RS) and correctly drawn (CD) yes-vote percentages across all sites for the total VW (top) and OBIA (bottom) modeled polygon sets. For CD, the percentages shown represent the proportion of polygons, out of the subset of polygons that received yes-votes for RS, also received yes-votes for CD.

Commented [HB207]: VW is best on avg for both. However low CD values look to be a concern.

Inner -Ggorges

Commented [PJ208]: Please check the whole document for consistent use of dashes.

Figure 3536 shows the proportion of potential inner gorge polygons to receive yes-votes for prompts 1 & 2 across all sites for both models. The yes-vote percentages were higher than those for bedrock hollows at all sites, for both models. The models also had lower variability in yes-vote percentages between each other for inner -gorges compared to the variability seen for bedrock hollows. For example, the difference between the models for RS yes-vote percentages in bedrock hollows ranged between 12-29%, whereas the difference for inner gorges ranged between 0-7%.

Additionally, the CD yes-vote percentages were higher overall for inner gorges than the bedrock hollows.

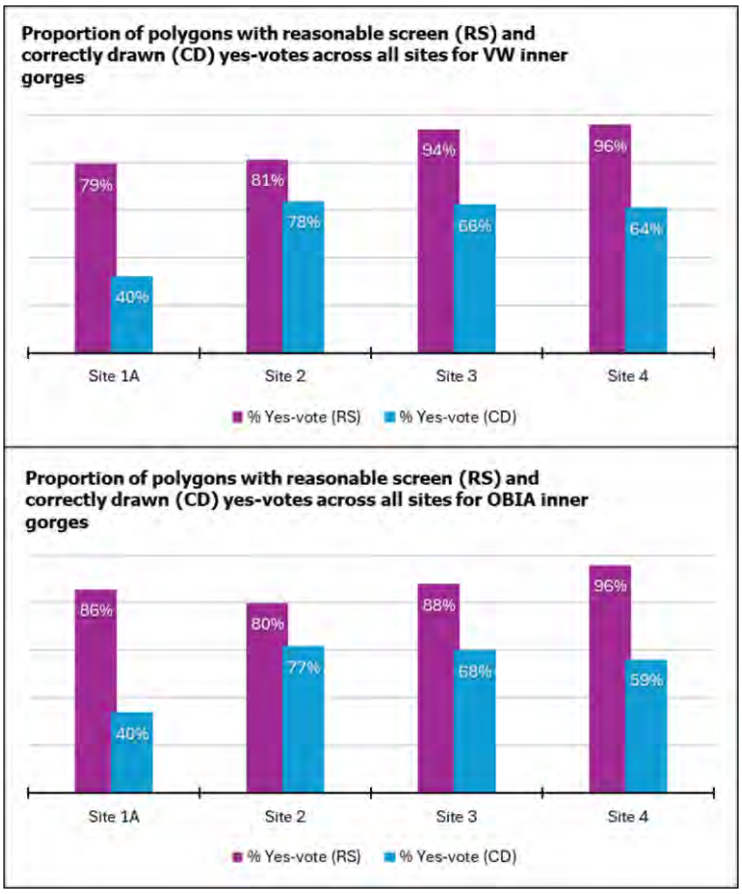


Figure 3635: Inner gorge reasonable screen (RS) and correctly drawn (CD) yes-vote percentages across all sites for the total VW (top) and OBIA (bottom) modeled polygon sets. For CD, the percentages shown represent the proportion of yes-vote RS polygons that received CD yes-votes.

Commented [HB209]: OBIA = VW but low CDs are a concern.

Key observations from the Phase 2 exercise:

- For the VW model, polygons at Site 1A had the lowest average RS and DC yes-vote percentages for both bedrock hollows and inner gorges.
- For the OBIA model, bedrock hollow polygons at Site 3 received the lowest percentage of yes-votes for reasonable-screen.

Commented [HB210]: True but what is the importance? If it is a site by model interaction, can site differences be adequately classified in order to determine which model to use?

- The mappers indicated that on average 70%, or more, of both model's landform polygons were a reasonable screen for bedrock hollows and inner ~~-~~gorges.
- The mappers indicated that 39% of bedrock hollow polygons and 62% of inner gorge polygons, that represented a reasonable screen, were correctly drawn.

The results of the voting exercise at each site are described below.

Site 1A

Bedrock Hollows

The voting results for Site 1A potential bedrock hollow polygons for each model are shown in Tables 1~~1~~2 & 1~~3~~2. The mapper's averaged votes indicated they agree 65% of the 81 selected VW polygons represented a reasonable screen for a bedrock hollow. Votes for "correctly drawn" were then cast for those 65%, ~~Of tho~~and of those, ~~se 65%,~~ they considered 35% to be correctly drawn.

Table 12: Voting results for Site 1A VW potential bedrock hollow polygons.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/81)	%Yes (CD/RS)	%Yes (CD/81)
M1	49	0	60%	0%	0%
M2	51	15	63%	29%	19%
M3	45	35	56%	78%	43%
M4	66	21	81%	32%	26%

Of the OBIA potential bedrock hollow polygons, the mappers averaged votes indicated that 77% of the 79 selected polygons represented a reasonable screen for a bedrock hollow, 65% of which they considered to be correctly drawn.

Table 13: Voting results for Site 1A OBIA potential bedrock hollow polygons.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/79)	%Yes (CD/RS)	%Yes (CD/79)
M1	69	0	87%	0%	0%
M2	71	28	90%	39%	35%
M3	50	34	63%	68%	43%
M4	52	11	66%	21%	14%

The mappers voted "yes" unanimously on 70 (44%) of the 160 total potential bedrock hollow polygons produced by both models. Figure 3~~7~~6 shows the number of yes-votes each polygon in the set received. The mappers unanimously agreed that 60 represented a reasonable screen for a bedrock hollow (22 VW and 38 OBIA), and that 10 of them did not (3 VW and 7 OBIA). Figure 3~~8~~7 shows the distribution of mapper agreement on RS designation.

Commented [HB211]: If you are trying to determine which model is best why did you confound the two in the conclusions? For bedrock hollows the VW model looks on average better for both RS & CD.

Commented [HJ212]: These "correctly drawn" percentages need to be adjusted to represent the number of correctly drawn polygons out of the whole population of polygons.

Commented [HJ213]: Yellow - The %Yes (CD/81) should also be shown.

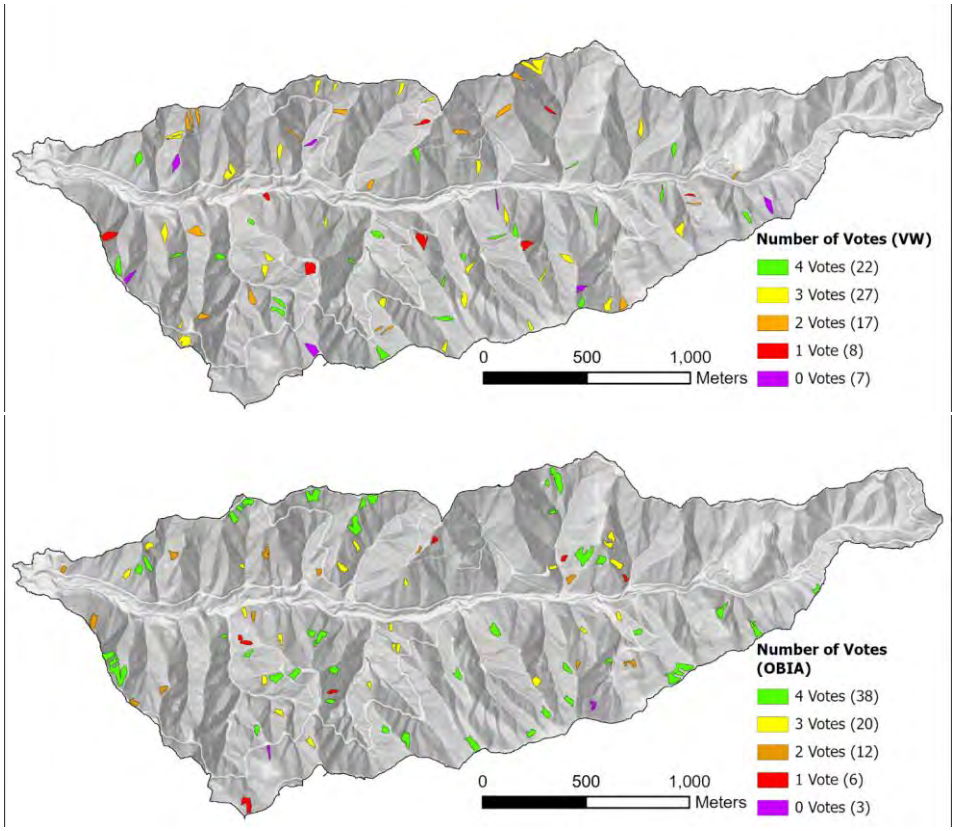


Figure 3736: Number of yes-votes received for each polygon at Site 1A in the potential bedrock hollow sets (VW top and OBIA bottom) and how many polygons received that number of votes.

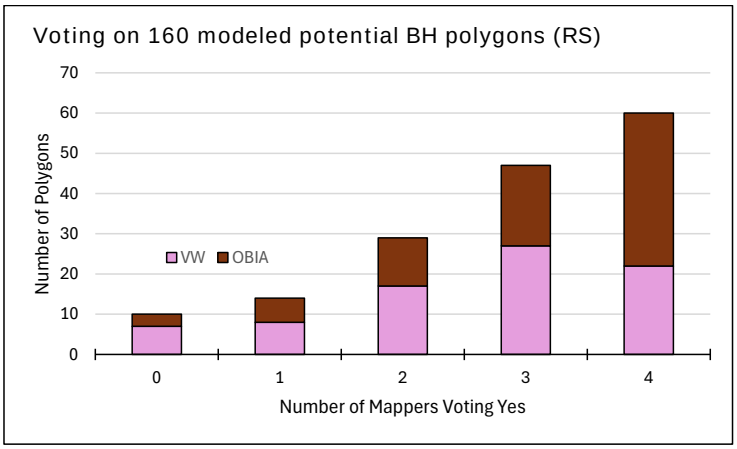


Figure 3837: Distribution of mapper agreement on RS designation votes for Site 1A. The x-axis represents the number of mappers who voted "yes" for a given polygon, while the y-axis shows the count of polygons that received each corresponding number of yes-votes.

Inner gorges

To evaluate model performance in delineating inner gorges, we summarized experienced practitioner (EP) voting results for both the VW and OBIA models across all study sites. The voting results for Site 1A potential inner gorge polygons for each model are shown in Tables 134 & 154. The mappers' averaged votes indicated they agreed 79% of the 81 selected VW polygons represented a reasonable screen for an inner gorge. Of those 79%, they considered 36% to be correctly drawn.

Commented [HJ214]: Yellow - The %Yes (CD/81) should also be shown.

Table 14: Voting results for VW potential inner gorge polygons at Site 1A.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/81)	%Yes (CD/RS)	%Yes (CD/81)
M1	78	18	96%	23%	22%
M2	74	46	91%	62%	57%
M3	72	38	89%	53%	47%
M4	33	2	41%	6%	2%

Of the OBIA potential inner gorge polygons, the mapper's averaged votes indicated they agree 86% of the 76 selected polygons represented a reasonable screen for an inner gorge, 40% of which they considered to be correctly drawn.

Table 15: Voting results for OBIA potential inner gorge polygons at Site 1A.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/76)	%Yes (CD/RS)	%Yes (CD/76)
M1	72	8	95%	11%	11%
M2	71	35	93%	49%	46%
M3	68	43	89%	63%	57%
M4	49	17	64%	35%	22%

The mappers voted "yes" unanimously on 79 (50%) of the 157 total potential inner gorge polygons produced by both models. Figure 398 shows the number of yes-votes each polygon in the set received. The mappers unanimously agreed that 76 of the polygons represented a reasonable screen for an inner gorge (32 VW and 44 OBIA), and that 3 of them did not (2 VW and 1 OBIA). Notably, 85% of the selected VW polygon set and 90% of the selected OBIA polygon set received at least two yes-votes. Figure 4039 shows the distribution of mapper agreement on RS designation.

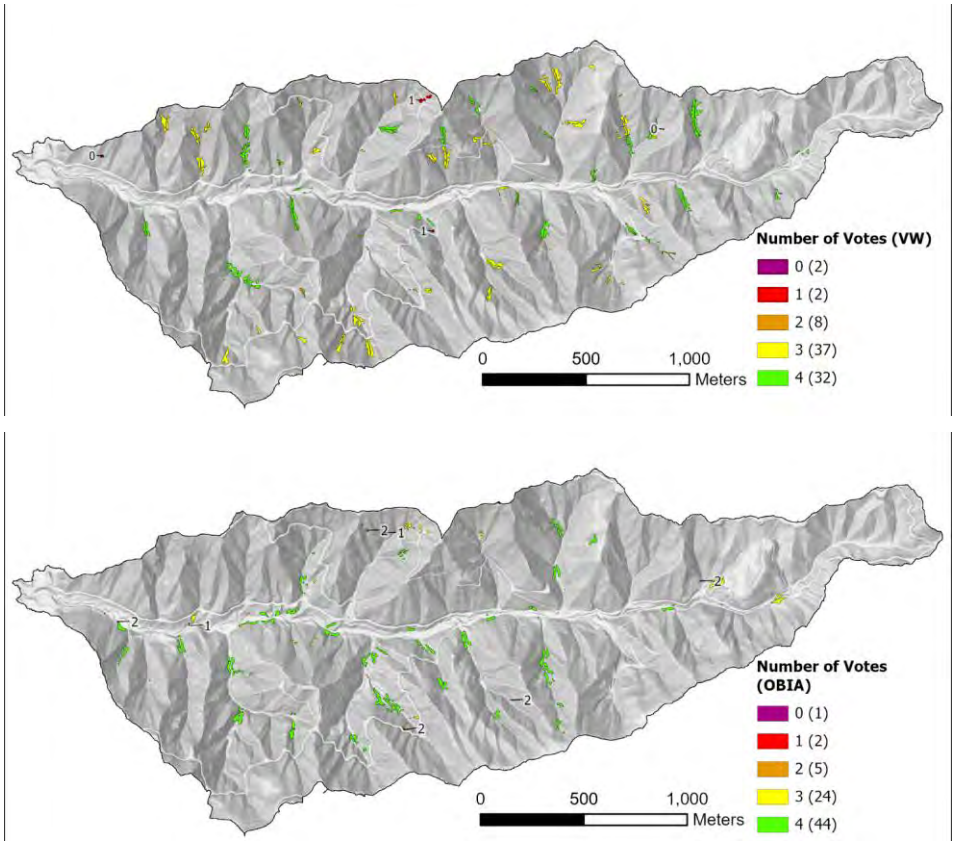


Figure 3938: Number of yes-votes received for each polygon in the potential inner gorge sets (VW top and OBIA bottom) for Site 1A and how many polygons received that number of votes.

Commented [PJ215]: These paired figures are too small to useful for the reader (applies to all of the maps in this section).

Commented [JM216R215]: I agree.

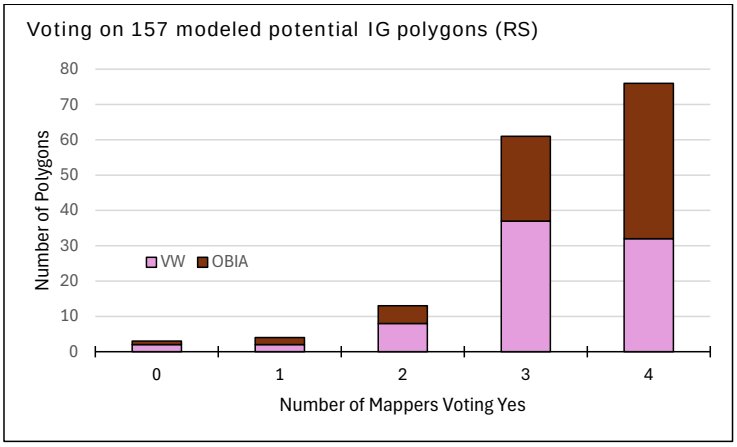


Figure 4039: Distribution of mapper agreement on RS designation votes for Site 1A. The x-axis represents the number of mappers who voted "yes" for a given polygon, while the y-axis shows the count of polygons that received each corresponding number of yes-votes.

Site 2

Bedrock Hollows

The voting results for Site 2 potential bedrock hollow polygons for each model are shown in Tables 156 & 176. The mapper's averaged votes indicated they agree 91% of the 51 selected VW polygons represented a reasonable screen for a bedrock hollow. Of those 91%, they considered 37% to be correctly drawn.

Table 16: Voting results for VW potential bedrock hollow polygons at Site 2.

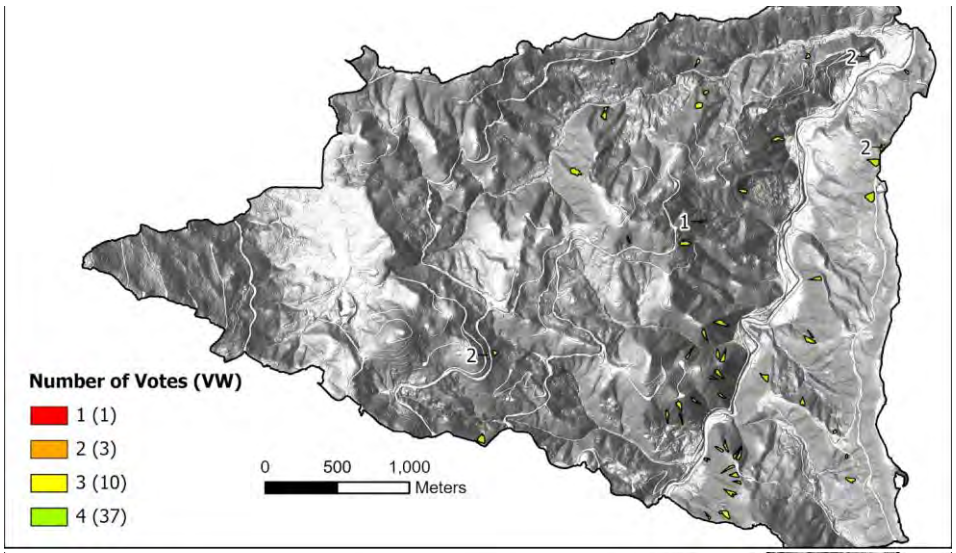
	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/51)	%Yes (CD/RS)	%Yes (CD/51)
M1	51	8	100%	16%	16%
M2	41	15	80%	37%	29%
M3	49	34	96%	69%	67%
M4	44	12	86%	27%	24%

Of the OBIA potential bedrock hollow polygons, the mappers' averaged votes indicated they agree 76% of the 66 selected polygons represented a reasonable screen for a bedrock hollow, 53% of which they considered to be correctly drawn.

Table 17: Voting results for OBIA potential bedrock hollow polygons at Site 2.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/66)	%Yes (CD/RS)	%Yes (CD/66)
M1	58	13	97%	22%	22%
M2	41	28	62%	68%	42%
M3	56	54	85%	96%	82%
M4	41	11	62%	27%	17%

The mappers voted "yes" unanimously on 64 (55%) of the 117 total potential bedrock hollow polygons produced by both models. Figure 410 shows the number of yes-votes each polygon in the set received. The mappers unanimously agreed that 62 represented a reasonable screen for a bedrock hollow (37 VW and 25 OBIA), and that 2 from the OBIA set did not. Figure 421 shows the distribution of mapper agreement on RS designation.



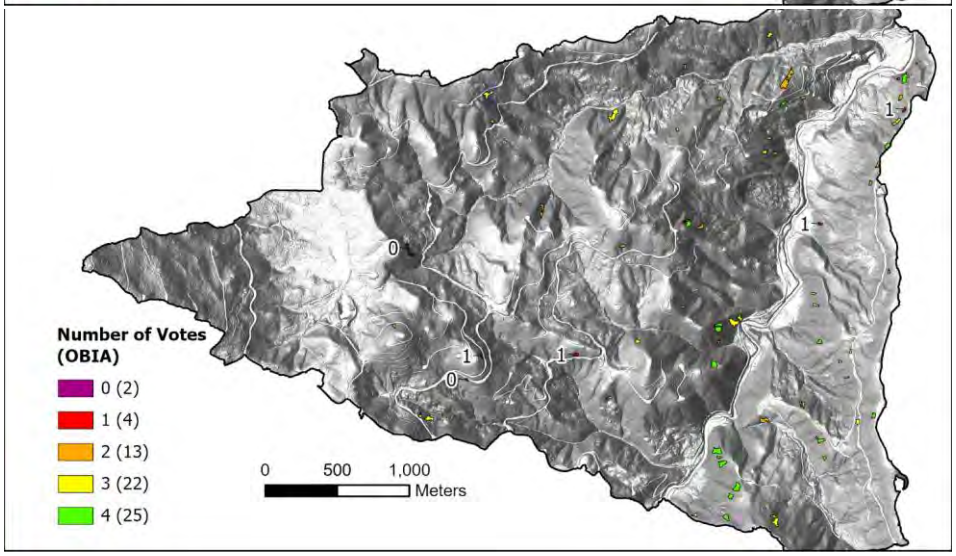


Figure 4140: Number of yes-votes received for each polygon in the potential bedrock hollow sets (VW top and OBIA bottom) for Site 2 and how many polygons received that number of votes.

Commented [JM217]: Hard to read map.

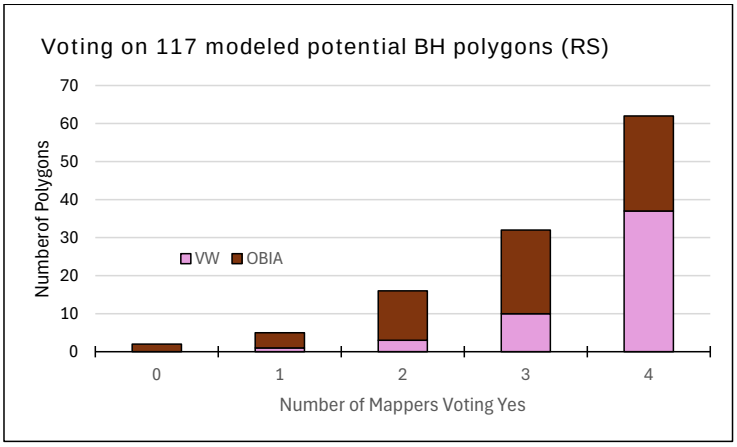


Figure 4241: Distribution of mapper agreement on RS designation votes for Site 2. The x-axis represents the number of mappers who voted "yes" for a given polygon, while the y-axis shows the count of polygons that received each corresponding number of yes-votes.

Inner gorges

The voting results for Site 2 potential inner gorge polygons for each model are shown in Tables 187 & 198. The mappers' averaged votes indicated they agreed 83% of the 84 selected VW polygons represented a reasonable screen for an inner gorge. Of those 83%, they considered 75% to be correctly drawn.

Table 18: Voting results for VW potential inner gorge polygons at Site 2.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/84)	%Yes (CD/RS)	%Yes (CD/84)
M1	71	67	90%	94%	85%
M2	76	63	90%	83%	75%
M3	79	67	94%	85%	80%
M4	47	17	56%	36%	20%

Of the OBIA potential inner gorge polygons, the mapper's averaged votes indicated they agreed 83% of the 81 selected polygons represented a reasonable screen for an inner gorge, 75% of which they considered to be correctly drawn.

Table 19: Voting results for OBIA potential inner gorge polygons.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/81)	%Yes (CD/RS)	%Yes (CD/81)
M1	65	61	92%	94%	86%
M2	71	50	88%	70%	62%
M3	74	65	91%	88%	80%
M4	49	24	60%	49%	30%

The mappers voted "yes" unanimously on 85 (52%) of the 165 total potential inner gorge polygons produced by both models. Figure 43~~2~~ shows the number of yes-votes each polygon in the set received. The mappers unanimously agreed that 82 represented a reasonable screen for an inner gorge (41 from each model), and that 3 of them did not (1 VW and 2 OBIA). Figure 4~~3~~34 shows the distribution of mapper agreement on RS designation.

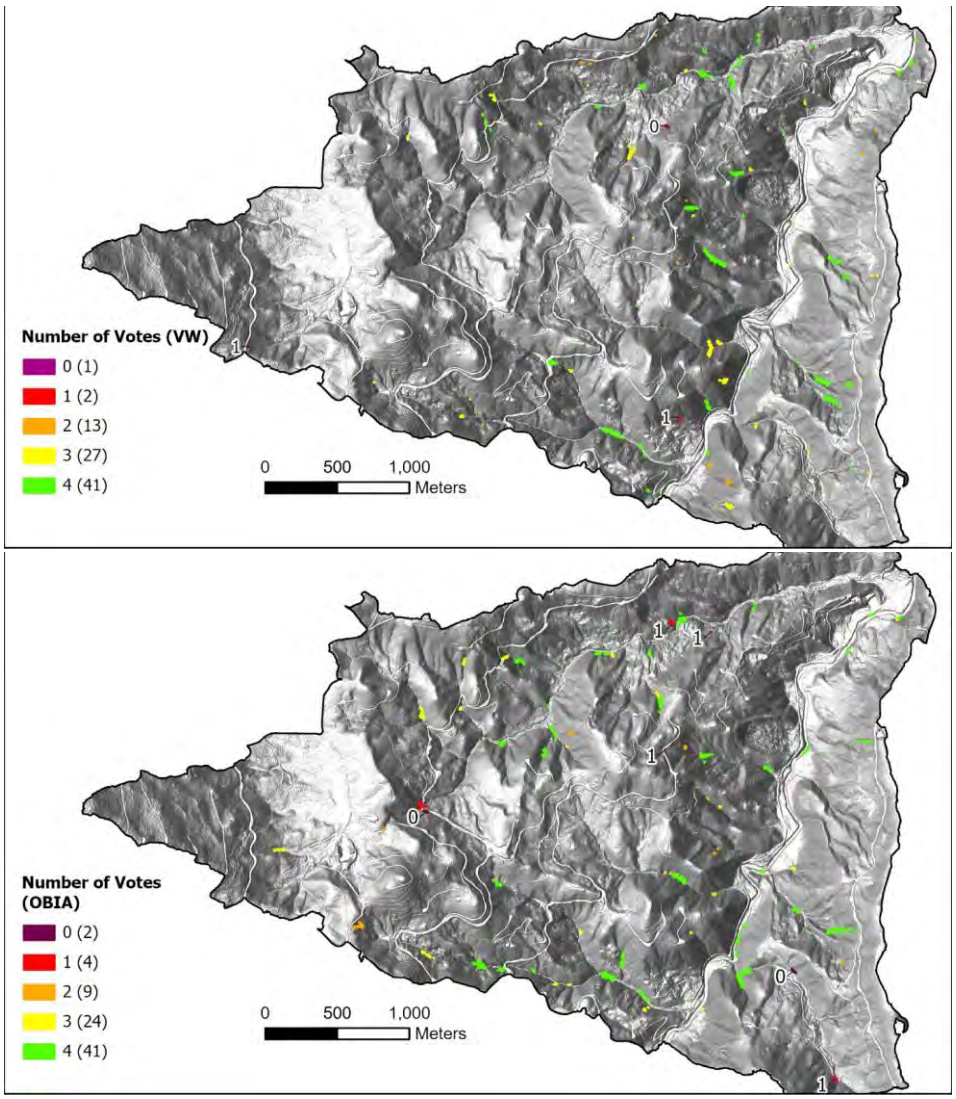


Figure 4342: Number of yes-votes received for each polygon in the potential ~~bedrock hollow~~inner gorge sets (VW top and OBIA bottom) for Site 2 and how many polygons received that number of votes.

Commented [PJ218]: Should this say inner gorges?

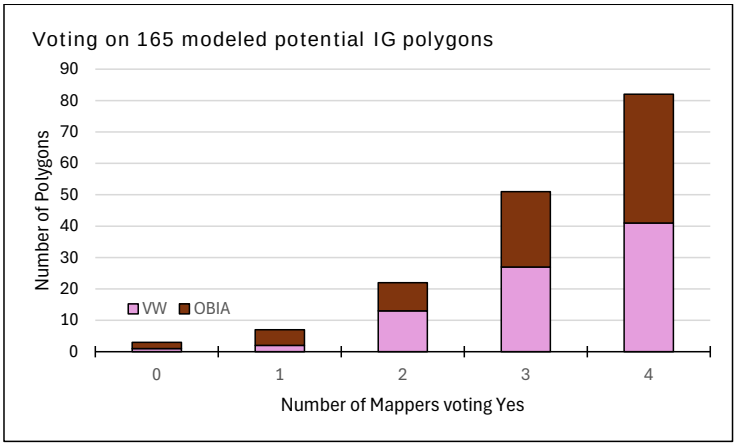


Figure 4443: Distribution of mapper agreement on RS designation votes for Site 2 inner gorges. The x-axis represents the number of mappers who voted "yes" for a given polygon, while the y-axis shows the count of polygons that received each corresponding number of yes-votes.

Site 3

Bedrock Hollows

The voting results for Site 3 potential bedrock hollow polygons for each model are shown in Tables 2019 & 210. The mapper's averaged votes indicated they agreed 83% of the 78 selected VW polygons represented a reasonable screen for a bedrock hollow. Of those 83%, they considered 36% to be correctly drawn.

Table 20: Voting results for VW potential bedrock hollow polygons at Site 3.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/78)	%Yes (CD/RS)	%Yes (CD/78)
M1	77	17	99%	22%	22%
M2	67	34	86%	51%	44%
M3	63	33	81%	52%	42%
M4	52	10	67%	19%	13%

Of the OBIA potential bedrock hollow polygons, the mapper's averaged votes indicated they agreed 54% of the 90 selected polygons represented a reasonable screen for a bedrock hollow, 31% of which they considered to be correctly drawn.

Table 21: Voting results for OBIA potential bedrock hollow polygons at Site 3.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/90)	%Yes (CD/RS)	%Yes (CD/90)
M1	77	12	86%	16%	13%
M2	41	21	46%	51%	23%
M3	36	11	40%	31%	12%
M4	42	12	47%	29%	13%

The mappers voted "yes" unanimously on 74 (44%) of the 168 total potential bedrock hollow polygons produced by both models. Figure 454 shows the number of yes-votes each polygon in the set received. The mappers unanimously agreed that 66 represented a reasonable screen for a bedrock hollow (49 VW and 17 OBIA), and that 8 did not (1 VW and 7 OBIA). Figure 456 shows the distribution of mapper agreement on RS designation.

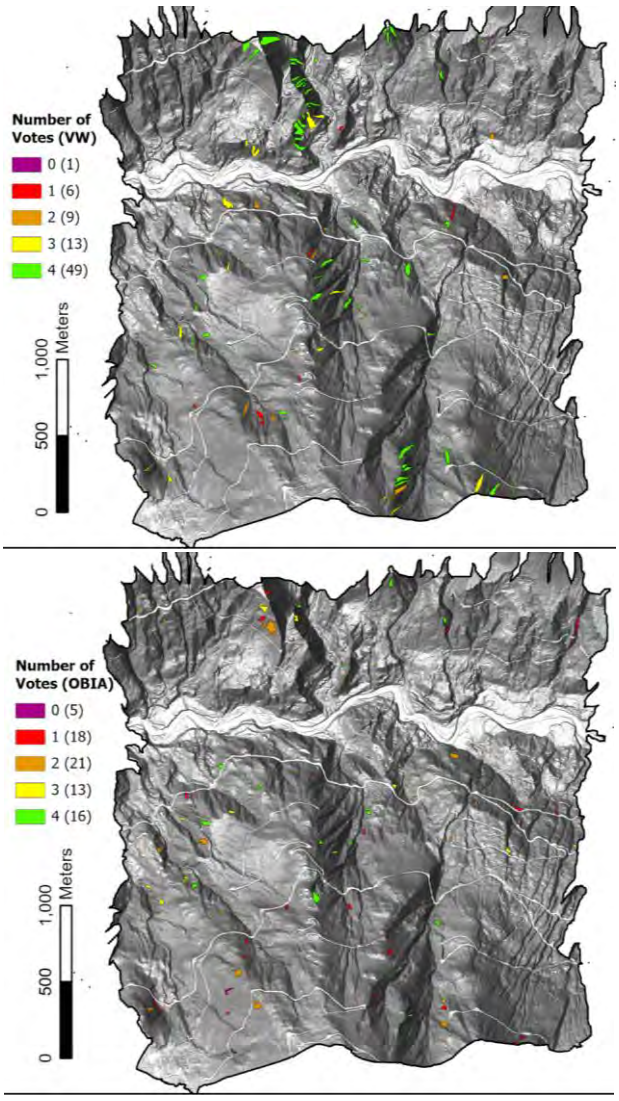


Figure 4544: Number of yes-votes received for each polygon in the potential bedrock hollow sets (VW top and OBIA bottom) for Site 3 and how many polygons received that number of votes.

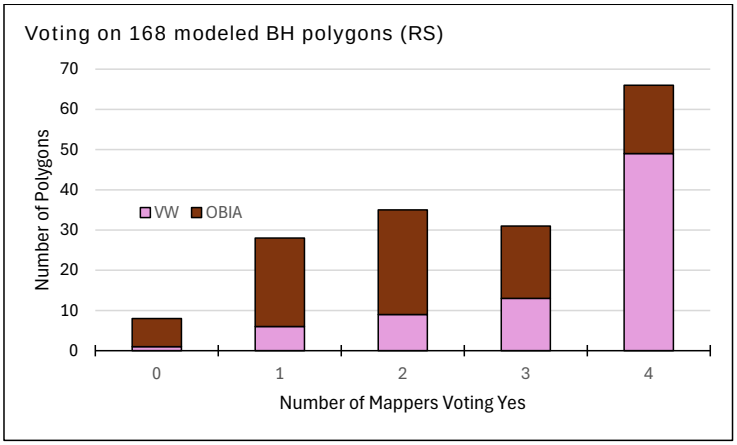


Figure 4645: Distribution of mapper agreement on RS designation votes. The x-axis represents the number of mappers who voted "yes" for a given polygon, while the y-axis shows the count of polygons that received each corresponding number of yes-votes.

Inner gorges

The voting results for Site 3 potential inner gorge polygons for each model are shown in Tables 2~~1~~2 & 2~~2~~3. The mapper's averaged votes indicated they agreed 94% of the 88 selected VW polygons represented a reasonable screen for an inner gorge. Of that 94%, they considered 63% to be correctly drawn.

Table 22: Voting results for VW potential inner gorge polygons at Site 3.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/88)	%Yes (CD/RS)	%Yes (CD/88)
M1	88	79	100%	90%	90%
M2	88	80	100%	91%	91%
M3	85	52	97%	61%	59%
M4	69	8	78%	12%	9%

Of the OBIA potential inner gorge polygons, the mapper's averaged votes indicated they agreed 88% of the 80 selected polygons represented a reasonable screen for an inner gorge, 68% of which they considered to be correctly drawn.

Table 23: Voting results for OBIA potential inner gorge polygons at Site 3.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/80)	%Yes (CD/RS)	%Yes (CD/80)
M1	80	55	100%	69%	69%
M2	72	56	90%	78%	70%
M3	68	49	85%	72%	61%
M4	61	32	76%	52%	40%

The mappers gave unanimous yes-votes on 125 (74%) of the 168 total potential inner gorge polygons produced by both models (69 VW and 56 OBIA). The mappers did not give any unanimous no-votes. Figure 476 shows the number of yes-votes each polygon in the set received. Figure 487 shows the distribution of mapper agreement on RS designation.

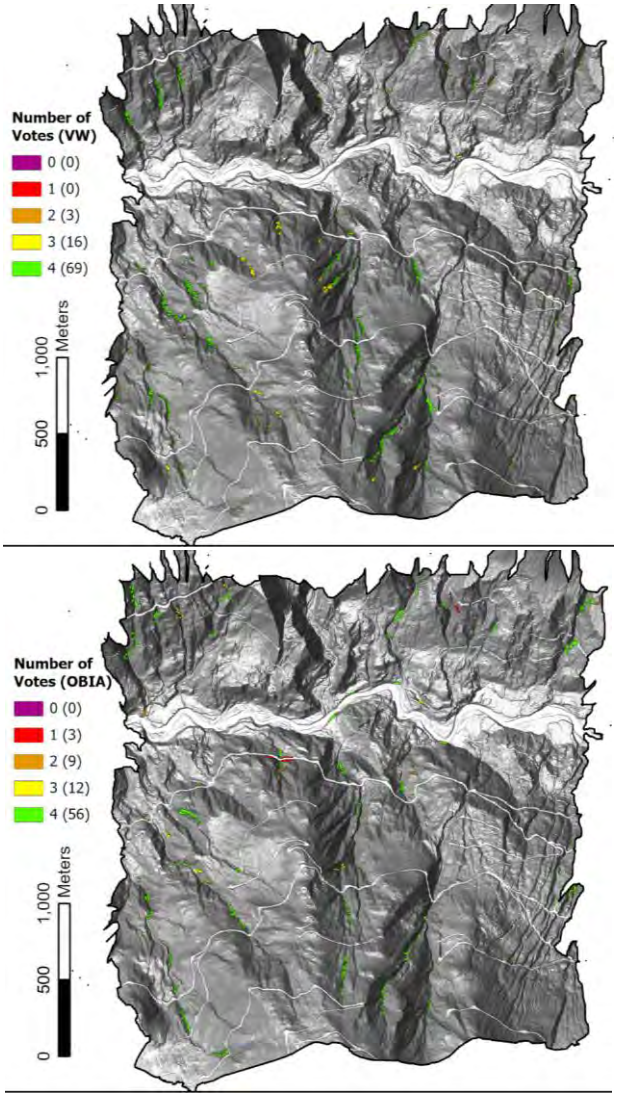


Figure 4746: Number of yes-votes received for each polygon in the potential ~~bedrock hollow~~inner gorge sets (VW top and OBIA bottom) for Site 3 and how many polygons received that number of votes.

Commented [PJ219]: inner gorge?

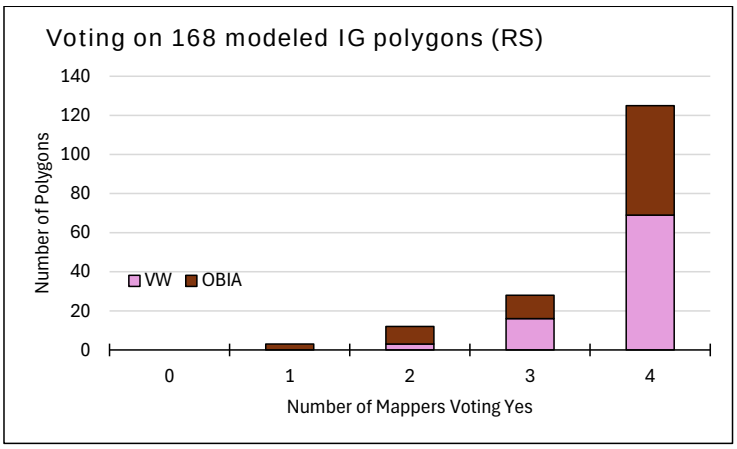


Figure 4847: Distribution of mapper agreement on RS designation votes. The x-axis represents the number of mappers who voted "yes" for a given polygon, while the y-axis shows the count of polygons that received each corresponding number of yes-votes.

Site 4

Bedrock Hollows

The voting results for Site 4 potential bedrock hollow polygons for each model are shown in Tables 243 & 254. The mapper's averaged votes indicated they agreed 88% of the 27 selected VW polygons represented a reasonable screen for a bedrock hollow. Of those 88%, they considered 40% to be correctly drawn.

Table 24: Voting results for VW potential bedrock hollow polygons at Site 4.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/27)	%Yes (CD/RS)	%Yes (CD/27)
M1	27	2	100%	7%	7%
M2	22	15	81%	68%	56%
M3	26	18	96%	69%	67%
M4	20	3	74%	15%	11%

Of the OBIA potential bedrock hollow polygons, the mapper's averaged votes indicated they agreed 75% of the 59 selected polygons represented a reasonable screen for a bedrock hollow, 57% of which they considered to be correctly drawn.

Table 25: Voting results for OBIA potential bedrock hollow polygons at Site 4.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/59)	%Yes (CD/RS)	%Yes (CD/59)
M1	54	18	92%	33%	31%
M2	48	35	81%	73%	59%
M3	46	42	78%	91%	71%
M4	30	9	51%	30%	15%

The mappers voted "yes" unanimously on 43 (50%) of the 86 total potential bedrock hollow polygons produced by both models. Figure 4~~8~~9 shows the number of yes-votes each polygon in the set received. The mappers unanimously agreed that 42 represented a reasonable screen for a bedrock hollow (17 VW and 25 OBIA), and that 1 OBIA polygon did not. Figure 50~~49~~ shows the distribution of mapper agreement on RS designation.

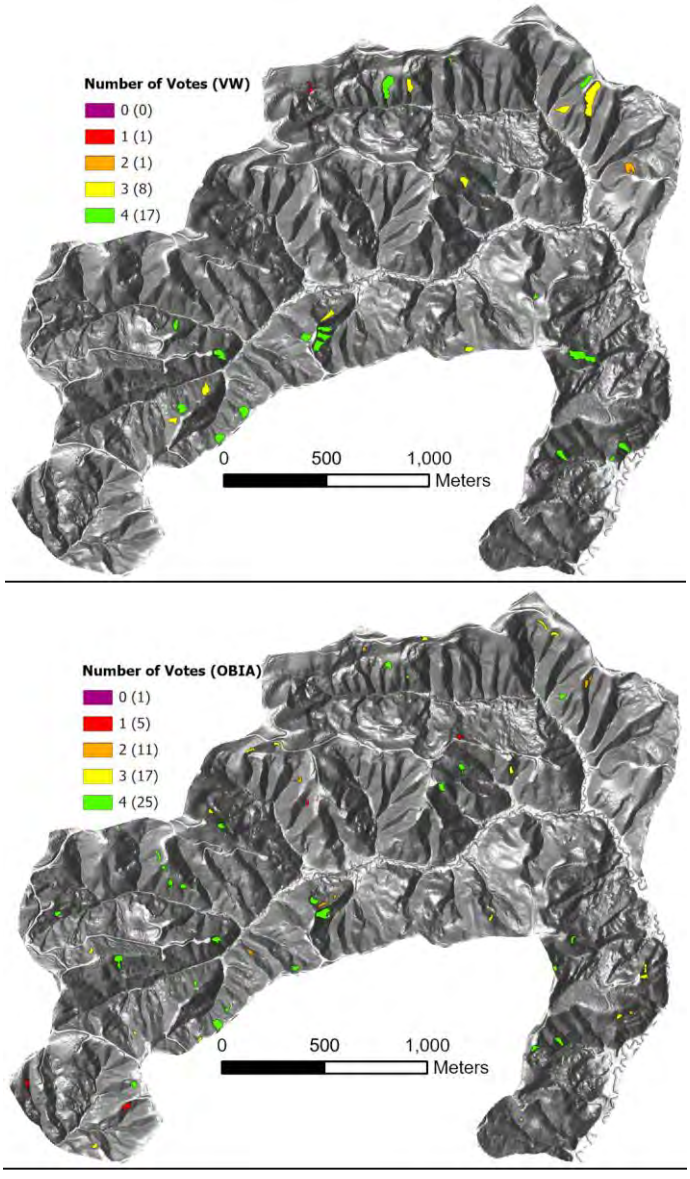


Figure 4948: Number of yes-votes received for each polygon in the potential bedrock hollow sets (VW top and OBIA bottom) for Site 4 and how many polygons received that number of votes.

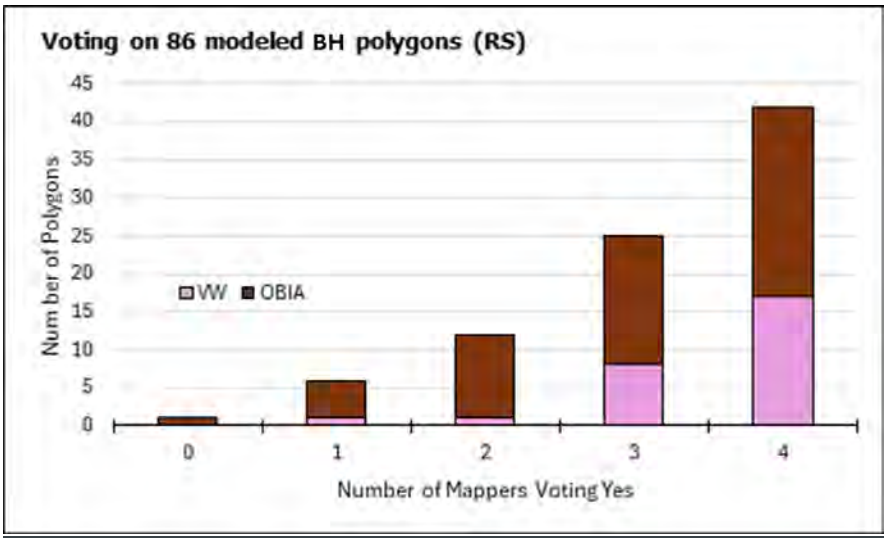


Figure 5049: Distribution of mapper agreement on RS designation votes for Site 4. The x-axis represents the number of mappers who voted "yes" for a given polygon, while the y-axis shows the count of polygons that received each corresponding number of yes-votes.

Inner Gorges

The voting results for Site 4 potential inner gorge polygons for each model are shown in Tables 256 & 276. The mapper's averaged votes indicated they agreed 96% of the 68 selected VW polygons represented a reasonable screen for an inner gorge. Of those 96%, they considered 62% to be correctly drawn.

Table 26: Voting results for VW potential inner gorge polygons at Site 4.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/68)	%Yes (CD/RS)	%Yes (CD/68)
M1	68	61	100%	90%	90%
M2	68	59	100%	87%	87%
M3	68	40	100%	59%	59%
M4	56	6	82%	11%	9%

Of the OBIA potential inner gorge polygons, the mapper's averaged votes indicated they agreed 96% of the 64 selected polygons represented a reasonable screen for an inner gorge, 58% of which they considered to be correctly drawn.

Table 27: Voting results for OBIA potential inner gorge polygons at Site 4.

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/64)	%Yes (CD/RS)	%Yes (CD/64)
M1	63	50	98%	79%	78%
M2	63	33	98%	52%	52%
M3	64	49	100%	77%	77%
M4	55	12	86%	22%	19%

The mappers gave unanimous yes-votes on 110 (83%) of the 132 total potential inner gorge polygons produced by both models (56 VW and 54 OBIA). The mappers did not give any unanimous no-votes. Figure 510 shows the number of yes-votes each polygon in the set received. Figure 521 shows the distribution of mapper agreement on RS designation.

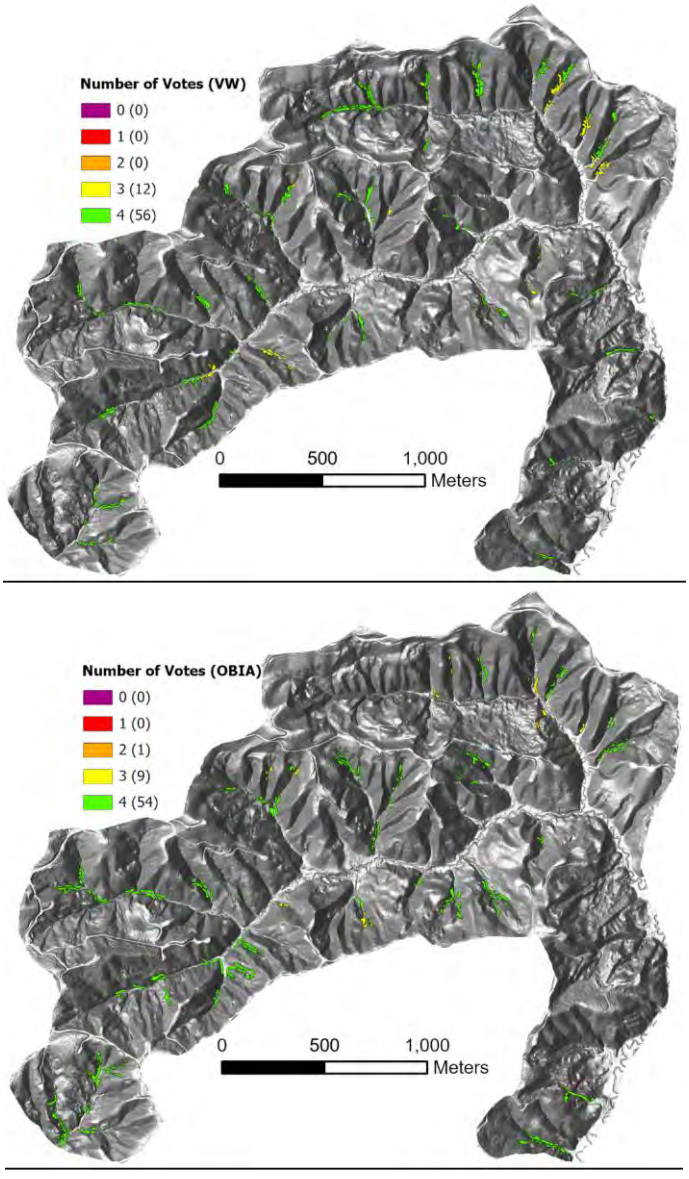


Figure 5150: Number of yes-votes received for each polygon in the potential ~~bedrock hollow~~inner gorge sets (VW top and OBIA bottom) for Site 4 and how many polygons received that number of votes.

Commented [PJ220]: inner gorge?

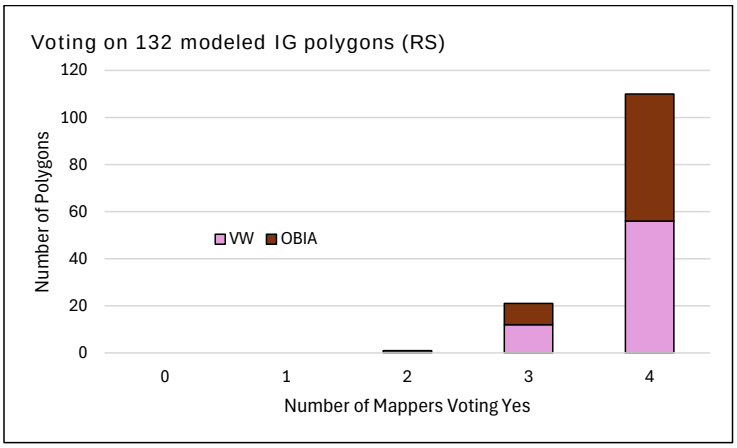


Figure 5251: Distribution of mapper agreement on RS designation votes. The x-axis represents the number of mappers who voted "yes" for a given polygon, while the y-axis shows the count of polygons that received each corresponding number of yes-votes.

4.2.3 PHASE 2 CONCLUSIONS

Phase 2 of this project evaluated the performance of the computer-based models in delineating bedrock hollow and inner gorge landforms across four study areas, using experienced practitioner (EP) assessments as a benchmark. By removing reliance on hand-drawn reference maps and employing structured yes/no voting across defined prompts, the evaluation focused on each model's ability to (1) reasonably screen (RS) for the presence of target landforms and (2) correctly draw (CD) their boundaries.

Results indicated that both models generally performed well in identifying landform presence, particularly for inner gorges. Across all sites, expert mappers judged that 70% or more of the modeled landforms were reasonable screens for either bedrock hollows or inner gorges. Careful examination of comments where the EPs thought the modeled bedrock hollows were invalid shows that a lack of delivery potential (i.e., no way for a landslide from a >70%, concave area to route into the channel network) was a frequent consideration. This delivery issue particularly impacted the OBIA model results because the model does not locate bedrock hollows by finding zero-order basins, and this issue was particularly noticeable in younger topographic settings such as Site 3 where recent glaciation has limited channel dissection.

~~D~~D~~However, d~~elineation accuracy showed reduced model performance and greater variability, with a marked difference between models and landform types. Thirty-nine percent of bedrock hollow polygons and 62% of inner gorge polygons were designated to be correctly drawn (CD). Inner gorge polygons consistently received higher CD ratings than bedrock hollows, and VW model polygons tended to perform slightly better on average for both RS and CD criteria.

Commented [HJ221]: These "correctly drawn" percentages need to be adjusted to represent the number of correctly drawn polygons out of the whole population of polygons.

The variability in correctly drawn ratings highlights areas for model refinement, especially for bedrock hollow delineations. EP comments and site-specific patterns suggest that some modeled polygons failed to capture the necessary geomorphic boundaries or sufficient topographic convergence expected by practitioners. These outcomes suggest targeted improvements to input parameterization, particularly those controlling convergence and planform curvature thresholds are needed to improve the models' abilities to accurately delineate bedrock hollows and inner gorges..

Commented [PJ222]: Was this done?

Overall, Phase 2 demonstrates that automated mapping tools can reliably support initial screening for potential bedrock hollow and inner gorge landforms but require further refinement to improve delineation accuracy. These findings inform the next phase of model development and underscore the importance of iterative evaluation methods.

5.0 Discussion

~~*Potential Causes of Automated and Manual Mapping Differences*~~

~~Our original intent was to evaluate computer-generated landform maps by comparing them with landform maps drawn by experienced practitioners (EPs). Our mapping team consisted of five landslide subject-matter experts. Three members of this team were involved with collecting and analyzing the data initially used to define the current RIL criteria and have each been professionally mapping landforms for forest-practice applications, both remotely and on the ground, for decades. Despite the knowledge and expertise of the EPs, the maps they produced were inconsistent.~~

~~Manual mapping of RILs using DTM basemaps relies on interpretation of topographic attributes measured from the DTMs or inferred from DTM derivatives. As described in Section 3.1.1, mappers used shaded-relief images, slope classes, and contour lines derived from DTMs. Shaded-relief images provide a view of the terrain from which basic landform features and their locations relative to each other are visible: valleys can be distinguished from ridges, and convergent slopes from planar and divergent slopes. To map RILs, a mapper must decide precisely where the RIL criteria are met and draw a line delineating those areas inside and outside of each RIL. Each type of RIL is found in specific landscape locations. For example, hollows occur on steep hillslopes and are separated from the valley floor by steep-sided small channels that serve as debris-flow corridors. Inner gorges occur at the base of steep slopes adjacent to channels (See Table 1 for RIL definitions). These general locations can be determined from the shaded relief. Mappers then use variations in the shaded relief image, the slope classes, and the contour lines to identify where slope thresholds are met, where slope breaks occur, and to determine the degree of hillslope convergence. Their choices are guided and constrained by the RIL descriptions provided in the Board Manual, but each mapper relies on their own unique field observations and experience mapping RILs in making those choices. Hence, each mapper will have a unique understanding and interpretation of the data available from a DTM, which can translate to differences in the RIL polygons drawn by different mappers. Additionally, each mapper brings their own degree of meticulousness, patience, and GIS (digitizing) skills to this task, which can also contribute to differences in RIL polygons drawn by different mappers, or even by the same mapper at different times.~~

~~Perhaps this disagreement arises from the ambiguity and leeway allowed in interpretation of the RIL criteria?. Are Tthe mappers may be including different levels or ranges of topographic attributes within the landform polygons they map.? We looked at frequency distributions of gradient and tangential curvature within the mapped landform polygons (Appendix C2). There is more overlap than difference in the range and shape of these distributions among mappers,~~

Commented [PJ223]: Toward the end of this section, some of the information doesn't relate to "potential causes of mapping differences." Instead it's about next steps. I recommend dividing this section into multiple parts with clear thesis statements to each section

Commented [PJ224]: This is an oversimplification. They are also required to visit the locations in the field.

Commented [HJ225]: Use the board manual definitions of each RIL, abbreviated if necessary.

~~suggesting they are all interpreting the criteria similarly (the differences we do see could, however, be used to suggest more precise RIL criteria).~~

~~Perhaps their lack of agreement arose from differences in mapping protocols?. Through visual examination of bedrock-hollow RIL polygons produced by the four mappers at Site 1A, the mapping team identified potential sources of variability, tightened up their protocol, and then five mappers produced bedrock-hollow RIL maps and four mappers produced inner-gorge RIL maps for a fifth site. Pairwise agreement of these maps was essentially no better than that for the adjacent site prior to tightening of the protocols.~~

~~So why the persistent low level of agreement? The voting exercises showed persistent disagreement among mappers as to what was and was not an RIL. In particular, voting on modeled outliers (Figure 33 showed that every mapper was missing some proportion of the bedrock-hollow RILs. Given the very small proportion of the study sites mapped as RILs (Figure 12), if each of a pair of mappers is including some polygons that are not actually RILs that the others properly did not include, or missing some RILs that the other properly caught, measures of recall and precision will be low, particularly for sites where relatively small numbers of RILs are mapped (i.e., Site 3).~~

~~If mappers are improperly including polygons that do not meet RIL definitions or missing areas that do, that biases comparisons with the computer models. The computers do not include areas that do not meet the definitions specified by the model parameters or miss any RILs that do meet the definitions specified by the model parameters. RIL polygons improperly mapped by a human mapper will drive down measures of recall when compared to a computer-generated map; RIL polygons properly mapped by the computer but missed in the manual map will drive down measures of precision. With the voting exercises, mappers agreed that a majority of the bedrock-hollow sites identified by the computer models as RILs were, in fact, RILs (Figure 34).~~

~~The computers will not draw any polygons where topographic attributes derived from the DTM do not meet RIL criteria specified with the model rules and parameter values. The computers will not miss any polygons where the RIL criteria specified with the model rules and parameters are met. Can we ensure that the specified rules and parameter values are correct?~~

~~With the Susceptibility and Runout sub-project, we will use landslide inventories collected using lidar differencing to determine how measurable terrain elements are related to the density of delivering landslides. We expect to find a suite of terrain elements associated with landslide initiation sites and runout extents. Meaningful terrain elements are likely to exhibit a range of values described by a frequency distribution, like the frequency distributions of gradient and tangential curvature displayed in this report for mapped RIL polygons. They are likely to include factors that provide geomorphic context, such as location of zero-order basins and adjacency to a stream channel. The suite of terrain elements identified with that project and their associated~~

Commented [HJ226]: Yellow - What was the baseline or "truth map" used to make this determination. That needs to be better explained if this statement is going to be retained.

Commented [HJ227]: Probably?

Commented [HJ228]: Yellow - Replace "RIL's" with "a reasonable screen for RIL's."

Commented [HJ229]: Redundant

Commented [HJ230]: The rest of this section appears to address the next sub-project and should be moved to a future work section.

~~frequency distributions will clarify the criteria for identifying high-hazard landforms; the criteria may be specific to regions or lithologies. That is essentially the same approach used initially to define criteria for the current RILs, first by following the Watershed Analysis methodologies (CH-222-22 WAC), and then by following the more formalized Landslide Hazard Zonation (LHZ) Project Protocol (UPSAG, 2006). Here is the analysis procedure specified in the LHZ Protocol (UPSAG, 2006):~~

~~*The analysis comprises a three-step procedure. Each step is discussed in the detailed instructions below.*~~

~~*a. Conduct a Landslide Inventory.*~~

~~*b. Amend the GIS derived Preliminary Landform map. Use the landslide inventory map, statistical analysis of the landslide data and geomorphic information provided by the GIS to create a landform map. Landforms will be differentiated based on slope gradient and shape, lithology, landslide density and sensitivity to forest practices. Specific rule identified high hazard areas will also be individually identified in this process.*~~

~~*c. Set the hazard ratings for landforms. Hazard ratings will be assigned to the landforms based on observed landslide density over time.*~~

~~Look again at the critical question for the USC Project:~~

~~*What modifications to the unstable slopes criteria and delivery assessment methods would result in more accurate and consistent identification of:*~~

~~*i. unstable slopes and landforms,*~~

~~*ii. unstable slopes and landforms sensitive to forest practices related changes in landslide process, and*~~

~~*iii. locations susceptible to impacts from upslope landslides such that an unstable slope and landform conditions where landslide runout would likely have an adverse impact to public resources or a threat to public safety is possible?*~~

~~o We can use the sequence of tasks described with the LHZ protocol to answer these. With the two models, we can generate maps that delineate bedrock hollows and inner gorges and flag those that meet current RIL criteria. As noted, the two computer-generated maps are not exactly the same; we have two starting hypotheses. These maps are generated using topographic attributes derived from the lidar DTMs. Anywhere we run the models, they use the same rules, the same input parameter values, and the same DTM derivatives. The landforms they delineate will be consistent with those rules and parameters, which we set to be consistent with current RIL criteria. We can also incorporate the information gathered from Phase 2 of this project regarding convergence parameters and hydrography. However, those landforms will vary in shape and abundance from~~

Commented [PJ231]: It is not great to add new information to a discussion section. This procedure should be in a different part of the report, either the introduction with the history, or in a recommendations section where you provide key recs/steps for the next phases of the USC project.

Commented [AW232]: Yellow. It is still unclear to me how a landslide inventory will be produced that will cover more than 2 possible landform types? Is there a GIS preliminary landform map will be used to provide "more accurate and consistent identification" of the large range of unstable slopes and landforms where shallow landslides might occur in Washington state?

Commented [HJ233]: Yellow - Were landform maps ever generated for the LHZ Project? If so, why didn't the project team use that methodology to map landforms from the start?

Commented [PJ234]: Please simplify this paragraph. Bullet points?

Commented [HJ235]: Yellow - It is unclear who will "flag" the bedrock hollows and inner gorges that meet current RIL criteria and exactly how that will be done.

Commented [HJ236]: Yellow - It is unclear how landforms that are not bedrock hollows or inner gorges, but have landslides, will be addressed.

Commented [HJ237]: Shouldn't slope be considered as well?

~~region to region depending on the terrain characteristics represented by the DTM. Those are our preliminary maps. We then collect landslide inventories and generate statistics that relate landslide locations and runout extents with topographic attributes measured from the DTMs and other ancillary data. We use those statistics to evaluate those preliminary maps and look for ways to modify the rules and parameters used to generate them that better resolve differences in delivering landslide density among landform types. The closer we keep those rules and parameters to attributes that can be measured or recognized by practitioners on the ground, the more useful they will ultimately be for field-based identification and delineation of high-hazard landforms. The issue is not how well the computer maps match some other maps, but how well this methodology will work to address the critical questions, which will be determined during the implementation of the Susceptibility and Runout sub-project.~~

~~So, do we need expert opinion? We do. Computer models are consistent, so if the input parameters are not quite right or the models themselves do not quite work the way we expect, they produce consistently wrong results. The models rely on com~~

~~puter code written by human developers, who are not infallible. Software has bugs; the implementation may not exactly follow the conceptual model on which it is based. We need ways to check computer model outputs. One way is to use different models to do the same thing and compare the results. We will do that. Another is to look at the results. We will do that.~~

Findings from Phase 1 and Phase 2 of this sub-project demonstrate the challenges and opportunities involved in automating the identification and delineation of potential RILs like bedrock hollows and inner gorges. In Phase 1, we observed ~~substantial~~ variability among experienced practitioners when manually delineating landforms from lidar-derived topographic data. Despite shared guidance and criteria, individual interpretation of terrain features led to ~~low~~unsatisfactory spatial agreement, suggesting that manual mapping alone introduces ~~significant~~ subjectivity. Voting exercises further revealed that mappers ~~often~~ recognized valid landforms in model-generated polygons they had not originally mapped, ~~indicating that manual methods likely omit legitimate RILs.~~

In Phase 2, we addressed these limitations by shifting from manual reference maps to a structured expert evaluation of model-generated polygons. This approach allowed for more objective comparisons across models, focusing separately on a polygon's ability to represent a reasonable landform and whether its delineation matched expected boundaries. Results showed that both the OBIA and VW models performed consistently well in identifying potential RIL locations, especially for inner gorges. However, delineation accuracy, particularly for bedrock

Commented [HJ238]: In my opinion, this kind of banter-like writing style has no place in a technical report. It seems to be most common in this section and should be cleaned up. Aside from that, the salient points here have already been mentioned several times in this report.

Commented [HJ239]: Context is needed if "low" is going to be retained in this sentence.

hollows, was more variable and highlighted specific model limitations related to topographic convergence, slope thresholds, and stream adjacency.

The divergence between manual and automated mapping underscores the importance of refining model parameters and validating them against independent criteria. Models will consistently reproduce results based on their rules, so ensuring that those rules reflect actual geomorphic processes is critical. Phase 2 revealed useful pathways for refinement, such as parameterizing curvature thresholds and adjusting slope criteria, which can now be tested systematically in the Susceptibility & Runout Project (Section 6).

The Susceptibility and Runout sub-project will incorporate expert review of the landform maps used. The Project Team itself can perform that task but also has the option to contract with independent external expert review (e.g., by one or more state certified Qualified Experts). UPSAG has engaged the services of Neptune and Company, Inc., a statistical consulting company, who will provide guidance for the methods used and review of the results obtained for statistical analyses used in that sub-project.

6.0 Future Work

6.1 Automated Mapping Limitations

Reviewing the mapper's comments and answers to prompts 3, 4, & 5 during the Phase 2 exercise helped the team identify some common attributes correlating a no-vote at each of the sites. Identifying these issues provides valuable insights into ways we can tweak the models in the next phase of our project.

Commented [PJ240]: This jumps right to Phase 2.... but I think these points in Section 5.1 are conclusions from the whole project, not just what was learned in Phase 2. Correct?

At all sites, the mappers indicated that both models produced some bedrock hollow polygons that appeared too planar or lacked sufficient convergence. This issue relates to the Board Manual's definition of RILs which states that a category A bedrock hollow is found in a convergent headwaterall or topography. No quantifiable classification of convergence is defined in the Board Manual, so the convergence necessary for an RIL designation is up for individual interpretation and completed with necessary field indicators such as location of emerging water. However, because planform topographic curvature is a geomorphic proxy for shallow groundwater convergence and sediment storage, one remote and field-based interpretation applied to bedrock hollows is that planform curvature allows soil pore water pressures, relative to adjacent low or non-convergent slopes, to increase within the unchannelized colluvium material in the hollow before groundwater exfiltration (seepage) occurs at or near the channel head. As described in section 4.1.2, the models use tangential curvature to set specific convergence

parameters. With the next sub-project, we will collect landslide inventories using lidar differencing. Model thresholds for tangential curvature can then be based on the range of tangential curvatures found for delineated hollows that contain landslides. We will also- look at how curvature varies within the bedrock-hollow polygons. It may be that the relationship between the degree of curvature in the upslope and downslope portions of the polygon are important in distinguishing if there is a routing mechanism. Specific convergence values and the distribution of those values within the polygons can be evaluated relative to the sites with and without mapped landslides in the next project.

Commented [GS241]: Why was this not tackled in this project? The goal of the project was to 'develop automated methods for mapping landforms to be used in subsequent sub-projects.' It seems like tuning the model parameters to best fit mapper expectations fits squarely into this goal.

In a few sites, particularly for the potential inner gorge polygon sets, mappers noted that the polygons were not stream adjacent, or mapped within road prisms adjacent to a road instead of a stream. For example, the model detected some road cut slopes as inner gorges. The VW model used a stream layer produced using the method described for delineating channel heads and flood plains in Appendix B. One way this issue could be addressed in the next project would be for both the VW and OBIA models to use the same stream layer to produce their polygons. Spatially accurate road networks can be used in future classification strategies to minimize false positives caused by model identification of cut slopes.

Commented [PJ242]: seems like a new topic, consider making it a new paragraph.

6.2 CONTIGUOUS LANDFORM MAPPING

With the next sub-project in the USC sequence, Susceptibility and Runout, we will map landslide locations and runout extents using lidar differencing. These mapped landslides will then be overlaid on landform maps, produced by the OBIA and VW models described in this report, to rank landform types by landslide density and the proportion of total mapped landslides occurring within each type.

Although both models are capable of delineating a wide range of landform types, this report focused only on model results for bedrock hollows and inner gorges, as specified by the Study Design. While using a set of human-drawn maps as a baseline resulted in persistent ambiguities, we have demonstrated that the computer models can be adapted by adjusting the input parameters to accommodate a range of landform delineation options. It is this model adaptability that will serve the needs of the next sub-project.

Both the OBIA and VW models delineate landforms using topographic attributes derived from lidar DTMs, including gradient, curvature, contributing area, topographic position, feature size, and the distance between features. These same attributes, referred to as "*terrain elements"* in the Susceptibility and Runout study design, are included in the set of data sources for determining landslide susceptibility. The relationships found between landslide locations, runout extent, and those terrain elements will be quantified in terms of landslide density and proportion.

The same topographic attributes used by the OBIA and VW models to identify and delineate landforms will thus be used in the next study to quantify susceptibility to landslide initiation and runout. We aim to use the observed correlations relating terrain elements with landslide density/proportion to refine current landform definitions and define the criteria for delineating other landform types.

For example, if the range of gradients, curvatures, and contributing areas associated with landslides within bedrock hollows varies regionally, we can adjust the model parameters to accommodate regional trends. Similarly, if a certain proportion of landslides originate on steep slopes that do not meet the criteria for a bedrock hollow or an inner gorge, we can use the terrain-attribute values associated with those landslides to define criteria for identifying and delineating landform types associated with the source areas and runout zones for those landslides.

By associating the terrain elements used by the OBIA and VW models to measures of landslide density and proportion, we will be able to define landform types in terms of a hierarchical range of density and proportion. We will also be able to identify the range and combination of terrain-element values where no landslides are observed. If needed, both the OBIA and VW models can then delineate landforms contiguously across the landscape, ranked in terms of landslide density and proportion.

In this next sub-project, landform mapping will be done solely to evaluate and refine criteria for landform identification and delineation based on landslide density and proportion. ~~There is no intention at this time to use~~ Tthese landform maps are not intended for use as screening tools and have not been designed to perform in that manner. We will, however, strive to use terrain elements that, although measured remotely, have a clear association with ground-based measurements. For example, we can use upslope facing and downslope facing DTM measures of gradient over a specified slope distance, similar to how gradient is measured on the ground using a clinometer (as described in Appendix E).

6.3 Model Sensitivity

Even among experienced experts using the same data source and landform definitions, we found substantial variation in the number, size, and shape of landform polygons drawn on their maps. This likely reflects –an inherent level of uncertainty and ambiguity in how terrain features associated with landslide occurrences are interpreted. We anticipate that the landslide locations identified using lidar differencing will also be associated with a range of terrain-element values, introducing similar variability into any quantitative model that relies on terrain elements to identify and delineate landforms. This variability creates uncertainty about landform boundary delineation.

By ranking landforms according to landslide density and the proportion of total landslides they contain, we can quantify that uncertainty. As we have shown, adjusting parameter values used by the models can alter the number, location, shape, and size of delineated landforms, which affects the landslide density and proportion calculated for each landform type. We can seek a set of optimum parameter values to maximize the landslide density and proportion associated with some landforms while minimizing the density and proportion associated with others. We can also look at how the density and proportion change as we vary the parameter values.

As an example, suppose we determine that landslides within bedrock hollows are found on gradients ranging from 60% upwards, with most occurring on gradients of 70% or more. We could evaluate how changing the minimum gradient threshold for landforms classified as bedrock hollow (e.g., from 55%, to 60%, to 65%, to 70% and so on) affects the landslide density and proportion. If the resulting density and proportion change little as we adjust this parameter, then the model is not sensitive to this parameter and an approximation would suffice. However, if the resulting density and proportions change significantly, then we need to carefully consider the parameter value to use.

We will use model sensitivity to specific terrain elements and threshold values to guide the development and refinement of criteria for identifying and delineating landform types. This approach will help ensure that model outputs reflect meaningful geomorphic patterns while accounting for the natural variability present in landslide-prone terrain.

6.4 Susceptibility and Runout Implementation Overview

The Study Design details the implementation of the next sub-project in the USC sequence, *Empirical Evaluation of Shallow Landslide Susceptibility, Frequency, and Runout by Landform*, which is summarized here. Because this sub-project did not fulfill Objective 2, this overview is intended to demonstrate how the objective will be achieved in the subsequent sub-project. Potential field verification and consultation of independent statisticians and QEs will be threaded throughout the implementation steps of the Susceptibility and Runout project.

1. *Collect a landslide inventory*. Using lidar differencing to obtain this inventory, landslide locations can be precisely correlated with pre-landslide terrain features. We will start with the "Post-Mortem" study area because it has a field-based inventory of landslides associated with the December 2007 storm. We can compare the inventory constructed with lidar differencing to the field-based inventory to determine what level of confidence to place in the lidar-based inventory (e.g., what is the smallest landslide surface area or volume that can be detected?).
2. *Find statistics that relate landslide density to terrain features*. These statistics can be calculated from the frequency distribution of values. We will include terrain features that are measurable by a field practitioner. Likewise, all the chosen terrain features must serve

as inputs to the automated landform-mapping models presented here. Features with field-based analogs include topographic indices such as upslope- and downslope-looking gradient and differences in those gradients (indicating a break-in-slope), proximity to a stream, location relative to valley floor and ridgetop (topographic position index), slope length from the ridge top and slope length from the valley floor. We have not listed all possibilities here, but these and others can be directly measured for observed landslides and incorporated into future iterations of the landform models. There are a variety of topographic measures that can be obtained from a DEM but are not readily measured on the ground. These include upslope contributing area and curvatures. We may find that these are important indicators of landslide location and will therefore want to include them in landform criteria. As discussed in this report, these attributes can form an important component in a QE's determination of a landform location and boundary. For example, convergent curvature is a key component of a bedrock hollow landform that a mapper keys into, but the criteria do not specify any curvature values and there are no simple methods for measuring curvature directly on the ground.

3. *Assign each inventoried landslide to a landform category in a preliminary list of candidate landform type*s. This assignment will be derived directly from terrain elements and then confirmed by a human. Terrain elements (e.g., gradient, curvature, contributing area, topographic position index, channel-network hydrography, and distance to a channel), are a collection of lidar-derived data that will be used to describe and hopefully classify distinct areas on the landscape that are prone to landslide initiation, runout and/or deposition. For example, a landslide may have terrain elements that place it in a steep, convergent zone in the upper third of the relief between valley and ridge top; this should lie within a bedrock hollow, and we can see on a map product (such as a lidar hillshade) if it does or does not. A landslide adjacent to a stream below a slope break would be associated with inner gorges. A landslide on an 80% planar slope would be associated with steep, planar slopes. A landslide on a 60% planar slope would be associated with moderately steep planar slopes. The number and type of landforms needed in this preliminary sorting of landslide types depends on the diversity of terrain features found for the inventoried landslides and includes a degree of judgement by the Project Team. The preliminary list of landform types will be based on the extensive history of landslide and landform mapping in Washington State and elsewhere. The types of mass-wasting-map units used for Watershed Analyses and the Landslide Hazard Zonation Project provide a starting point – and by "mass-wasting-map units" we more broadly include basic landforms that cover the entire landscape (e.g., ridgetops, moderate gradient hillslopes, valley bottoms which are low hazard areas). We favor lidar derivatives that can be compared to field measurements, but do not limit our choices only to such features.

4. *Produce a set of preliminary, computer-generated landform maps that identify potential high-hazard sites based on current RIL criteria*. Our current models provide these preliminary maps of the RILs of interest – bedrock hollows and inner gorges. We can compare these maps to the landslide inventory in two ways: 1) overlay the inventory polygons on the landform polygons, and 2) compare the frequency distribution of terrain-element values within the RIL polygons to those found for the landslide inventory polygons. The landslide polygons are precisely defined by the LCD results. The RIL polygons will be designed to meet the variable opinions of a group of individuals. Experts' interpretation of RIL criteria will be tabulated and the relating model parameters and thematic classification to which the interpretations are correlated will be clearly stated. We will work with independent statisticians and qualified experts to report this correlated data. We will also work with them to design an expert voting exercise to be used in this step to evaluate the modeled landforms in terms of current RIL descriptions.

5. *Overlay the inventoried landslide locations on the delineated landforms.* This will show if there are landform types that we have not included in the computer models where inventoried landslides fall outside the landform boundaries with which they are associated ~~with~~. Multiple iterations of landform polygons will be created by adjusting modeled parameter values such as slope threshold. These landform polygons will be used as spatial overlays with the landslide polygons. Using this overlay-approach, we can adjust the landform rules to increase landslide density (area of landslides/area of landform). In some areas, this approach might lead to larger landforms while in others it may lead to smaller. We will see which landslides fall outside of the RIL polygons and will see the location of those landslides relative to the RIL polygons. This will allow us to see if changes to the landform criteria applied to the models can increase the size and shape of the modeled polygons in such a way as to increase the proportion of landslides included in the modeled RIL polygons. Likewise, we will see if there might be changes to the rules to decrease the size of the modeled RIL polygons without reducing the proportion of landslides they contain.
6. Test and amend the preliminary maps using the results from Step 5 and the statistics gained from the landslide inventory. For example, the landslides will be associated with a range of topographic indices, such as gradient, curvature, contributing area, proximity to a stream, etc. The landform polygons will encompass a range of the same topographic indices. We can compare the frequency distributions of values found for landslides to the frequency distributions of values encompassed within the mapped landform polygons and adjust the computer model inputs to ensure the modeled landforms encompass the same range of attributes. We can group these into unique combinations of terrain-element values using principal component analyses or similar methods (e.g., Netra Regmi et al., 2017). These

combinations can be used to define rules for delineating non-RIL landforms that contain inventoried landslides. We will repeat a similar voting exercise to that used in the previous step; however, instead of relating the classifications to existing RILs they will be evaluated for the feasibility of the new suite of landforms to be distinctly identified in the field.

7. Steps 5 and 6 will provide rule sets for the models so that the delineated landslides are contained within the appropriate landforms based on the landform type assigned to each landslide according to terrain elements. By overlaying the landslide inventory on the mapped landforms, we can calculate landslide density for each landform type and the proportion of all landslides occurring within each landform type. We will then look at how sensitive the densities and proportions are to variations in the rules used to delineate the landforms. For example, the set of rules will probably include threshold values for slope gradient. As we vary the threshold gradient, we will see how landslide density and proportion change and at what threshold values some landslides are no longer contained within the landform they should be associated with. This exercise provides several types of information. We can try to optimize the combination of threshold values to best discriminate low from high-density landforms. We can also see how important our choice of threshold is to measures of density and proportion. Importantly, we can see if there is some small proportion of landslides that have an outsized influence on measures of density. For example, imagine an inventory of bedrock-hollow landslides for which 98 fall within a range of gradients between 80 and 120%, but there are an additional 2 that occur at 65%. We may decide that those 2 should be assigned to a different landform or fall below a density of concern if there are lots of bedrock hollows that measure 65% but rarely fail.
8. *Review of the resulting sets of landform maps and the landslide inventory by experienced practitioners*. Based on results of this landform-mapping sub-project, we expect that there is no single best map that everyone would agree on and that there is a range of landform shapes and extents that would be acceptable. The sensitivity analysis described in Step 7 seeks to address that range. We still need review by experienced human eyes to ensure that the computer models are producing reasonable results. Reviews will be done as team exercises because we expect that there are insights to be gained through group discussions. We also expect this will be an iterative exercise; that evaluation of the modeled landforms relative to the observed landslides will result in further adjustment to the rules applied for the computer models.

7.0 CONCLUSIONS

While the models produced results that are broadly consistent with expert-drawn maps, ~~substantial variation remains both~~there is variation between model outputs and among expert

interpretations. This variation could highlights the need for clearer, more quantifiable landform definitions and underscores the limitations of using subjective reference maps to evaluate model performance. The project's findings demonstrate that model sensitivity to rule parameters can significantly influence the number, size, and shape of mapped landforms, which in turn affects any derived metrics like landslide density. Future work will apply statistical comparisons with mapped landslides to evaluate and refine classification rules.

Through the development and evaluation of the OBIA and VW models, we demonstrated foundational approaches for ~~delineating~~mapping the RILs specified in the Study Design, namely inner gorges and bedrock hollows, based on terrain attributes and rule-based classification. The models successfully generated interpretable and repeatable landform maps, satisfying the first project objective. Based on the results of Phase 2, the models reliably identify these two distinct landforms. They do this using rule sets unique to each landform type, demonstrating the ability of the models to reliably differentiate different types of landforms. As described in Section 1.3, with the next subproject, additional rulesets will be defined to delineate other landform types based on the landslide locations found with lidar differencing. Just as the models proved capable of adjustment to better match the expectations of expert reviewers, adjustments will be used to determine how sensitive the models are to the rules applied. Uncertainty associated with input parameters will be explored via model sensitivity analysis (described in Section 6.3) in the next project phase.

~~Automated models offer consistency and scalability, but their effectiveness hinges on the specification of input parameters—parameters that are necessarily grounded in human interpretation. Our review of mapper feedback and topographic attribute distributions provided actionable insights into where and why models produce inconsistent outputs, offering a roadmap for refining these methods in the next phase of this project. The upcoming Susceptibility and Runout sub-project provides a critical opportunity to ground automated methods in empirical evidence by linking mapped landforms to actual landslide occurrences.~~

~~The results suggest that, with clear definitions and calibrated rulesets, these automated methods offer scalable and consistent methods for landform delineation across diverse landscapes and thus provide a baseline geomorphic context from which to evaluate landslide susceptibility and runout in the next sub-project.~~

~~Ultimately, the results of this project satisfy the objectives stated in the Study Design (Section 1.3) to identify methods for consistent automated delineation of landforms. Furthermore, based on the results of a voting exercise and comparison of topographic characteristics within manually~~

Commented [HJ243]: Need some mention of the Objective Rewrite here.

Commented [JM244]: Will consistent output from the models produce better forest practices on the ground?

Commented [HJ245]: Yellow - The report did not present a clear roadmap as stated here.

Commented [HJ246]: Red - This report has not clearly demonstrated that.

Commented [JM247R246]: I agree.

Commented [AW248]: Red. This report does provide useful information on the use of the manually mapped LiDAR to automate a landform inventory. Please further clarify how this will answer the USC critical question for this project at scale.

~~and computer-mapped landforms, the automated mapping models developed in this report do so in a way that adequately replicates manually mapped landforms for use as a screen but not for use as definitive landform boundaries (Section 4.1.4). The automatically mapped landforms will provide baseline geomorphic context from which to evaluate landslide susceptibility and runout. To account for uncertainty associated with input parameters, uncertainty in that geomorphic context will be explored via model sensitivity analysis (described in Section 5.2) in the next project phase.~~

~~With that next project phase, we expect to find landslides located outside of delineated bedrock hollows and inner gorges. Attributes of those locations will be used to identify the parameters and threshold values associated with other landslide-prone terrain. Those parameters and threshold values will point to potential modifications to the rules used to identify and delineate bedrock hollows and inner gorges and to other potential unstable landform definitions.~~

The information assembled to fulfill these objectives along with the deliverables provided in this report collectively provide the basis for the next phase of this work. Using expert information to inform the plausible range of values and probability distributions for the input parameters for the automated methods can allow for the development of probabilistic output. In this context, machine-learning based sensitivity analysis algorithms can be implemented to defensibly identify which parameters are most important with respect to explaining the observed variance in key aspects of the output from the automated methods. Once the most important parameters are identified through sensitivity analysis, the greatest return for additional effort regarding refinement of input parameter distributions can be clearly demonstrated. The use of subject matter information to constrain reproducible computer-based methods for landform identification represents the best of both the human-based and computationally-based approaches.

Commented [GS249]: In my opinion, this is the most important finding from the report and should be highlighted better throughout.

Commented [HB250]: A brief reminder of what this screen will do would be helpful here.

Commented [PJ251]: automated, not automatically

Commented [HJ252]: Yellow - How will landslide densities for landforms other than bedrock hollows and inner gorges be calculated without delineating discrete boundaries for other landforms?

Commented [DM253R252]: These computer models can delineate discrete boundaries for other landforms once we specify the rules to use for identifying those landforms and their boundaries. The landslide locations found with the next project phase will provide those rules.

Commented [HB254]: Is this what the screen will do?

Commented [HB255R254]: Is the information referred to in the previous sentence or something else?

Commented [HJ256]: Yellow - The deliverables of this sub-project should be clearly summarized.

8.0 LITERATURE CITED

Anderson, C. J., 2019, Comparing the Slope–Area Threshold for Stream Initiation in Primeval and Managed Forests of Northern Michigan: Forest Science.

Avcioglu, B., Anderson, C. J., and Kalin, L., 2017, Evaluating the Slope-Area Method to Accurately Identify Stream Channel Heads in Three Physiographic Regions: JAWRA Journal of the American Water Resources Association, v. 53, no. 3, p. 562-575.

Barquin, J., Benda, L. E., Villan, F., Brown, L. E., Bonada, N., Vietes, D. R., Battin, T. J., Olden, J. D., Hughes, S. J., Gray, C., and Woodward, G., 2015, Coupling virtual watersheds with ecosystem services assessment: a 21st century platform to support river research and management: WIREs Water, v. 2, no. 6, p. 609-621.

Benda, L., Miller, D., Barquin, J., McCleary, R. J., Cai, T., and Ji, Y., 2016, Building virtual watersheds: a global opportunity to strengthen resource management and conservation: Environmental Management, v. 57, p. 722-739.

Benda, L., Miller, D. J., Andras, K., Bigelow, P., Reeves, G. H., and Michael, D., 2007, NetMap: A new tool in support of watershed science and resource management: Forest Science, v. 53, no. 2, p. 206-219.

Blaschke, T., 2010, Object based image analysis for remote sensing: ISPRS Journal of Photogrammetry and Remote Sensing, v. 65, no. 1, p. 2-16.

Blaschke, T., and Strobl, J., 2003, Defining landscape units through integrated morphometric characteristics: Landscape modelling: Digital techniques for landscape architecture, p. 104-113.

Booth, A. M., Roering, J. J., and Perron, J. T., 2009, Automated landslide mapping using spectral analysis and high-resolution topographic data: Puget Sound lowlands, Washington, and Portland Hills, Oregon: Geomorphology.

Bue, B. D., and Stepinski, T. F., 2007, Machine Detection of Martian Impact Craters From Digital Topography Data: IEEE Transactions on Geoscience and Remote Sensing, v. 45, no. 1, p. 265-274.

Cochran, W.G., 1977, Sampling Techniques. 3rd Edition, John Wiley & Sons, New York.

Clarke, S. E., Burnett, K. M., and Miller, D. J., 2008, Modeling streams and hydrogeomorphic attributes in Oregon from digital and field data: Journal of the American Water Resources Association, v. 44, no. 2, p. 459-477.

Clubb, F. J., Mudd, S. M., Milodowski, D. T., Hurst, M. D., and Slater, L. J., 2014, Objective extraction of channel heads from high-resolution topographic data: Water Resources Research, v. 50, p. 4283-4304.

Dieu, J. and Shelmerdine, W. 1997. Module A - Sedimentation Assessment. In: North Fork Calawah Watershed Analysis. Rayonier Timberlands Operating Company in cooperation with USDA Forest Service Olympic National Forest and Pentec Environmental, September 5, 1997. Hoquiam, WA. 38 pp. plus appendices.

Dieu, J. and Toth, S.E., 2009. A History and Evaluation of Washington State's Forest Practices Rules for Unstable Slopes: Geological Society of America Abstracts with Programs. Vol. 41, No. 7, p.254.

Dieu, J., Miller, D., Turner, T. and Stewart, G., 2018, Unstable Slopes Criteria Project: Study Design for the Object-based Mapping with High-Resolution Topography, Cooperative Monitoring Evaluation and Research Report. CMER 2018.01.23, Washington State Department of Natural Resources, Olympia, WA.

Dietrich, W.E., Wilson, C.J. and Reneau, S.L. (1986). "Hollows, colluvium, and landslides in soil-mantled landscapes". In Abrahams, A.D. (ed.), Hillslope Processes. London: Allen & Unwin, pp. 361-88

Divakar, S., 2021, Determination of Sample Size and Sampling Methods in Applied Research, Proceedings on Engineering Services, v. 3, p. 25-32.

D'Oleire-Oltmanns, S., Marzolff, I., Tiede, D., and Blaschke, T., 2014, Detection of Gully-Affected Areas by Applying Object-Based Image Analysis (OBIA) in the Region of Taroudannt, Morocco: Remote Sensing, v. 6, no. 9, p. 8287-8309.

Dai, C., Zhang, Z., and Lin, D., 2020, An Object-Based Bidirectional Method for Integrated Building Extraction and Change Detection between Multimodal Point Clouds: Remote Sensing, v. 12, no. 10.

Darwish, A., Leukert, K., and Reinhardt, W., 2003, Image segmentation for the purpose of object-based classification: IGARSS, v. 2003, p. 2039-2041.

Dietrich, W. E., and Dunne, T., 1993, The channel head, *in* Beven, K., and Kirkby, M. J., eds., Channel Network Hydrology: New York, John Wiley & Sons, Ltd.

Dikau, R., 1989, The application of a digital relief model to landform analysis in geomorphology, *in* Raper, J., ed., Three dimensional application in Geographical Information Systems: London, Taylor & Francis, p. 51-77.

Dragovich, J. D., Logan, R. L., Schasse, H. W., Walsh, T. J., Lingley, W. S., Norman, D. K., Gerstel, W. J., Lapen, T. J., Schuster, E., and Meyers, K. D., 2002, Geologic Map of Washington - Northwest Quadrant.

Drăguţ, L., and Blaschke, T., 2006, Automated classification of landform elements using object-based image analysis: Geomorphology, v. 81, no. 3-4, p. 330-344.

Dragut, L., and Eisank, C., 2012, Automated object-based classification of topography frm SRTM data: Geomorphology (Amst), v. 141-142, no. 4, p. 21-33.

Dunne, T., 1991, Stochastic aspects of the relations between climate, hydrology and landform evolution: Transactions of the Japanese Geomorpphological Union, v. 12, p. 1-24.

~~Etzelmüller, B., and Sulebak, J. R., 2000, Developments in the use of digital elevation models in periglacial geomorphology and glaciology: Physische Geographie, v. 41, p. 35-58.~~

Evans, I. S., 2003, Scale-specific landforms and aspects of the land surface, *in* Evans, I. S., Dikau, R., Tokunaga, E., Ohmori, H., and Hirano, M., eds., Concepts and Modelling in Geomorphology: International Perspectives: Tokyo, TERRAPUB, p. 61-84.

Feizizadeh, B., Kazemi Garajeh, M., Blaschke, T., and Lakes, T., 2021, An object based image analysis applied for volcanic and glacial landforms mapping in Sahand Mountain, Iran: Catena, v. 198.

Franklin, Jerry F., Dyrness, C.T. 1973. Natural vegetation of Oregon and Washington. Gen. Tech. Rep. PNW-GTR-008. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Florinsky, I. V., 2016, Digital Terrain Analysis in Soil Science and Geology, Acedemic Press, Elsevier, 486 p.:

Garajeh, M. K., Feizizadeh, B., Blaschke, T., and Lakes, T., 2022, Detecting and mapping karst landforms using object-based image analysis: Case study: Takht-Soleiman and Parava Mountains, Iran: The Egyptian Journal of Remote Sensing and Space Science, v. 25, no. 2, p. 473-489.

Garrett, K. K., and Wohl, E. E., 2017, Climate-invariant area–slope relations in channel heads initiated by surface runoff: Earth Surface Processes and Landforms, v. 42, no. 11, p. 1745-1751.

Gerçek, D., and Uğur, Z., 2010, Object-based classification of landscape into land management units (LMUS), Third International Conference on Geographic Object-Based Image Analysis (GEOBIA), Volume 29.

Hattanji, T., Kodama, R., Takahashi, D., Tanaka, Y., Doshida, S., and Furuichi, T., 2021, Migration of channel heads by storm events in two granitic mountain basins, western Janap: Implication for predicting location of landslides: Geomorphology, v. 393.

Henriksen, D.A., 1956, Eocene stratigraphy of the lower Cowlitz River-eastern Willapa Hills area, southwestern Washington: Washington Division of Mines and Geology Bulletin 43, 122 p.

Hyde, K. D., Wilcox, A. C., Jencso, K., and Woods, S., 2014, Effects of vegetation disturbance by fire on channel initiation thresholds: Geomorphology, v. 214, p. 84-96.

Ijjasz-Vasquez, Ede J., and Bras, R. L., 1995, Scaling regimes of local slope versus contributing area in digital elevation models: Geomorphology, v. 12, p. 299-311.

Ilich, A. R., Misiuk, B., Lecours, V., and Murawski, S. A., 2023, MultiscaleDTM: An open-source R package for multiscale geomorphometric analysis: Transactions in GIS, v. 27, no. 4, p. 1164-1204.

Jenness, J., 2013, DEM Surface Tools, Jenness Enterprises.

Justice, T. E., 2021, Evaluation of Manual and Semi-Automated Deep-Seated Landslide Inventory Processes: Willapa Hills, Washington: Portland State University, 110 p.

Kresch, D. L., 1998a, Determination of upstream boundaries on western Washington streams and rivers under the requirements of the Shoreline Management Act of 1971, Water-Resources Investigations Report 96-4208: Tacoma, WA, U.S. Geological Survey.

Kresch, D. L., 1998b, Determination of upstream boundary points on northeastern Washington streams and rivers under the requirements of the Shoreline Management Act of 1971: U.S. Geological Survey.

Laprade, W.T. 1994. Appendix A – Mass Wasting Assessment. In: Stillman Creek Watershed Analysis. Shannon & Wilson in cooperation with Weyerhaeuser Company & Geomatics International Corp., September 29, 1994. Curtis, WA.

Lashermes, B., Foufoula-Georgiou, E., and Dietrich, W. E., 2007, Channel network extraction from high resolution topography using wavelets: Geophysical Research Letters, v. 34, no. L23S04, p. 6.

Li, J., Li, T., Zhang, L., Sivakumar, B., Fu, X., Huang, Y., and Bai, R., 2020, A D8-compatible high-efficient channel head recognition method: Environmental Modelling & Software, v. 125.

Li, X., Cheng, X., Chen, W., Chen, G., and Liu, S., 2015, Identification of Forested Landslides Using LiDar Data, Object-based Image Analysis, and Machine Learning Algorithms: Remote Sensing, v. 7, no. 8, p. 9705-9726.

Lin, S., Xie, J., Deng, J., Qi, M., and Chen, N., 2022, Landform classification based on landform geospatial structure – a case study on Loess Plateau of China: International Journal of Digital Earth, v. 15, no. 1, p. 1125-1148.

Logan, R. L., 2003, *Geologic map of the Shelton 1:100,000 quadrangle, Washington*: Washington Division of Geology and Earth Resources, Open File Report 2003-15, scale 1:100,000.

Magirl, C. S., and Olsen, T. D., 2009, Navigability potential of Washington rivers and streams determined with hydraulic geometry and a geographic information system: US Geological Survey, Scientific Investigations Report 2009-5122.

Maxwell, A. E., and Shobe, C. M., 2022, Land-surface parameters for spatial predictive mapping and modeling: Earth-Science Reviews, v. 226.

McCleary, R. J., Hassan, M. A., Miller, D., and Moore, R. D., 2011, Spatial organization of process domains in headwater drainage basins of a glaciated foothills region with complex longitudinal profiles: Water Resources Research, v. 47, no. W05505.

Metes, M. J., Jones, D. K., Baker, M. E., Miller, A. J., Hogan, D. M., Loperfido, J. V., and Hopkins, K. G., 2022, Ephemeral Stream Network Extraction from Lidar-Derived Elevation and Topographic Attributes in Urban and Forested Landscapes: JAWRA Journal of the American Water Resources Association, v. 58, no. 4, p. 547-565.

Miller, D., Benda, L., DePasquale, J., and Albert, D., 2015, Creation of a digital flowline network from IfSAR 5-m DEMs for the Matanuska-Susitna Basins: a resource for update of the National Hydrographic Dataset in Alaska.

Miller, D. J., 2003, Programs for DEM Analysis, Landscape Dynamics and Forest Management, General Technical Report RMRS-GTR-101CD: Fort Collins, CO, USA, USDA Forest Service, Rocky Mountain Research Station.

Miller, D. J., and Burnett, K. M., 2007, Effects of forest cover, topography, and sampling extent on the measured density of shallow, translational landslides: Water Resources Research, v. 43, p. 23.

Miller, D. J., and Burnett, K. M, 2008, A probabilistic model of debris-flow delivery to stream channels, demonstrated for the Coast Range of Oregon, USA: Geomorphology, v. 94, p. 184-205.

Montgomery, D. R., 1999, Process domains and the river continuum: Journal of the American Water Resources Association, v. 35, no. 2, p. 397-410.

Montgomery, D. R., and Foufoula-Georgiou, E., 1993, Channel network source representation using digital elevation models: Water Resources Research, v. 29, no. 12, p. 3925-3934.

Orlandini, S., and Moretti, G., 2009, Determination of surface flow paths from gridded elevation data: Water Resources Research, v. 45, no. W03417.

Orlandini, S., Moretti, G., Franchini, M., Aldighieri, B., and Testa, B., 2003, Path-based methods for the determination of nondispersive drainage directions in grid-based digital elevation models: Water Resources Research, v. 39, no. 6.

Othus, S. and Parks, D., 2009, Lower Wishkah Watershed, Grays Harbor Counry, Washington: Landslide Hazard Zonation Project, Mass Wasting Assessment, Washington State Department of Natural Resources.

Pater, D.E, Bryce, S.A, Thorson, T.D, Kagan, J., Chappell, C., Omernik, J.M, Azevedo, S.H., and Woods, A.J., 1998, Ecoregions of Western Washington and Oregon: U.S. Geological Survey.

Pazzaglia, F.J., & Brandon, M.T., 2001, A Fluvial Record of Long-term Steady-state Uplift and Erosion Across the Cascadia Forearc High, Western Washington State: American Journal of Science, 301(4-5), 385-431.

Petras, V., Petrasova, A., McCarter, J. B., Mitasova, H., and Meentemeyer, R. K., 2023, Point Density Variations in Airborne Lidar Point Clouds: Sensors (Basel), v. 23, no. 3.

PRISM Climate Group, 2024: https://prism.oregonstate.edu/normals/

Płaczkowska, E., Cebulski, J., Bryndza, M., Mostowik, K., Murawska, M., Rzonca, B., and Siwek, J., 2021, Morphometric analysis of the channel heads based on different LiDAR resolutions: Geomorphology, v. 375.

Putri, L. M., and Wicaksono, P., 2021, Mapping of Land Use Changes in the Core Zone of Parangtritis Sand Dunes Using Obia Method 2015-2020: Jurnal Geografi, v. 13, no. 1.

Qinghua, G., Li, W., Yu, H., and Alvarez, O., 2010, Effects of topographic variability and lidar sampling density on several DEM interpolation methods: Photogrammetric Engineering & Remote Sensing, v. 76, no. 6, p. 701-712.

Robb, C., Willis, I., Arnold, N., and Guðmundsson, S., 2015, A semi-automated method for mapping glacial geomorphology tested at Breiðamerkurjökull, Iceland: Remote Sensing of Environment, v. 163, p. 80-90.

Roering, J. J., Marshall, J., Booth, A. M., Mort, M., and Jin, Q., 2010, Evidence for biotic controls on topography and soil production: Earth and Planetary Science Letters, p. 183-190.

Rollerson, T.P., Thomson, B., and Millard, T. H., 1997, Identification of Coastal British Columbia Terrain Susceptible to Debris Flows: American Society of Civil Engineers, Debris-flow hazards mitigation: mechanics, prediction and assessment, San Francisisco, CA

Saha, K., and Van Landeghem, K. J. J., 2021, Evaluating an Automated Object-Oriented Method to Delineate Drumlins from Both Terrestrial and Submarine Digital Elevation Models: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, v. V-3-2021, p. 29-35.

Schuster, J. E., 2005, Geologic Map of Washington State: Washington Department of Natural Resources.

Shavers, E., and Stanislawski, L., 2018, Streams Do Work: Measuring the Work of Low-Order Streams on the Landscape Using Point Clouds: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, v. XLII-4, p. 573-578.

Shavers, E., and Stanislawski, L. V., 2020, Channel cross-section analysis for automated stream head identification: Environmental Modelling & Software, v. 132.

Shaw, S. C., Justice, T. E., McCarthy, R., Hinkle, J. C., Turner, T. R., Fransen, B. R., Jones, J. A., Giovanini, J., and Thornton, J., 2017, Automated, object-based image analysis (GEOBIA) model for landform detection and mapping, with applications to Pacific Northwest USA

landslide assessments, 3rd North American Symposium on Landslides: Roanoke, Virginia, Association of Environmental & Engineering Geologists.

Sherba, J., Blesius, L., and Davis, J., 2014, Object-Based Classification of Abandoned Logging Roads under Heavy Canopy Using LiDAR: Remote Sensing, v. 6, no. 5, p. 4043-4060.

Shi, X., Zhu, A.-X., Burt, J., Choi, W., Wang, R., Pei, T., Li, B., and Qin, C., 2007, An experiment using a circular neighborhood to calculate slope gradient from a DEM: Photogrammetric Engineering & Remote Sensing, v. 73, no. 2, p. 143-154.

Sofia, G., Tarolli, P., Cazorzi, F., and Fontana, G. D., 2011, An objective approach for feature extractions: distribution analysis and statistical descriptors for scale choice and channel network identification: Hydrol. Earth Syst. Sci., v. 15, p. 1387-1402.

Stewart, G., Dieu, J., Phillips, J., O'Connor, M., and Velduisen, C., 2013, The Mass Wasting Effectiveness Monitoring Project: An examination of the lanslide response to the December 2007 storm in Southwestern Washington: Cooperative Monitoring, Evaluation and Research committee of the Washington State Forest Practices Board.

Stock, J. D., and Dietrich, W. E., 2006, Erosion of steepland valleys by debris flows: Geological Society of America Bulletin, v. 118, p. 1125-1148.

Su, J., and Bork, E., 2006, Influence of vegetation, slope, and lidar sampling angle on DEM accuracy: Photogrammetric Engineering & Remote Sensing, v. 72, no. 11, p. 1265-1274.

Tabor, R. W.; Cady, W. M., 1978, Geologic map of the Olympic Peninsula, Washington: U.S. Geological Survey Miscellaneous Investigations Series Map I-994, 2 sheets, scale 1:125,000.

Tarboton, D. G., 1997, A new method for the determination of flow directions and upslope areas in grid digital elevation models: Water Resources Research, v. 33, no. 2, p. 309-319.

Tharwat, A., 2020, Classification assessment methods: Applied Computing and Informatics, v. 17, no. 1, p. 168-192.

U. S. EPA, 2012, Level III Ecoregions of Washington: U.S. EPA, National Health and Environmental Effects Research Laboratory.

UPSAG, 2006, Landslide Hazard Zonation Project Protocol: Olympia, Washington Department of Natural Resources.

Vamshi, G. T., Martha, T. R., and Vinod Kumar, K., 2016, An object-based classification method for automatic detection of lunar impact craters from topographic data: Advances in Space Research, v. 57, no. 9, p. 1978-1988.

Van Den Eeckhaut, M., Kerle, N., Poesen, J., and Hervás, J., 2012, Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data: Geomorphology, v. 173-174, p. 30-42.

USC TWIG, 2017, Unstable Slope Criteria – Research Alternatives, Unstable Slope Criteria Technical Working Group, Washinton State Department of Natural Resources, 47pp.

USFWS, 1999, Forests and Fish Report, United States Fish and Wildlife Service, 182pp.

Walsh, T. J., Koresec, M. A., Phillips, W. M., Logan, R. L., and Schasse, H. W., 1987, Geologic Map of Washington - Southwest Quadrant: Washington Department of Natural Resources.

Wang, S., Wu, Q., and Ward, D., 2017, Automated delineation and characterization of drumlins using a localized contour tree approach: International Journal of Applied Earth Observation and Geoinformation, v. 62, p. 144-156.

WADNR, 2006, Forest Practices Habitat Conservation Plan, Washington State Department of Natural Resources, 1,133 pp.

Washington Forest Practices Board, 2011, Board Manual: Standard Methodology for Conducting Watershed Analysis.

Wilson, J. P., Aggett, G., Yongxin, D., and Lam, C. S., 2008, Water in the landscape: a review of contemporary flow routing algorithms, *in* Zhou, Q., Lees, B., and Tang, G.-a., eds., Advances in digital terrain analysis: Berlin, Springer, p. 213-236.

Wohl, E., 2013, Migration of channel heads following wildfire in the Colorado Front Range, USA: Earth Surface Processes and Landforms, v. 38, p. 1049-1053.

Wohl, E., 2018, The upstream extent of a river network: A review of scientific knowledge of channel heads: U.S. Army Engineer Research and Development Center.

Wood, J., 2009, Geomorphometry in Landserf: Developments in Soil Science, v. 33, p. 333-349.

Wu, J., Liu, H., Wang, Z., Ye, L., Li, M., Peng, Y., Zhang, C., and Zhou, H., 2021, Channel head extraction based on fuzzy unsupervised machine learning method: Geomorphology, v. 391.

Yeomans, C. M., Middleton, M., Shail, R. K., Grebby, S., and Lusty, P. A. J., 2019, Integrated Object-Based Image Analysis for semi-automated geological lineament detection in southwest England: Computers & Geosciences, v. 123, p. 137-148.

Zhang, H., Loáiciga, H. A., Feng, L., He, J., and Du, Q., 2021, Setting the Flow Accumulation Threshold Based on Environmental and Morphologic Features to Extract River Networks from Digital Elevation Models: ISPRS International Journal of Geo-Information, v. 10, no. 3.

Zhou, H., Zhang, C., Xin, Z., Chang, Q., Wu, C., Ye, L., and Wu, J., 2019, Spatial variation of channel head curvature in small mountainous watersheds: Hydrology Research, v. 50, no. 5, p. 1251-1266.

Zylshal, P. D., and Haryono, E., 2013, An object based image analysis approach to semi-automated karst morphology extraction: Asian Conference on Remote Sensing, v. 34.

APPENDIX A. OBJECT-BASED IMAGE ANALYSIS (OBIA)~~BIA VIA IMAGE SEGMENTATION METHODS~~

With OBIA, we use specific terrain features to delineate different landforms. These terrain features and the associated landforms are listed in Table A1.

Table A1. ~~Lidar derivatives used to characterize landform morphometry~~OBIA model landforms and terrain features~~.~~

Landform	Morphometric Characteristics, Terrain Features
Terrace Valley Bottom Bench	-flat (<40%) -large area -defined by break-in-slope at boundary
Stream	-flat, or significantly less steep than neighboring valley walls -low density, i.e., shaped like a filament instead of a box (area/object radius) -generally contained within tightly convergent areas (maximum curvature)
Ridge	-divergent areas
Road	-low gradient -relatively long length of mainline -not divergent, i.e., ridge -main direction perpendicular to flow -flat or significantly less steep than neighboring areas
Inner Gorge	-significantly more steep than neighboring pixels -steep (>70%) -downslope of measurable break-in-slope -adjacent to stream
Bedrock Hollow	-convergent in planform direction -steep (>70%) -near top of basin -not parallel to stream -compact features

Commented [PJ257]: It would be helpful if these two methods were split into different appendices.

Commented [PJ258]: It would be helpful to start the appendix with introductory text to describe what is in the appendix before the table.

Commented [PJ259]: Is this table just for OBIA?

The primary data source for the OBIA routine is a lidar DTM. ~~All other derivatives used in the feature extraction process are created from this DTM and are listed in the far-right column of Table A1.~~ Lidar provides the opportunity to create high-precision, ground-surface digital terrain models at small grid spacings. At very high resolutions (e.g., less than 1m), the finite differences of first- and second- order derivatives of elevation (i.e., gradient and curvature) exhibit considerable noise (Lashermes et al., 2007) and lidar DTM accuracy can be diminished further by data interpolations algorithms and source data density characteristics (e.g., Qinghua et al., 2010). In forested landscapes, some topographic variability is associated with pits caused by the upheaval or decay of tree roots (Roering et al., 2010) and with dense vegetation or woody debris that is misclassified as bare earth (Lashermes et al., 2007). Shaw et al. (2017) recommend using a 2-m lidar DTM cell size for Pacific Northwest terrain because higher resolutions appear to capture pits and other micro-topographic features (e.g., fern clumps) on steep, densely vegetated slopes that influence DTM accuracy.

Derivatives used in the feature extraction process are created from this DTM. Lidar derivatives were created using published algorithms within ESRI ArcGIS Pro 3.0 when possible. This included the flow accumulation rasters and the gradient raster created using a standard 3X3 moving window. We developed executables in C++ to calculate curvature data based on definitions in Florinsky (2016) for: (a) tangential curvature (also known as horizontal curvature), the curvature tangential to a contour line and also a measure of flow convergence and divergence; and, (b) maximum curvature, the highest value of Gaussian curvature of normal slope sections at a point (i.e., ridges). Both curvature datasets were created using an 11X11 moving window. Curvature data were scaled from 0 to 255 floating point values, using published algorithms within ArcGIS Pro, in order to more closely represent 8-bit images that are more readily processed by eCognition. Edge detection data was created within eCognition using the Canny Algorithm using Gaussian Convolution factors of 2.5 and 6. Calculation of the mean difference to neighbor ~~(DTN)~~ pixels or objects was completed within the eCognition software using 5X5 and 15X15 moving windows.

Feature extraction in image analysis first begins by segmenting an image, i.e., lidar derivatives, into objects and then classifying those objects using statistical data values, geometry, or relationship to other objects. The ruleset, as designed, contains over 150 segmentation and classification algorithms to allow for analysis at multiple scales. For the purpose of this study only inner gorge and bedrock hollow extraction process will be covered in detail. However, the morphometric characteristics of the features extracted prior to the inner gorges and bedrock hollows are listed in Table B1.

We extract inner gorges and bedrock hollows following the outline in the tables below.

Commented [PJ260]: It would be helpful to start the appendix with introductory text to describe what is in the appendix before the table.

Commented [PJ261]: This paragraph needs to be divided into shorter paragraphs. It looks like the topics include:
-Derivatives
-How the derivatives were created
-Sources of noise

Table A2. Image analysis routine to classify inner gorges.

Commented [MM262]: It would be easier to understand if this was more plain text, similar to the Virtual Watershed table.

Landform	Image Analysis Process	Threshold Values	Features Could Interfere with Extraction
Inner Gorge	Threshold Segmentation	DTN(5)>10	-Bedrock outcrops perpendicular to stream, including bedrock cascades -Oversteepened sidecast -Steep, planar slopes not being actively incised
	Grow into neighboring pixels	DTN(15)>15 & Grad(3) >70	
	Grow into neighboring objects	Higher DEM value & below break in slope	
	Find enclosed by class	Absorb any object smaller than 50 pixels into inner gorge class	
	Remove objects	> Relative border to ridge or road –or- Distance to stream >30 pixel –or- Mean gradient <70% -or- <10 pixels	

Inner Gorges: OBIA delineation of inner gorges ~~The routine~~ begins with a multi-threshold segmentation algorithm, which segments and classifies an image into objects based on input data values - in this case the distance-to-neighbor (DTN) image derived from the DEM. Bright pixels in the DTN image correspond to pixels that are significantly steeper than the surrounding pixels within a 5X5 window. The threshold value for the bright pixels, 10, was determined manually to represent the length scale at which hillslopes could reasonably be expected to be influenced by

Commented [PJ263]: Is this just for image classification or does this section also apply to OBIA?

Commented [PJ264]: Please separate into multiple paragraphs for readability

channel incision given the potential energy of the streams in the test basins. It is also possible to select a segmentation scale to prevent individual pixel values from being classified. We chose a scale of 5 based on visual inspection of resulting classes.

The initial classified objects are then grown into neighboring pixels that are both steep (>70%) and significantly steeper than neighboring pixels using a moving window of 15, again based on visual inspection. Next, we extract slope breaks into an edge class using the same multi-threshold segmentation algorithm, this time using bright pixels in an image produced by running the Canny edge detection algorithm on a gradient raster. Bright pixels correspond to any abrupt changes in slope. The Gaussian convolution filter parameter in the edge detection algorithm determines the sensitivity of the algorithm to detect a break-in-slope. For inner gorge detection we used a value of 2.5. Again, this step is based on visual inspection of slope breaks noted adjacent to the channels in the test basins.

The initial inner gorge class and the edge class are segmented into 2-meter pixels using a chessboard segmentation algorithm, enabling a pixel-based object resizing algorithm to grow the ~~inner-gorge~~inner gorge seed class upslope into the edge class. Criteria for pixel expansion is defined in a customized algorithm that specifies the candidate object must have a larger DTM value. In effect, this grows the initial inner gorge class up to the first break-in-slope. The resulting inner gorge object is then merged and any holes in the classified object less than 50 pixels are absorbed into the inner gorge class.

The final step is to remove misclassified objects. First, any objects classified as inner gorges that are greater than 30 pixels from any stream classification are removed. This value was chosen after visual inspection of channel characteristics, i.e., wide valley bottoms that were not classified as streams. Second, any objects with mean gradients of less than 70% are removed. Third, inner gorge features with a relative border to a ridge or road that is greater than the relative border to a stream are removed. And finally, we chose a minimum mapping unit of 10 pixels for this class.

Table A3. Image analysis routine to classify bedrock hollows.

Landform	Image Analysis Process	Threshold Values	Features Could Interfere with Extraction
Bedrock Hollow	Threshold Segmentation	Convergent Tangential Curvature (<0)	-Bedrock outcrops -Steep, broadly convergent swales that do not converge at channel head -Steep cutslopes
	Find enclosed by class	Absorb any object smaller than 1000 pixels into inner gorge class	
	Multiresolution Segmentation Algorithm (MRS)	MRS: 40	
	Assign Class	Mean gradient >70% & Mean FAC <50 & Relative border to stream <0.3	
	Remove objects	Relative border to road >0.1 & Compactness >2.5 –or- Mean gradient <70% -or- <25 pixels	

Bedrock Hollows: The first step in the routine is a multi-threshold segmentation using the tangential curvature raster. The image is segmented and classified, at a scale of 5, resulting in a class with objects that are defined as convergent at a relatively coarse scale. The classified objects are reshaped using pixel-based object resizing algorithms to smooth the edges and all holes are

absorbed. A minimum mapping unit is set at 100 pixels for this class. The resulting class includes all topographic areas that are convergent at a coarse scale; this includes not only zero-order basins but also convergent channels further down in the basin. To extract only zero-order basins, the class is segmented using the multi-resolution segmentation algorithm at a scale of 40. Bedrock hollow classification is designated at the convergent objects that are greater than 70%, having low flow accumulation (<50), and a relatively low border to a stream object. This process pushes the steep, convergent class further up the basin. The final steps are to remove misclassified objects based on shape and proximity to the road, i.e., narrow, steep objects that are thought to be related to cutslopes. Small objects, <25 pixels, created as a biproduct of the MRS segmentation and classification are then removed.

APPENDIX B. VIRTUAL WATERSHED METHODS

Delineation of bedrock hollows and inner gorges was done using the following steps. Parameter values used are provided in Table B1.

1) Trace the channel network. A channel network is derived using flow directions inferred from the DTM. The methods used are described in Miller et al., 2015. For this study, ~~W~~we wanted the traced channels to extend into zones interpreted as bedrock hollows, so we forced the upstream extent of traced flow paths upslope of likely channel heads and into zero-order basins (Dietrich et al., 1986). Traced flow paths are represented using a linked-node data structure (Figure B1).

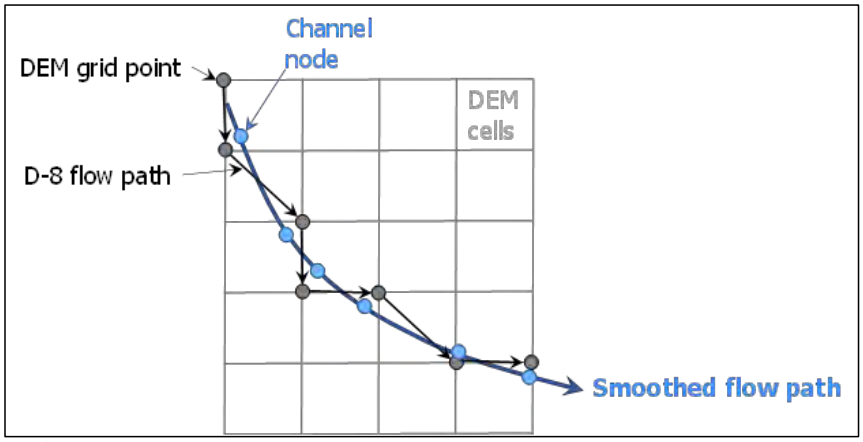


Figure B1. ***Channel centerlines are represented in a virtual watershed using a linked node data structure. Once the channel initiation criteria are met along any flow path, D8 flow directions are used to preclude dispersion and a node is created for each DTM grid point along the channelized flow path. The flow path is smoothed to provide improved estimates of channel length. Each node has an associated set of attributes, such as channel gradient, contributing area, channel geometry, and flood-plain width.***

2) Divide the traced flow-path network into channelized and upslope unchannelized (zero-order) portions. We sought~~eek~~ to initiate traced flow paths within the unchannelized zero-order basin and then identify the downstream channel head where the flow path likely transitions into a distinct channel. Bedrock hollows form within the zero-order basins, defined as the contributing area to the channel head, and inner gorges form adjacent to the channelized flow paths. Location of the channel head was based on combined thresholds in gradient, plan curvature, and contributing area times the gradient squared, discussed in more detail below. Contributing area is estimated from the DTM using the D-infinity flow-routing algorithm (Tarboton, 1997). Once the channel-initiation criteria are met, flow directions are based on a D-8 algorithm to preclude dispersion of channelized flow.
3) Delineate the flood plain or channel-adjacent zone for all channelized flow paths (downstream of the channel head). This is done by calculating the "height-above-channel"

Commented [PJ265]: This section is written partly as a directive for how other people should do this in the future, and also as a document for what was done for this project. It needs to be clear whether this appendix is supposed to be a step-by-step method for others in the future or if it is supposed to be what was done for this project. For example, were the "optional" steps used for all of the study areas, or just some? Which ones?
Try to get the tenses consistent. Currently a mix of present and past.

Commented [PJ266]: Some of this information belongs in the main text. Right now the first step sounds like the channels are traced by a human clicking along the channel network.

and flagging all DTM cells within five channel depths of the channel elevation. Channel depth is estimated using regional regressions to drainage area and mean annual precipitation (Kresch, 1998a; Kresch, 1998b; Magirl and Olsen, 2009).

Commented [PJ267]: Could you add a few examples/ranges of channel depths used along the regressions?

4) Delineate "drainage wings" to each flow-path segment. These are the adjacent hillslope areas that drain to a flow-path node (Figure B2). These are delineated using the D8-LTD flow direction algorithm (Orlandini and Moretti, 2009; Orlandini et al., 2003), which corrects for the bias introduced by the limited (eight) number of flow-direction options.

Figure B2. Drainage wings, showing channel-adjacent contributing area to each channel node. Each colored patch delineates the adjacent contributing area to a single channel node. Blue lines show traced channels; black lines show extension of the channel network into zero-order basins. Hollows are constrained to drainage wings within zero-order basins, those draining to the black lines. Single-direction flow paths with no downslope dispersion, using the D8-LTD algorithm that corrects for along-path bias, are used to delineate

contributing-area boundaries. Note the effect of flow diversion along the road near the bottom of the image.

5) Calculate ground-surface gradient at each DTM grid point. Gradient was~~is~~ calculated twice, looking upslope and downslope from each grid point, over a slope length of 15m. This is about the distance visible when measuring gradient in dense understory on the ground using a clinometer.
6) Delineate inner gorges.
   a. Identify the closest traced flow-path node, measured using horizontal Euclidean distance, for each DTM cell (Figure B3).

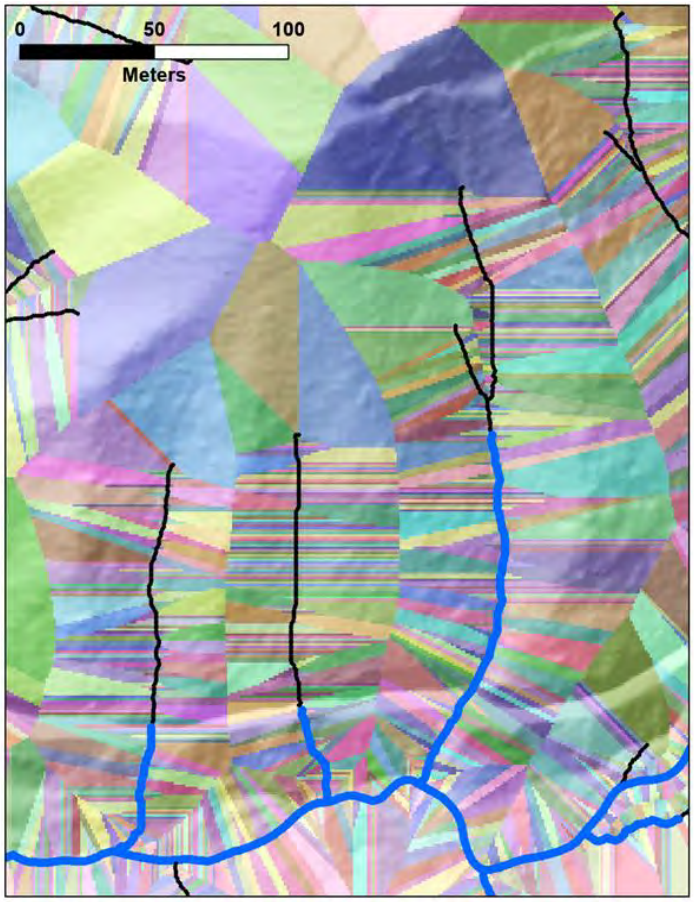


Figure B3. Closest node; each colored patch indicates the ID of the closest channel node, measured as horizontal Euclidean distance. Blue lines show traced channels; black lines show extension of the channel network

into zero-order basins. Inner gorges are constrained to closest-node patches associated with blue-line channels.

b. Examine ~~E~~each DTM cell ~~is examined~~. If it falls within a closest-node patch associated with a channelized flow path (blue lines in Figure B3) and is adjacent to a DTM cell mapped as a flood plain or channel-adjacent area in step two, then run a recursive flooding subroutine using that cell as the origin. The flooding routine examines all adjacent cells. Any cell that is within a closest-node patch associated with a channelized flow path that drains into the origin cell, that has an upslope-looking gradient greater than a specified threshold (e.g., 70%) and has an upslope-looking gradient greater than or equal to the downslope-looking gradient is flagged as a potential inner gorge. That cell then serves as the origin for the next iteration of the flooding routine. This will flag all contiguous valley-floor-adjacent areas that meet the inner -gorge criteria.

~~c.~~ c. The next three sub steps are for aesthetics~~; they~~ to make the delineated polygons look nicer.

i. Fill holes smaller than a specified minimum size within delineated inner gorges.
ii. Remove inner -gorge polygons smaller than a specified minimum size.
iii. Smooth inner -gorge polygon edges.

~~The following optional step was used to better match human-mapped inner gorges.~~

d. Allow inner -gorge polygons to extend upslope (along flow paths) into lower-gradient terrain beyond the slope break and relax the constraint that the upslope gradient must be greater than or equal to the downslope gradient. The final optional step was used to better match human-mapped inner gorges.

7) Delineate bedrock hollows.

~~d.~~a. Examine each ~~Each~~ DTM cell ~~is examined~~. If it meets the following two criteria, it is flagged as a hollow:

i. If it is within a zero-order basin, determined by being within a drainage wing that drains to a flow path node upstream of the channel head.
ii. If its upslope-looking gradient exceeds the specified hollow-gradient threshold (70%) and the upslope gradient is greater than or equal to the downslope gradient.

~~e.~~b. ~~Then the~~To improve ascetics:

i. Fill holes less than a specified size within the delineated polygons.
ii. Smooth polygon edges.
iii. Remove hollow polygons less than a specified size.

Commented [PJ268]: what size was used?

Commented [PJ269]: how?

Commented [PJ270]: what size?

~~f.~~c. The following ~~optional~~ steps were~~are~~ used to make polygons better match those drawn manually.

i. Increase the threshold gradient to match the maximum gradient found within mapped hollow polygons. This reduces the area and number of zones identified as potential hollows. Thresholds on curvature (tangential and normal) are also specified for those steepest-gradient points to be within the range of values found in the mapped polygons.

ii. Allow the polygon boundary to expand outward into lower-gradient and curvature zones, within the range of values included within the mapped polygons. The expanding polygon edges are precluded from crossing a drainage divide into a different zero-order (or channelized flow path) basin.

iii. Allow expanding polygon edges to extend laterally outside of the zero-order-basin boundary. The human-mapped hollow polygons did not follow zero-order basin edges, so they tended to be wider than the computer-generated polygons. Allowing the computer-drawn polygons to extend laterally outside of the zero-order basin, but still within the contributing area of the downstream channel, produced wider computer polygons.

iv. Infill embayments in the polygons. Human mappers tend to keep polygon boundaries smooth and include edge zones that may not meet the gradient criteria. Embayments are filled by tracing downslope flow paths out of a polygon edge; if they intersect the same polygon downslope, the intervening DTM cells are added to the polygon.

v. Eliminate computer-drawn polygons that have a proportion of their area less than 70% gradient that is greater than a specified minimum, based on the proportions in human-mapped polygons.

Frequency distributions of values from the human-mapped polygons were used to guide the choice of thresholds for mapping inner gorges and hollows.

The up- and downslope gradients measured on the DTM are sensitive to the length scale over which they are measured. The choice of 15 meters was predicated on both the typical line-of-site distance visible with thick understory vegetation and on the minimum dimension of the hollow and inner -gorge landforms we seek to delineate. With a measurement length of 15 meters, tree-fall pits and steep channel banks and terrace edges with slope lengths less than this will not be resolved and not included in the delineated hollow and inner -gorge polygons. This length scale can be adjusted to better match the spatial scale of the landforms of interest, if needed.

Commented [PJ271]: Do optional steps mean that they were only sometimes used? Why is it optional for this project? How would you decide whether or not to do this when processing data across a large area?

B1. THRESHOLD VALUES

Table B1. Steps in generating ~~bedrock hollow~~bedrock hollow and inner -gorge RIL polygons using a Virtual Watershed.

Task	Description	Parameters
Trace channel network	Create a linked-node data structure that delineates stream channels based on flow paths traced on a DTM. Flow paths are based on D-infinity until the channel initiation threshold is met, downstream of which D-8 flow paths are used. Channelized D-8 flow paths are constrained by topographic indicators of a channel using tangential curvature and deviation from mean elevation. This network is extended upstream of the channel head into zero-order basins.	Specific contributing area times gradient squared (AS^2) ~ $30m^2$, variable by region. This threshold is estimated using a log-log plot of channel density versus threshold value. A concave-up inflection in the channel density range of 1-100km/km^2 indicates feathering of traced channels onto planar hillslopes; this inflection identifies the transition from divergent to convergent topography. This threshold is calibrated directly from the DTM. Tangential Curvature > 0 (> 0 indicates concave up). Gradient and curvature are measured over centered circular windows 30m in diameter.
Identify channel head	The traced network extends into generally unchannelized zero-order basins. The channel head lies somewhere downslope of the traced network head. The channel head determines the extent of the zero-order basin, which are defined by the contributing area to the channel head. RIL hollows are only allowed within zero order basins and inner gorges only	Same parameters as the channel-initiation threshold above, but with larger threshold values: AS^2 ~$1500m^2$, tangential curvature ~ 0.005; may be variable by region. We do not have a reliable means of identifying these thresholds directly from the DTM (unless we obtain very high-density ground returns) and so we estimated them using the manually mapped bedrock hollow RIL polygons assuming, as per Board Manual

Commented [PJ272]: If a table extends onto multiple pages, the heading needs to be on each page.

Task	Description	Parameters
	allowed downstream of the channel head.	~~Section 6~~, that these extend to the channel head.
Delineate valley floor	The floodplain or active channel zone to which an inner gorge is adjacent.	Delineated as the channel-adjacent zone within five channel-depths elevation of the channel. Channel depth is based on regional regressions to contributing area and mean annual precipitation.
Buffer valley floor	If a steep (e.g., >70%) valley-side slope is within a specified distance of the delineated valley floor, buffer the valley floor zone out to that steep slope.	A maximum buffer width of 2.5m was used based on manually mapped inner gorges.
Delineate drainage wings	A "drainage wing" delineates the contributing area to a single channel node. It is delineated using the D8-LTD flow-direction algorithm, which corrects flow-path bias arising from the limited (eight) number of flow directions. D8-LTD is used rather than D-infinity to preclude dispersion, which would create overlapping drainage-wing boundaries.	
Calculate up- and downslope-looking gradient at each DTM cell	This is calculated using a smooth polynomial surface fit to 9 points, one at the DTM grid point and 8 on the	Length scale over which gradient is measured; 15 m to match typical sight distance on a vegetated slope.

Commented [PJ272]: If a table extends onto multiple pages, the heading needs to be on each page.

Task	Description	Parameters
	circumference of a circle of specified radius around that point. For upslope-looking gradient, the four downslope points are a flipped image of the four upslope points; for the downslope-looking gradient, the four upslope points are a flipped image of the four downslope points.	
Create a "closest channel node" raster based on Euclidean distance	Each DTM grid point is assigned the straight-line horizontal distance to the nearest channel node. DTM grid points at channel nodes have a distance of zero.	
Identify inner gorge "seed" cells	These are DTM grid points adjacent to a delineated valley-floor zone that meet the gradient threshold.s	Upslope-looking gradient > 70% and upslope-looking gradient ≥ downslope-looking gradient
Grow the inner gorge zone outward from the seed cell	Zone grows outward by following flow paths into DTM cells that meet the gradient thresholds and do not extend into closest-channel cells of a zero-order channel node.	Upslope-looking gradient may be relaxed below 70% and the ratio of upslope to downslope gradient may be allowed a value less than one based on manually mapped inner gorge polygons.
Fill small holes within inner gorge zones; delete small inner gorge patches; smooth polygon boundaries		Minimum hole size = 1000m^2, minimum patch size = 50m^2

Commented [PJ272]: If a table extends onto multiple pages, the heading needs to be on each page.

Task	Description	Parameters
Delineate zero-order basins	These are the drainage wings that drain to channel nodes upstream of the channel head	
Identify hollow "seed" zones	These are DTM grid points within the zero-order basins that meet the gradient and tangential curvature thresholds	Upslope-looking gradient > 85%; upslope-looking gradient ≥ downslope-looking gradient; tangential curvature > 0.005 (must be convergent). Seed zones must exceed 20 m^2 in area.
Grow hollow zone outward from the seed zone	Optionally allow growth outside the contributing area to the channel head, but not into closest-node zones for nodes at or below the channel head or crossing drainage divides.	The upslope-gradient and tangential-curvature thresholds may be relaxed; the ratio of upslope to downslope gradient may be allowed a value less than one. These relaxed thresholds are based on values within the manually mapped hollow polygons. Current gradient threshold = 0.5, up-to-down slope ratio = 0.5, tangential curvature = 0.
Fill small holes within the hollow zones; delete small hollow patches; infill embayments; smooth edges	Embayments are identified as zones where a flow path exits a hollow zone and then re-enters the same hollow zone downslope. These flow-path segments are then included into the hollow zone.	Minimum hole size = 1000m^2, minimum patch size = 300m^2

Commented [PJ272]: If a table extends onto multiple pages, the heading needs to be on each page.

Task	Description	Parameters
Delete hollow polygons that are smaller than a minimum size	Decision is based on manually mapped hollow polygons.	300m^2
Delete hollow polygons that have a proportion of their area < 70% greater than a specified value	Decision is based on manually mapped hollow polygons.	Proportion > 70% must be greater than 40%

Commented [PJ272]: If a table extends onto multiple pages, the heading needs to be on each page.

APPENDIX C. PHASE 1 MANUAL MAP COMPARISONS

Commented [PJ273]: The figures in this appendix are too small - they lose their value when you can't see the details.

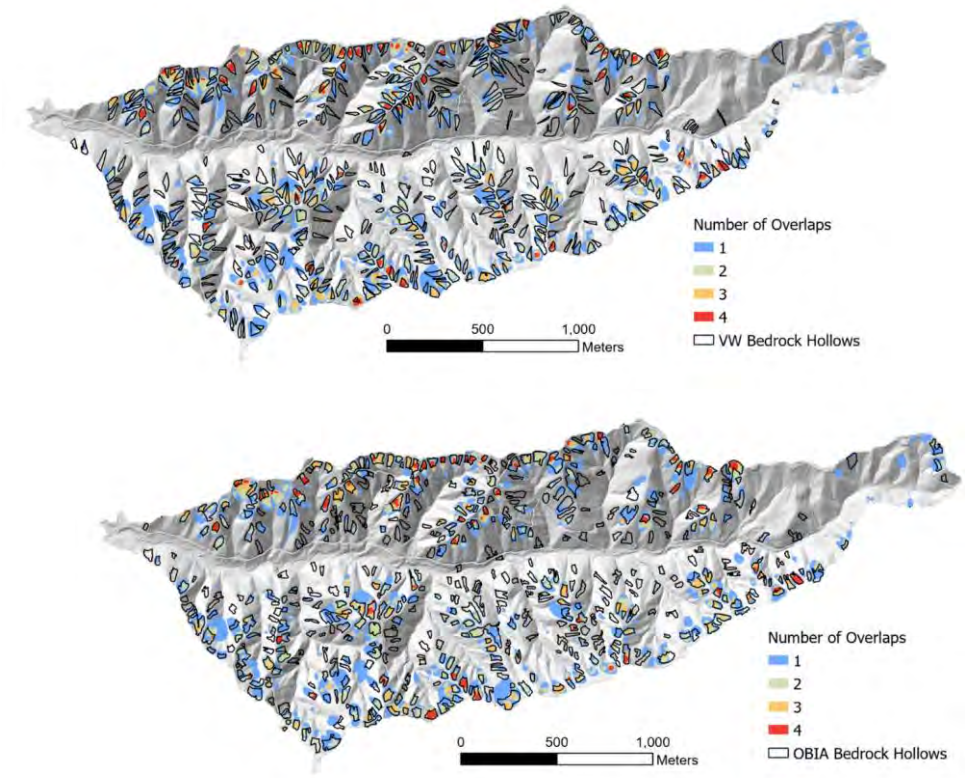


Figure C1: Overlap map of bedrock hollows in Site 1A. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

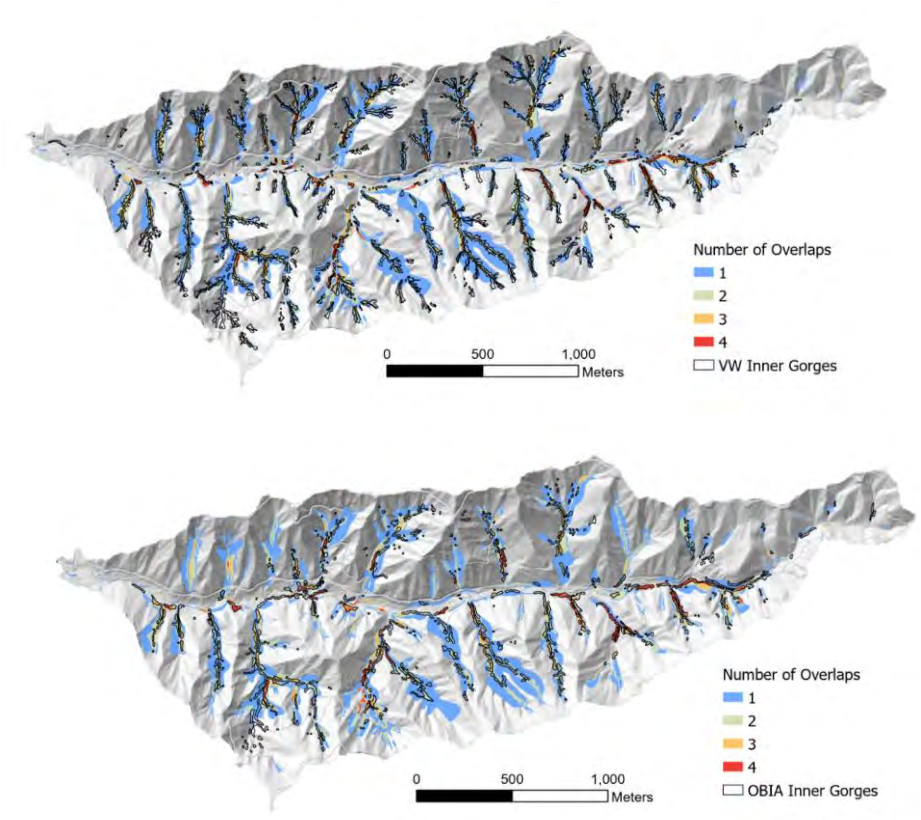


Figure C2: Overlap maps for inner gorges in Site 1A. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

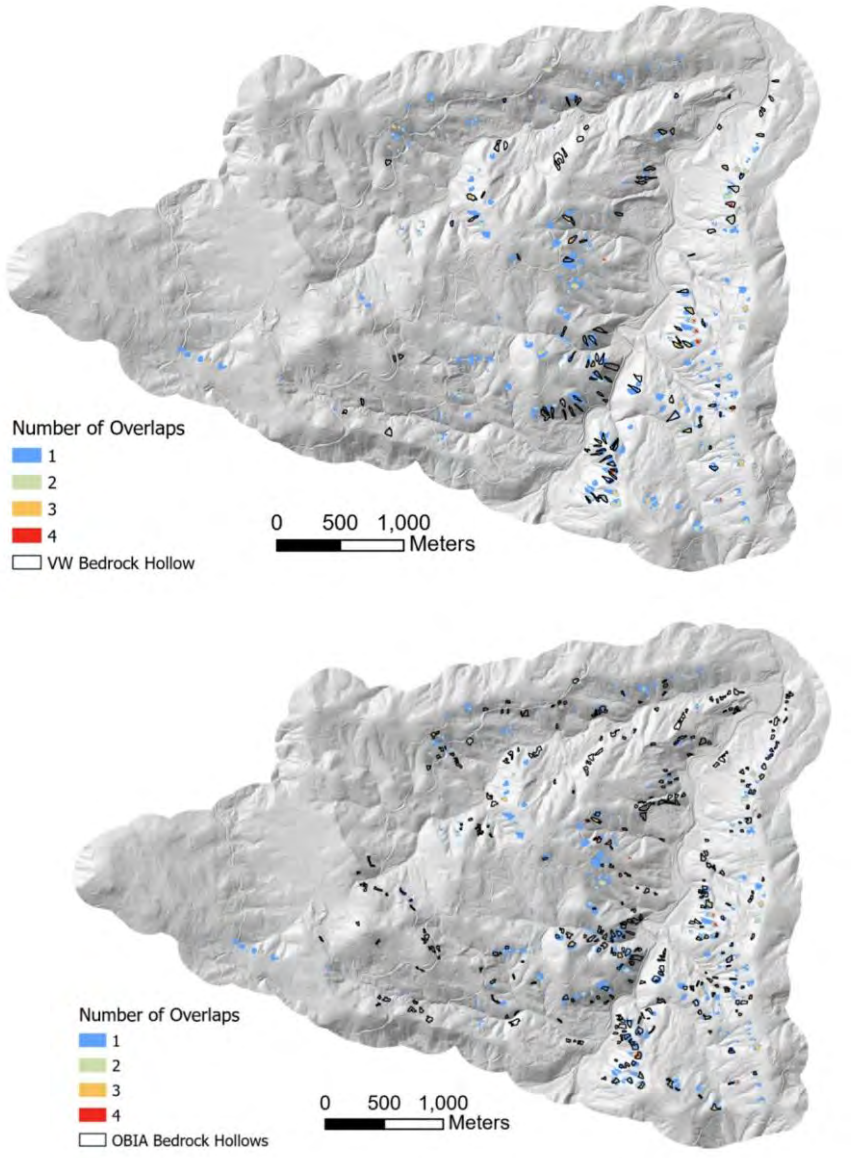


Figure C3: Overlap maps for bedrock hollows in Site 2. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

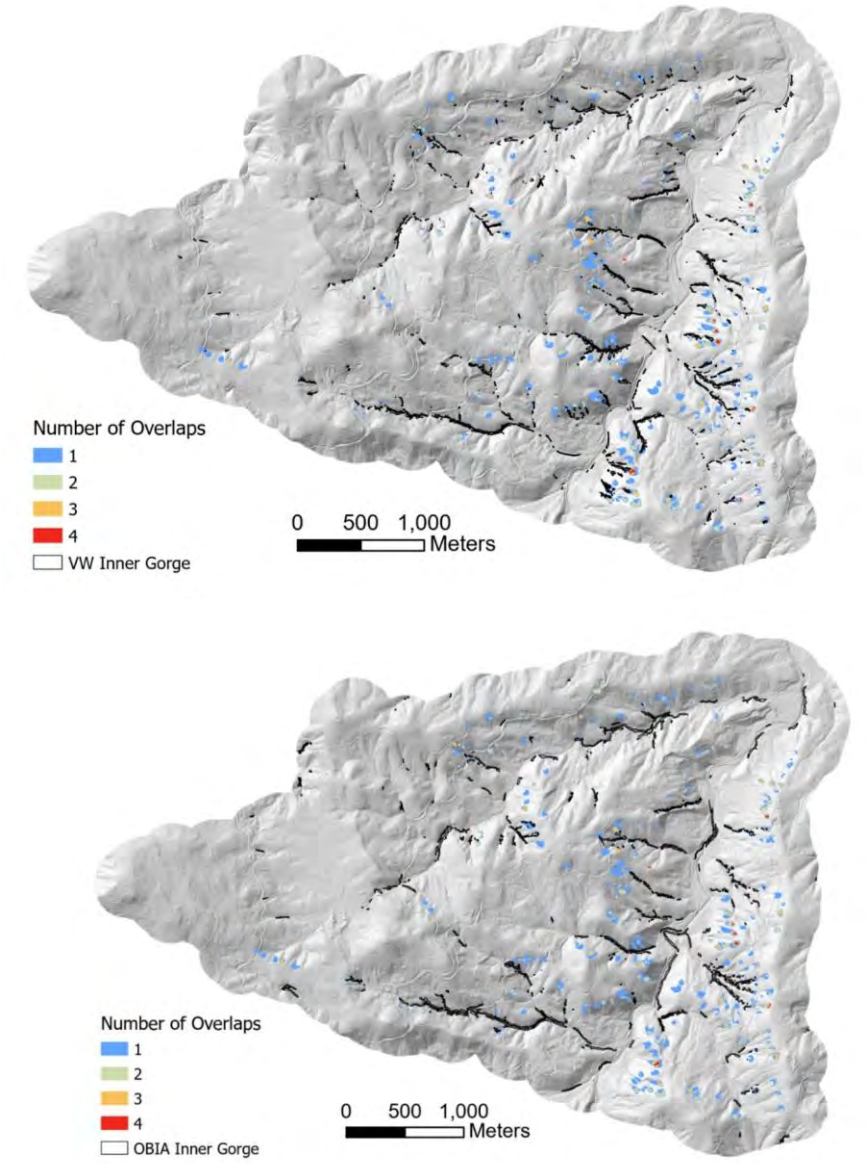


Figure C4: Overlap maps for inner gorges in Site 2. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

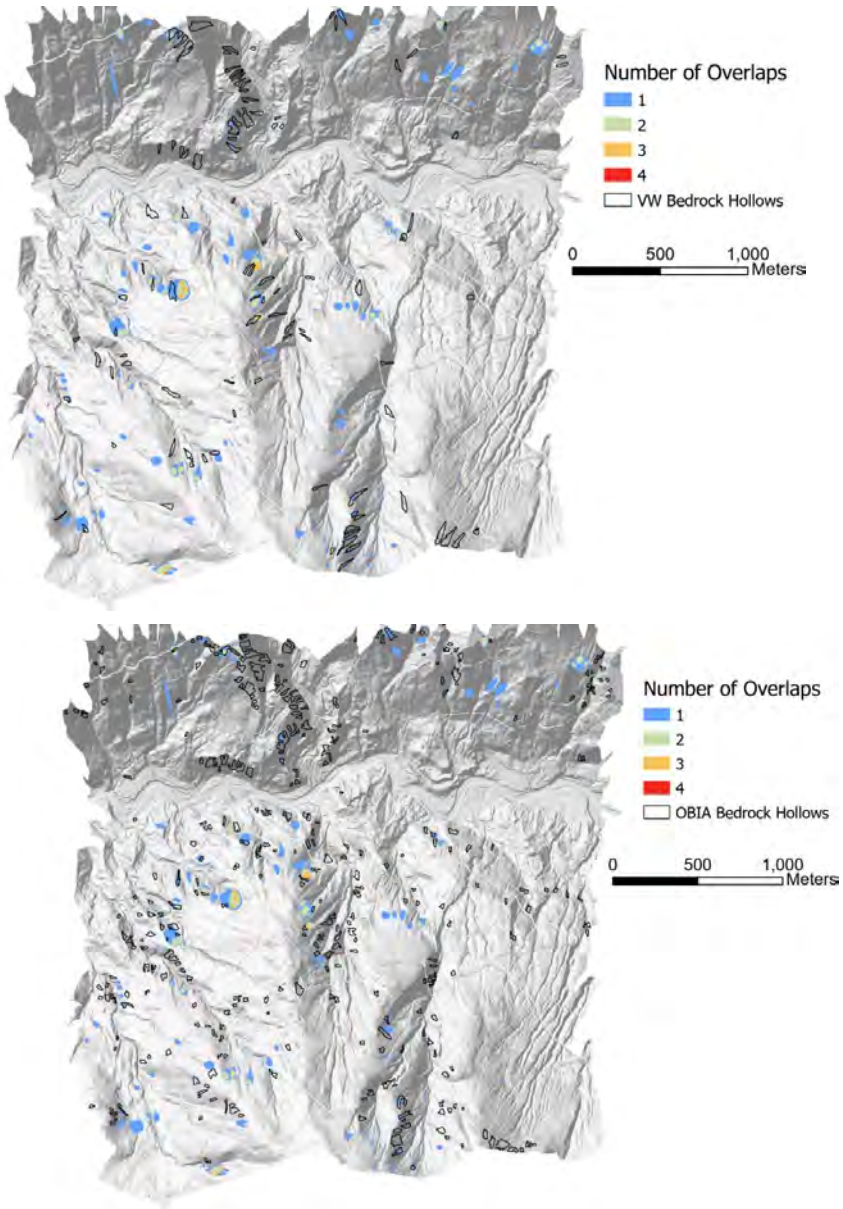


Figure C5: Overlap maps for bedrock hollows in Site 3. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

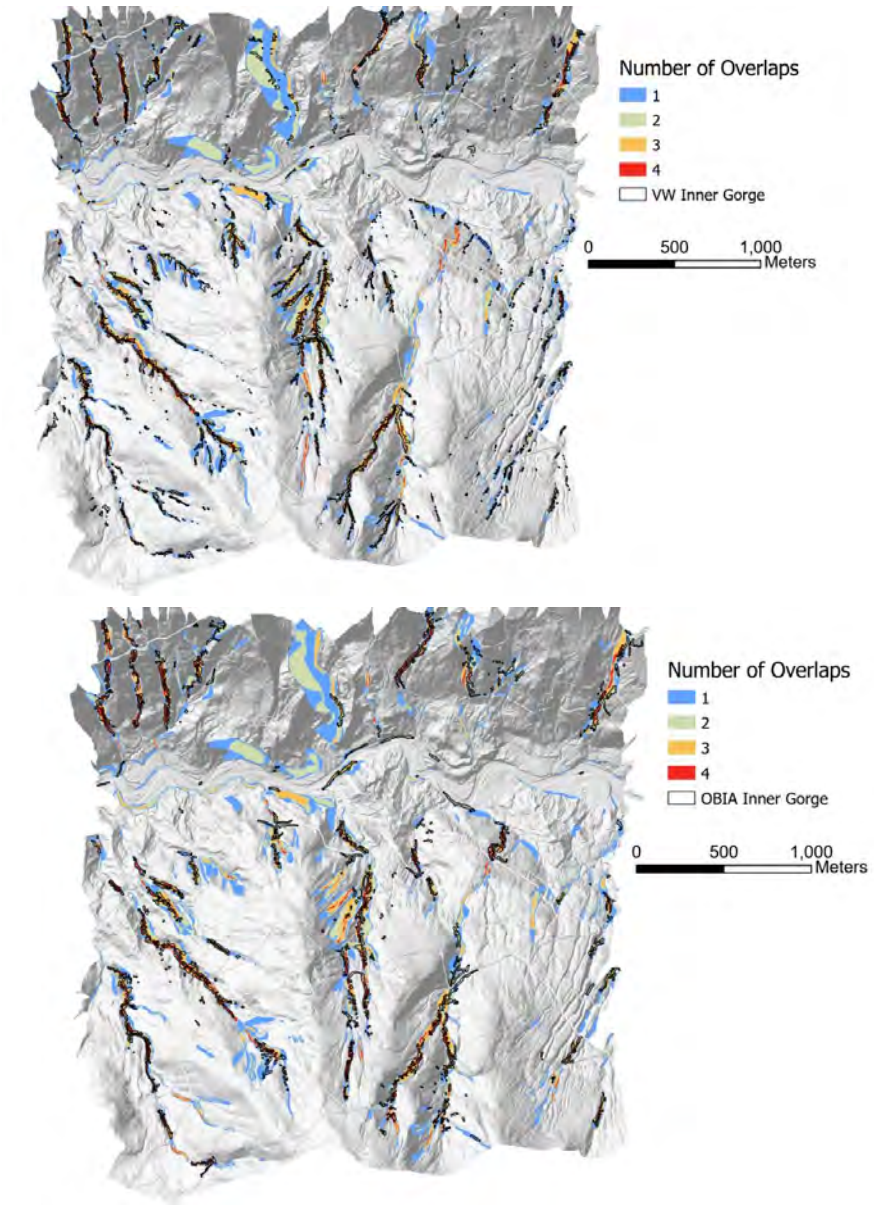


Figure C6: Overlap maps for inner gorges in Site 3. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

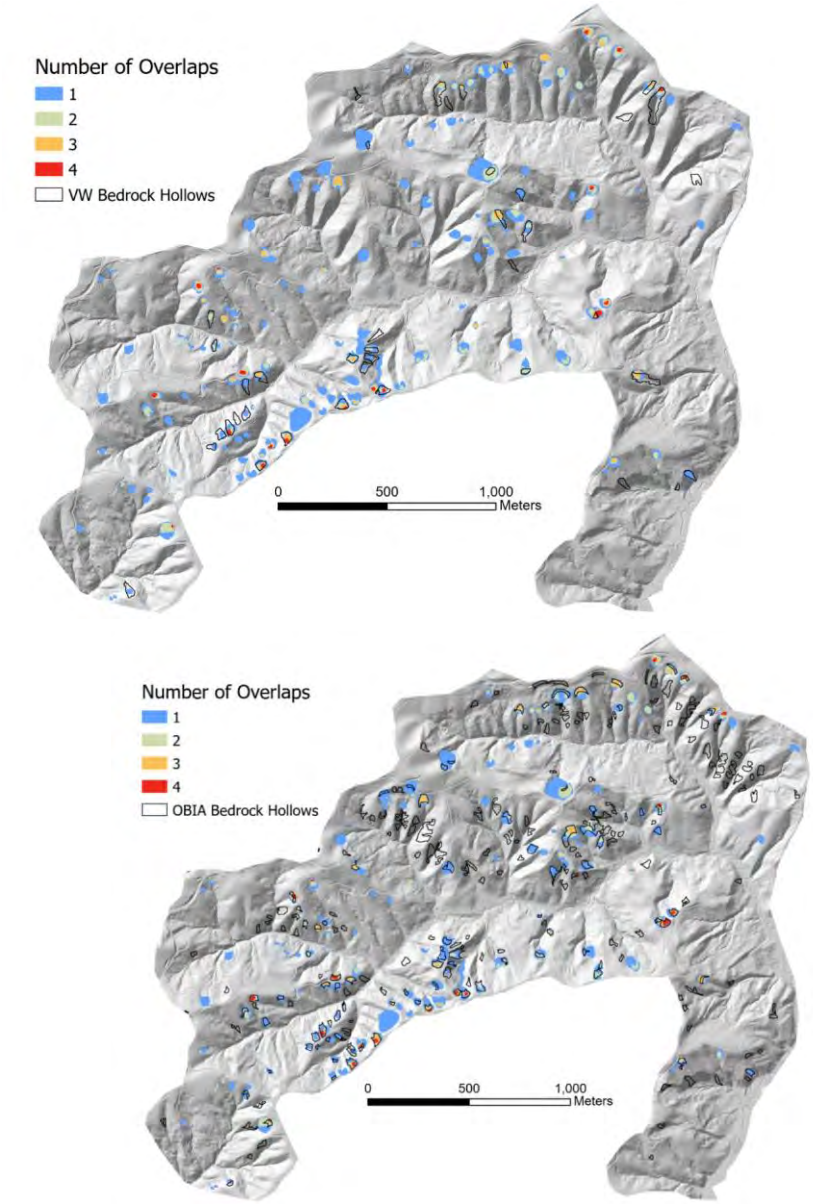


Figure C7: Overlap maps for bedrock hollows in Site 4. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

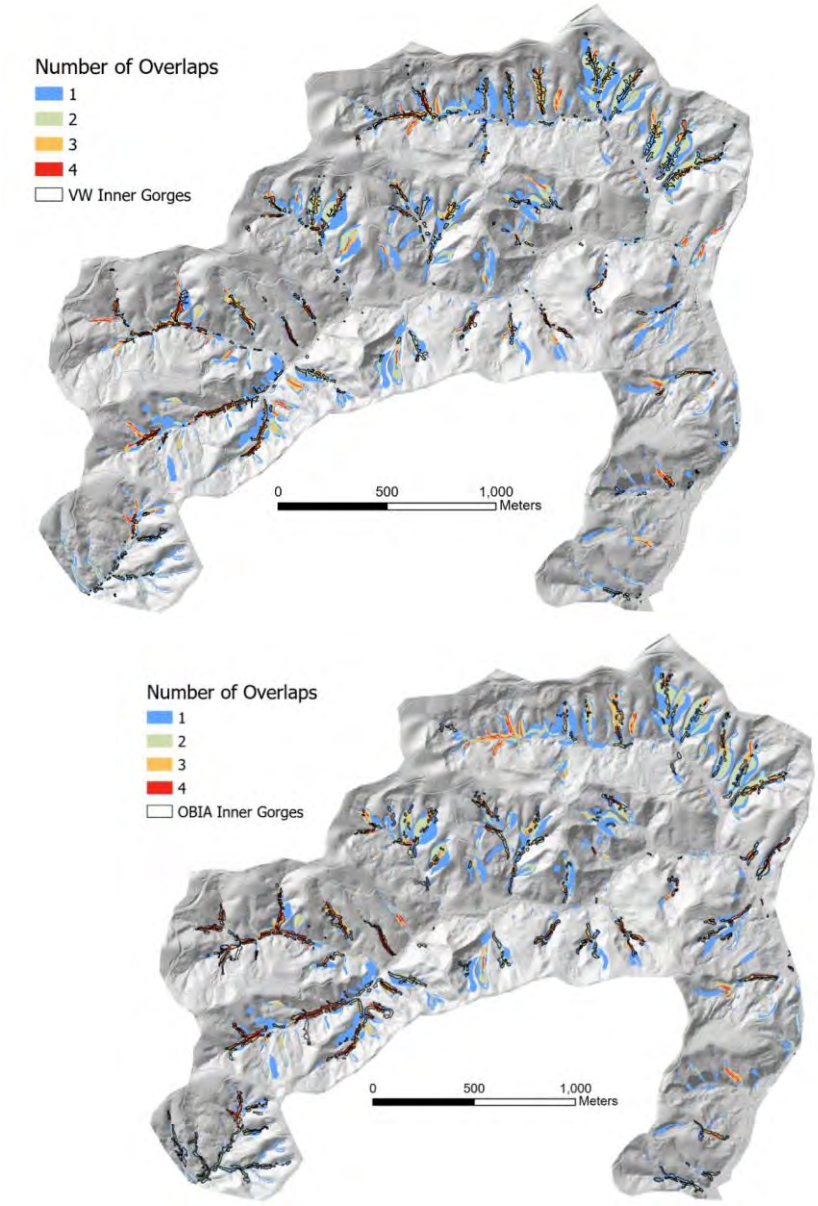


Figure C8: Overlap maps for inner gorges in Site 4. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

C1. Accuracy, Balanced Accuracy, Recall & Precision

The overlap maps were used to construct a confusion matrix for each of the 12 possible pairs of mappers, from which we calculated values for accuracy, balanced accuracy, recall, and precision for each pair of mappers.

To aid with interpretation of these results, Figure C.9 shows the distribution of mapped RIL and NonRIL area for each of the mappers at each of the study sites. This shows how dramatic the difference in area between the mapped RIL polygons and the remaining basin area – the NonRIL area – is for all mappers at all sites. The RIL polygons encompass a tiny fraction of the total study area at each site.

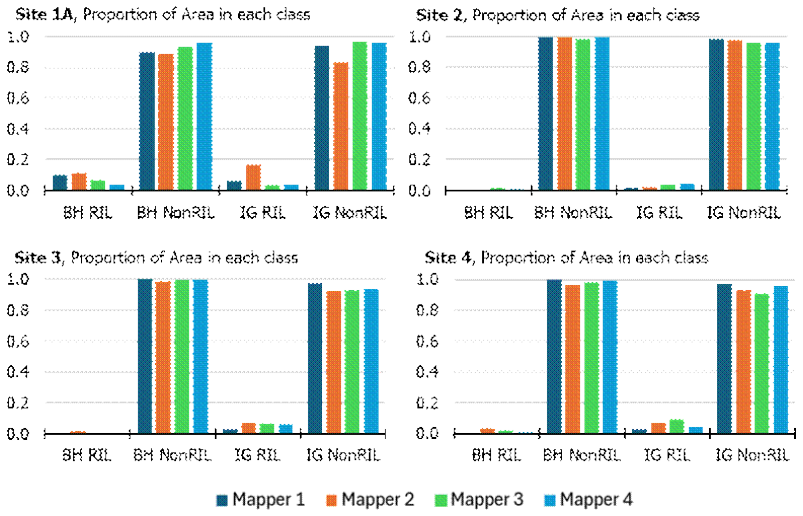


Figure C.9. The proportional area of mapped RIL polygons and NonRIL zones for Bedrock Hollows (BH) and Inner Gorges (IG) for each mapper at each study site. For both Bedrock Hollows and Inner Gorges, the proportion of area mapped as NonRIL is considerably higher than that mapped as RIL. This imbalance in area between the RIL and NonRIL classes complicates measures of accuracy.

Values of accuracy, balanced accuracy, recall, and precision calculated for each pair of mappers are shown in the bar charts in Figures C.10 through C.12. For any pair of mappers, the study area is divided into four zones (Table 2):

1. The area of overlap between the two sets of RIL polygons. This area is referred to as the True Positive (TP) area in the context of the confusion matrices.
2. The area where neither mapper drew a polygon. This is referred to as the True Negative (TN) area.

3. The area where the reference mapper had an RIL polygon and the comparison mapper did not. This area is referred to as the False Negative (FN) area – the comparison mapper missed it relative to the reference mapper.
4. The area where the reference mapper did not draw an RIL polygon but the comparison mapper did. This area is referred to as the False Positive (FP) area.

Note that when the reference and comparison mappers are switched, the TP and TN areas remain the same, but the FP and FN areas are switched. Looking at the definitions, the measure of accuracy will remain the same when the reference mapper and comparison mapper are switched ~~two mappers are switched between reference and comparison,~~ the previous measure of recall becomes the precision value, the precision value becomes the recall value, and the balanced accuracy will be slightly different.

Commented [PJ274]: Please revise this sentence to be clearer

Measures of accuracy (Figure C.10) are, as expected, very high, reflecting the predominance of NonRIL area at the study sites (Figure C.9). Measures of balanced accuracy (Figure C.11) tend to be lower and show variability among mappers. Note that about 50% of the calculated balanced accuracy still arises from the very large imbalance between RIL and NonRIL areas at the study sites. The NonRIL portion (TN/(TN+FP)) of balanced accuracy will be almost unity for every combination of mappers, because the values of TN are overwhelmingly larger than values of FP in all cases. For these reasons, measures of recall and precision, which do not rely on the value of TN, provide a more sensitive analysis of how well the mapped RIL polygons agree.

Measures of recall and precision are shown in Figures C.12 and C.13. Recall values indicate the proportion of the reference RIL area matched by overlap of the reference and comparison RIL polygons. Precision values indicate the proportion of the comparison RIL area matched by that overlap. If the reference-map RIL polygons are large compared to those of the comparison map, recall may be small and precision large, even if there is abundant overlap. Reference RIL polygons that are small compared to the comparison map would tend to produce larger values of recall and smaller values of precision. Cases where recall and precision vary inversely thus indicate a difference in the general sizes of the reference and comparison RILs. For example, at Site 1A, Mapper 2 tended to draw larger inner gorge RILs than the other mappers (Figure 12). Hence, measures of recall for inner gorges where Mapper 2's map served as the reference are small and measures of precision are large. Likewise, where Mapper 2's map served as the comparison, measures of recall are large and measures of precision are small. If both recall and precision are large, that indicates good spatial alignment of the RIL polygons, but we do not see any examples of that. If both recall and precision are small, that indicates poor alignment. This is particularly the case at Site 2.

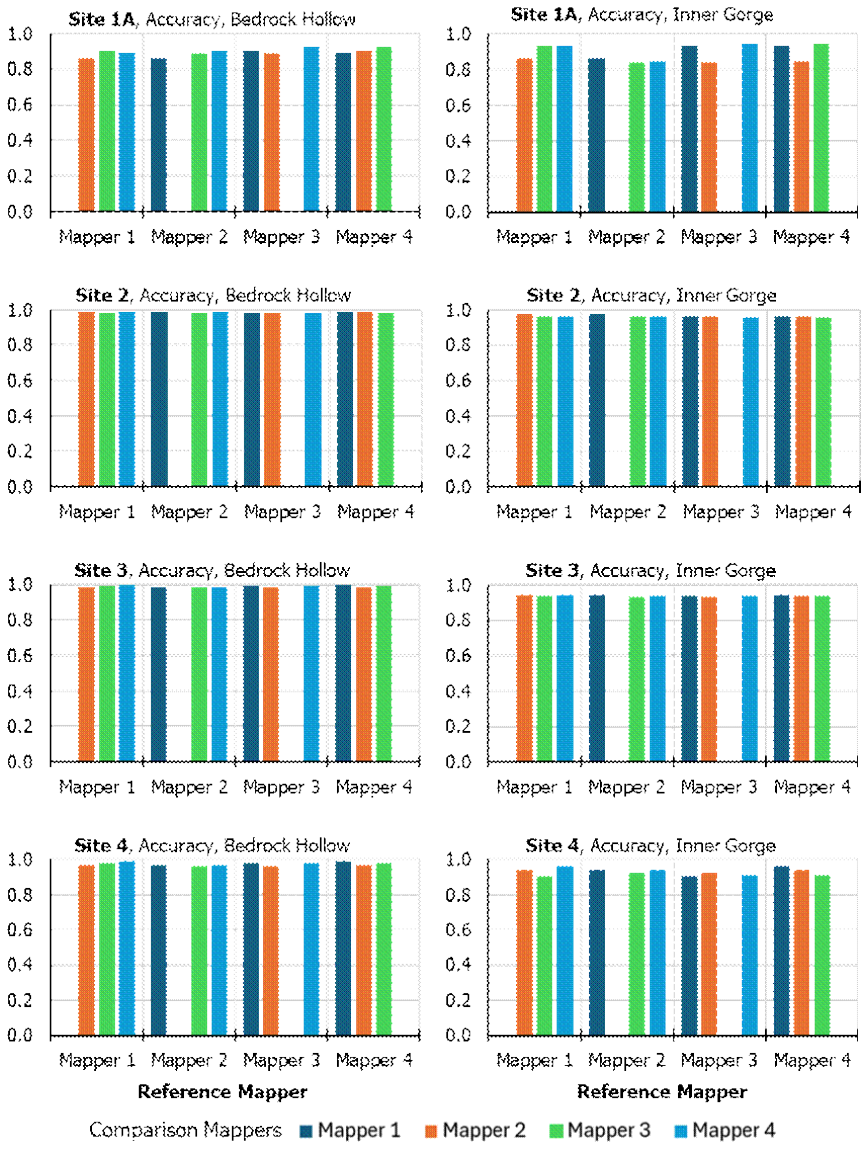


Figure C.10. Comparison across mappers using accuracy. The x-axis shows the reference mapper; the bars refer to the comparison mapper. Because the NonRIL area is much larger than the RIL area, this measure primarily reflects the agreement in NonRIL area between mappers.

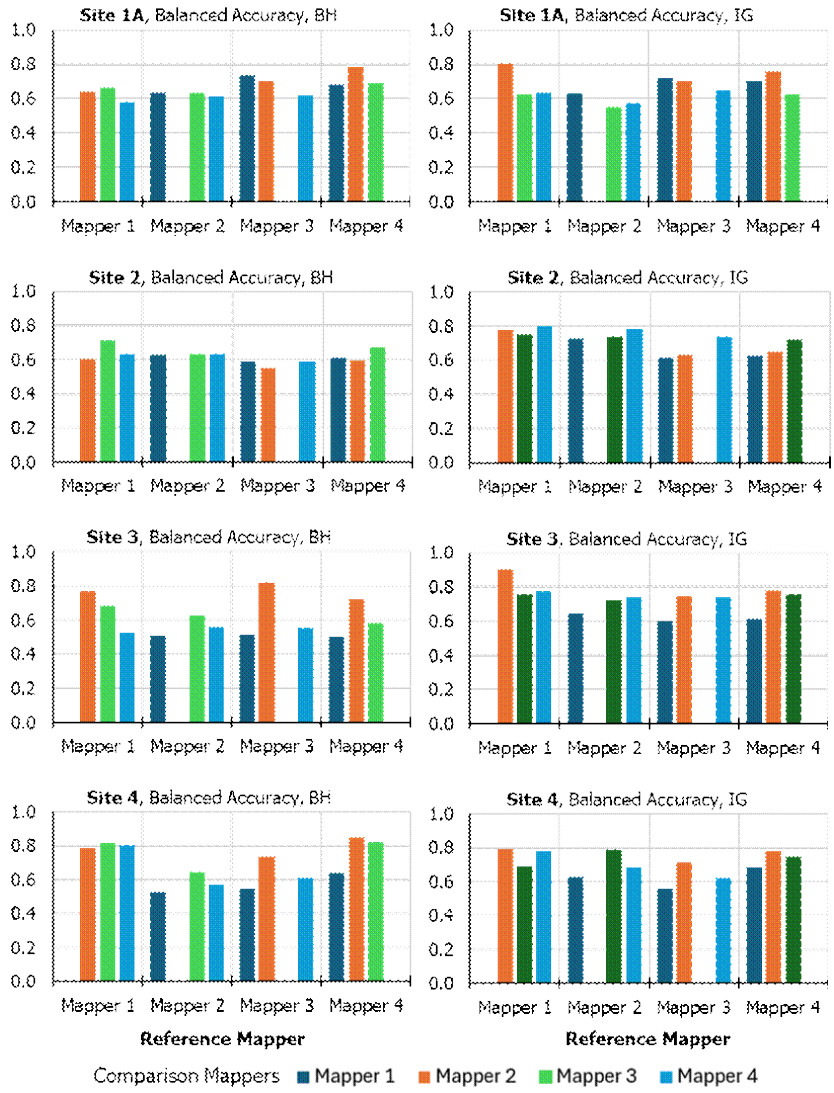


Figure C.11. Comparison across mappers using balanced accuracy. The x-axis shows the reference mapper; the bars refer to the comparison mapper. Balanced accuracy is similar to the accuracy shown in Figure C.9 for Site 1A, where the difference between RIL and NonRIL area is not so great. For Sites 2, 3, and 4, the balanced accuracy is much lower than the accuracy shown in Figure C.9, reflecting the large proportion of the study-site areas in NonRILs.

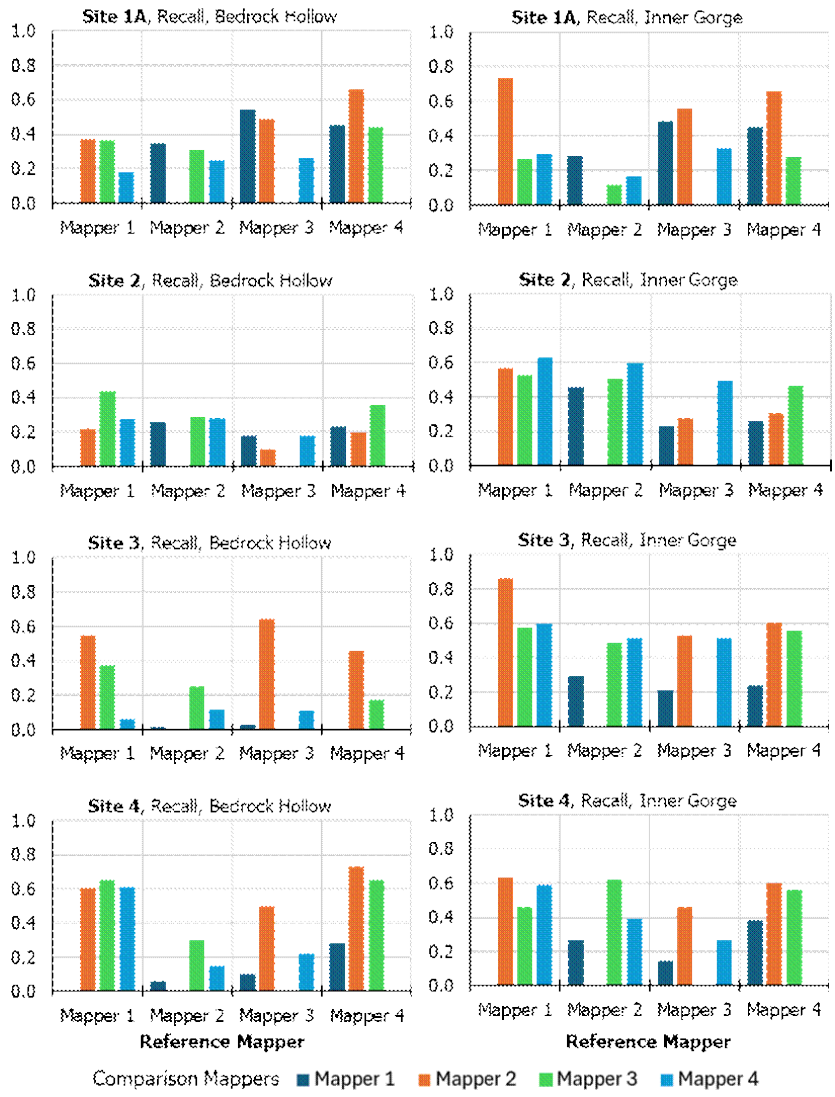


Figure C.12. Comparison across mappers using recall. The x-axis shows the reference mapper; the bars refer to the comparison mapper. Recall indicates the proportion of the reference RIL area matched by the comparison map. If the reference-map RIL polygons are large compared to those of the comparison map, recall may be small even if there is abundant overlap. If the reference polygons are small compared to the comparison map, recall will be small, even if there is abundant overlap.

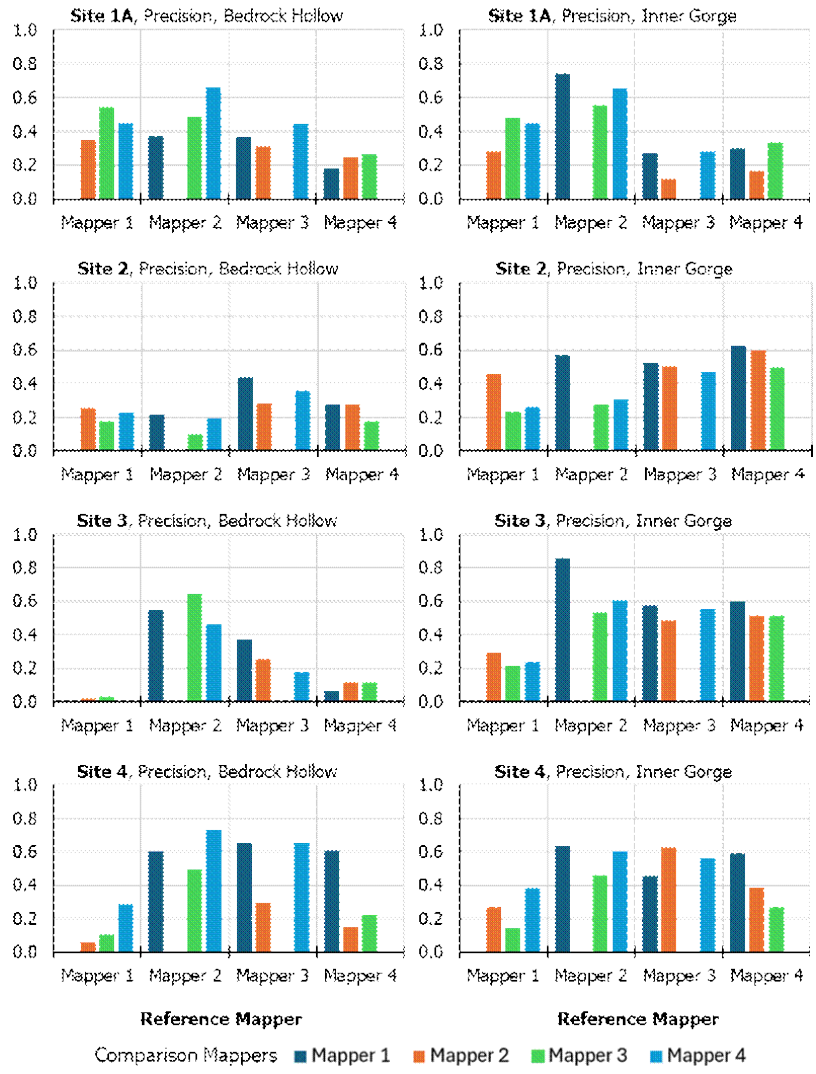


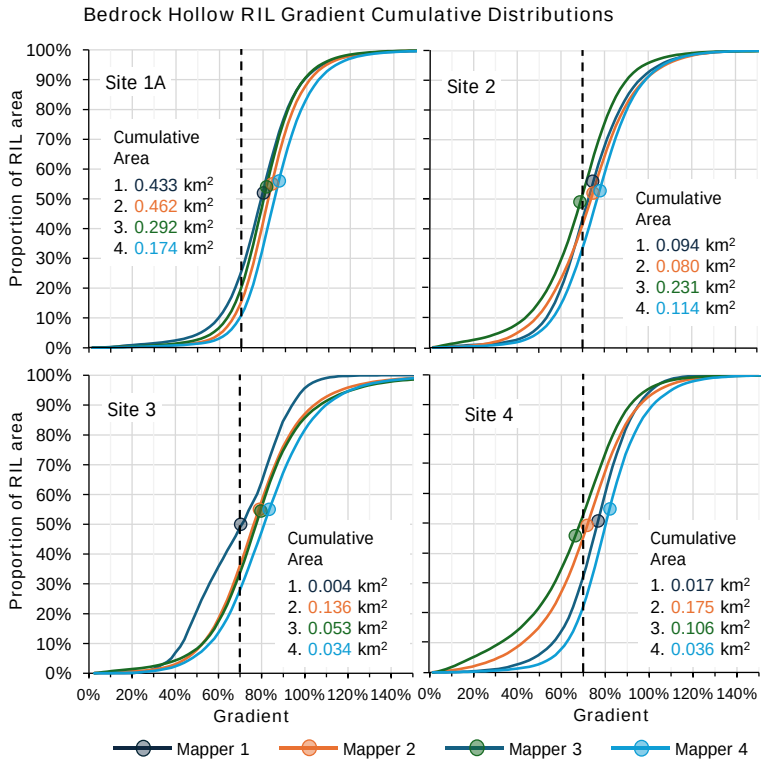
Figure C.13. Comparison across mappers using precision. The x-axis shows the reference mapper; the bars refer to the comparison mapper. Precision indicates the proportion of the comparison RIL area that matches the reference map. If the reference-map RIL polygons are large compared to those of the comparison map, recall may be small even if there is abundant overlap. If the reference polygons are small compared to the comparison map, recall will be small, even if there is abundant overlap.

Figures C.12 and C.13 show remarkable variability. The degree to which the maps agree varies both across mappers and across study sites. For some cases, there is almost no overlap of RIL polygons between two mapper's maps. At Site 3, Mapper 1 matched less than 3% of any other mapper's ~~bedrock hollow~~bedrock hollow RIL polygons (Figure C.13) and matched less than 1% of Mapper 4's.

C2. Cumulative Frequency Distributions

Here we look to see what differences exist in the frequency distributions of the relevant topographic attributes among the mapper's ~~bedrock hollow~~bedrock hollow RIL polygons across the four study sites. We focus on bedrock hollows because each hollow is represented by a distinct and separate polygon. Inner gorges are mapped as semi-continuous zones adjacent to channels and individual polygons are less well defined. Frequency distributions of topographic attributes for inner gorges are shown in Appendix F.

Figure C.14 shows cumulative distributions of gradient for the ~~bedrock hollow~~bedrock hollow RIL polygons for each of the mappers at each of the study sites. The cumulative distributions show ~~on the y-axis~~ the proportion of the total RIL area on the y-axis that falls below the value on the x-axis. The total area represented by each curve is also listed on the plots.



C.14. Cumulative distributions of ~~bedrock hollow~~bedrock hollow RIL gradient within each mapper's set of polygons for each study site. Circles indicate mean values. The total area within each mapper's polygons is listed for each site. Dashed lines indicate 70% gradient value along the x axis.

The Board Manual describes hollows as convergent (spoon-shaped) landforms with gradients of 70% or greater. The B~~b~~oard M~~m~~anual does not specify what proportion of the landform needs to meet or exceed 70%, nor does it give a qualifiable description of "convergent." From the cumulative distribution plots, we can see what proportion of the mapped hollow RIL area falls above and below 70%. Those proportions vary between mappers and between study sites. The mapped hollows tend to be steeper at Site 1A than at the other sites (as the site is overall steeper, this ~~is~~ reflects a real topographic difference); the proportion of the mapped hollows less than 70% falls between 10% (Mapper 4) to about 27% (Mapper 1), whereas at the other sites that proportion varies from 20% to over 50%. The circles on the cumulative-distribution lines indicate the mean value; the median is where the line crosses the 50% value on the y axis. Site 1A also has the highest mean gradient values, clustered between 80% to 90% over all the mappers (compare with Figure C.15). The shape of the curves and the range of values encompassed by the

distributions are similar, though Mapper 1 veers off a bit at Site 3. The mappers are consistent in their differences: Mapper 3 consistently includes lower gradients; mapper 4 consistently includes higher gradients. Site 4 shows the greatest variation among mappers, although the overall complexity of the Wishkah site probably contributes to this variance. These inconsistencies could be due to different interpretations of the rule and guidance – with each mapper considering a different proportion of side slope exceeding 70% necessary to fall under the rule criteria.

Figure C.15 shows box and whisker plots of the mean gradient within each polygon calculated for each set of mapper's polygons. The number of polygons represented by each plot is listed in parentheses in the legends. The steepness of the mapped hollow RILs span a broad range, with average values spanning 40% to 50% for individual mappers at each site, even for relatively few mapped polygons (see mappers 1 and 4 at Site 3). Nevertheless, there is overall considerable overlap in the distribution of values across all mappers.

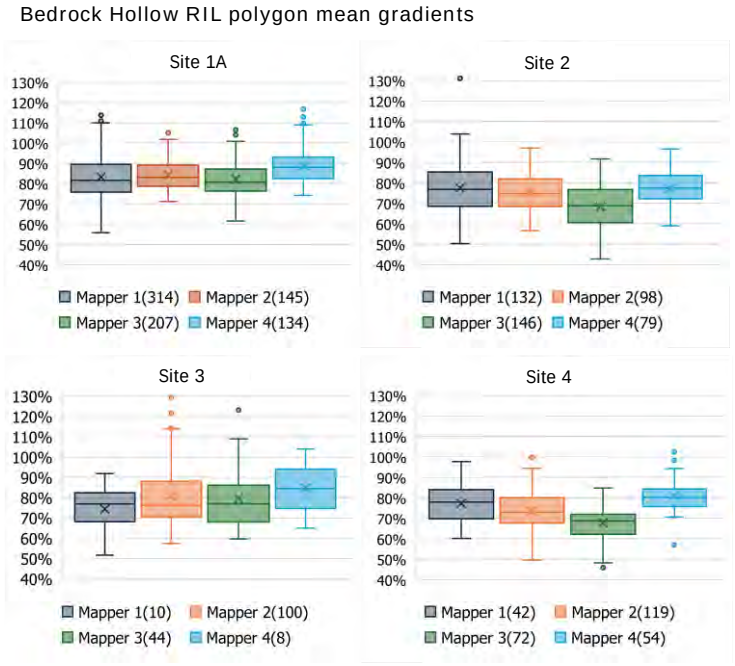


Figure C.15. Box and whisker plots showing distributions of mean gradient for each mapper's set of ~~bedrock hollow~~bedrock hollow RIL polygons for each study site. Boxes show extent of the 2nd to 3rd quartiles, whiskers show lower limit of the 1st quartile and upper limit of the 4th quartile, excluding points that fall outside 1.5 times the inter-quartile range (size of the box), which are plotted as circles.

Figure C.16 shows cumulative distributions of tangential curvature for each of the mapper's sets of ~~bedrock hollow~~bedrock hollow RIL polygons. Tangential curvature is measured cross slope and provides an indication of topographic convergence and divergence. Here we plot convergent zones as positive, divergent zones as negative. Planar slopes, even though they may be steep, have tangential curvature near zero. With the cumulative distributions, we can see what proportion of the mapped hollow area was convergent (positive values), what proportion was planar (near zero) and what proportion was divergent (negative values). Using zero as the cutoff between convergent and divergent topography, we see that the mapper's polygons included from about 25% to 35% of their area in divergent zones for Site 1A. These proportions differ by mapper and with study sites. It is important to note that, while the models used tangential curvature to delineate landform polygons, the EPs used visual observations of contour layers from DEMs and there is no minimum tangential curvature value defined for a convergent slope in the rule criteria.

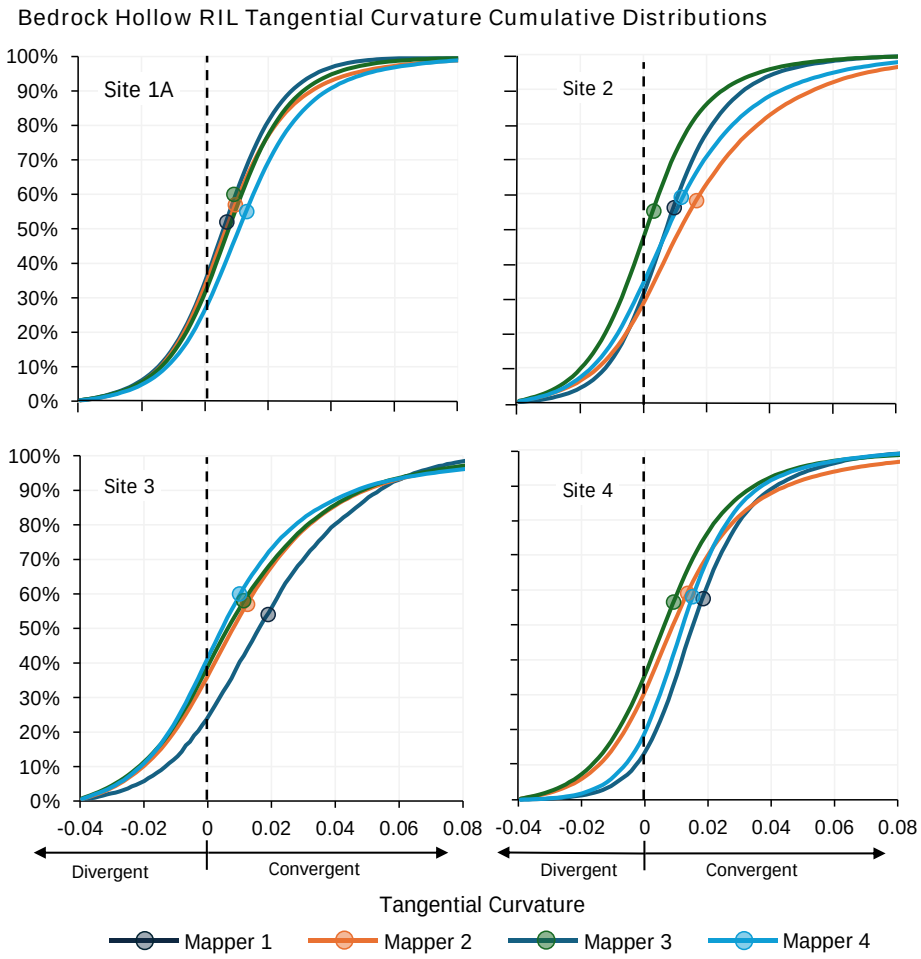


Figure C.16 Cumulative distributions of tangential curvature for the total area from each mapper's set of polygons for each study area. Circles show mean values. Dashed lines indicate zero curvature.

Figure C.17 shows box and whisker plots for the mean tangential curvature of each polygon in each mapper's sets of ~~bedrock-hollow~~bedrock hollow polygons. Similar to gradient, each mapper's polygons exhibit a range of mean values and these ranges differ between mappers.

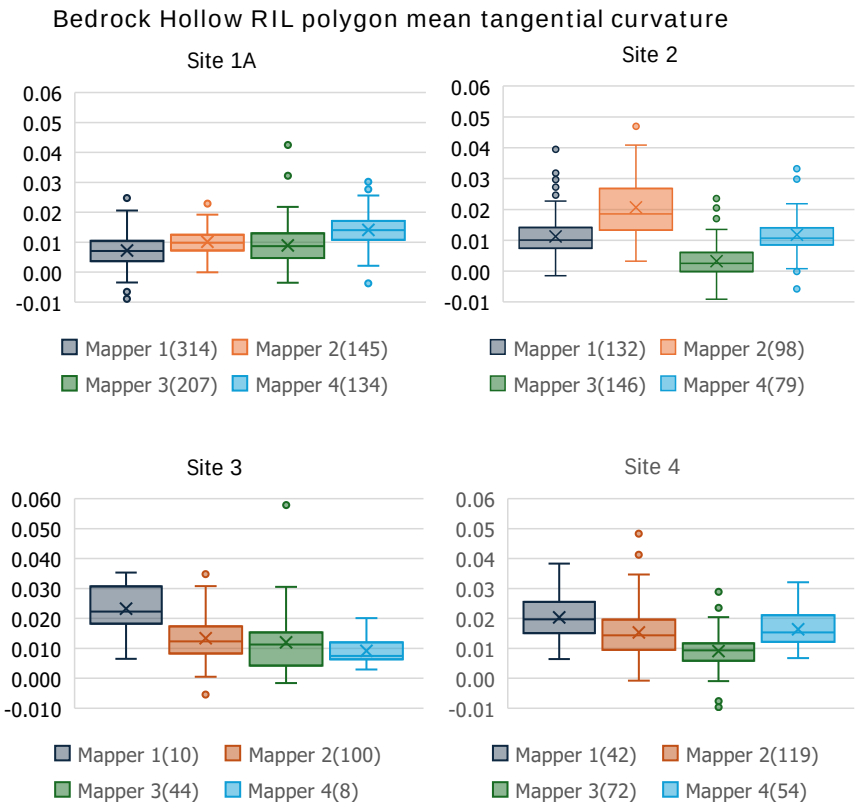


Figure C.17. Box and whisker plots of mean tangential curvature for each mapper's set of bedrock hollow RIL polygons at each study site.

We have looked at two of the primary topographic attributes used to delineate bedrock hollow RILs. The downslope extent of a hollow RIL must also be inferred from the DTM topography for this DTM-based mapping. As described in the Board Manual, the portion of a bedrock hollow most subject to soil failure does not extend further downslope than the point where a channel has formed (unless a landslide has recently occurred); that is, to the channel head. As discussed previously, channel-head locations are not static, but can migrate up and downslope depending on the history of storms and landsliding. Given the potential to resolve meter-scale features with lidar, it is feasible, with very high-quality lidar, to see a knick point along a swale axis marking the

location of a channel head[5]. Given the dense canopy cover found in these locations, however, it is not generally feasible to resolve channel extent directly and so the downslope extent of a hollow RIL must be inferred from other cues. This introduces a high level of subjectivity, which we see using frequency distributions of the maximum contributing area encompassed within each polygon. For any point on a hillslope, the contributing area is the upslope zone that can contribute infiltrating water from rainfall or snow melt that flows downslope through the soil column below that point. Contributing area increases downslope at a rate dependent on the degree of upslope convergence (see Appendix E), so the maximum upslope contributing area for a polygon depends on both the shape of the polygon, curvature within the polygon, and its downslope extent. C.18 shows box-and-whisker plots of the maximum contributing area for each mapper's sets of polygons. These are plotted in logarithmic scale, so each unit increase represents one order-of-magnitude difference in contributing area. We see a large range of values for each mapper, spanning three to four orders of magnitude. We also see that variability between mappers is inconsistent between sites, though Mapper 1 consistently has the smallest values.

[5] Given the potential for spatial intermittency of channel formation at the upslope extent of channel formation, e.g., associated with large wood or boulders trapping sediment, there may be more than one channel head.

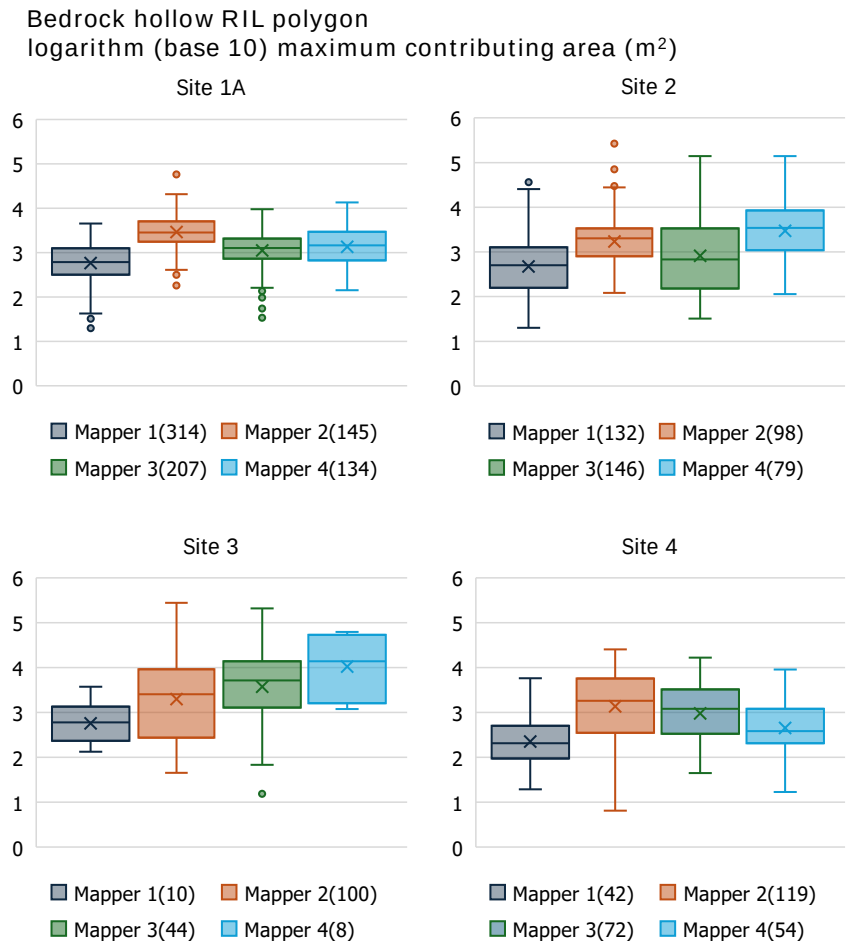


Figure C.18. Box and whisker plots of the base-10 logarithm of maximum contributing area (m²) for each mapper's set of ~~bedrock hollow~~bedrock hollow RIL polygons at each study site. Numbers in parentheses indicate the number of polygons in each mapper's set.

These plots all show some differences in the range and frequency distributions of topographic attributes both among mappers and study sites. They also show substantial overlap and similarities. Given these similarities, it is hard to explain why there is such little overlap between pairs of maps in some cases (e.g., Mapper 1 matched less than 1% of Mapper 4's bedrock hollow RIL polygons at Site 3). The lack of overlap between pairs of mapper's maps suggests perhaps that each mapper's map is incomplete – the mappers are missing some RILs.

APPENDIX D. MAPPING CRITERIA SPREADSHEET

Commented [PJ275]: This is not legible

	M1	M2	M3	M4	M5
SCALE RANGE USED					
TOPOGRAPHIC DERIVATIVES					
Slope class intervals					
Slope class density					
Contour interval					
Hillshade					
Curvature (data driven)					
GEOMORPHIC CRITERIA					
Planform curvature					
Profile curvature					
Slope gradient					
Landforms avoided or excluded					
Inner gorge slope break					
Inner gorge height					
Inner gorge relative to channel confinement					
LANDFORM CLASSES					
Inner gorge					
Bedrock hollow					
Deep-seated toe					
Outside meander bend					
Other					
OTHER LAYERS USED					
Air photos					
Landslide inventories					
Roads					
Streams					
Rayonier's linework (Calawah)					
COMMENTS					

D.1. Criteria spreadsheet used to refine manual mapping protocols.

APPENDIX E. ELEVATION DERIVATIVES

Commented [PJ276]: This appendix can be improved by simplifying language and using more active voice.
Also, need to be clear that only BHs and IGs were mapped, not all RILs. I'd search and replace RILs with bedrock hollows and inner gorges.

The manual mapping and computer delineation of ~~RILs~~bedrock hollows and inner gorges for this sub-project were based solely on interpretation of topography inferred from digital terrain models (DTMs) derived from lidar data. Lidar DTMs are available on the Washington Lidar Portal. These DTMs consist of regular grids of elevation values at a 3-foot horizontal spacing. The elevations are interpolated from the lidar laser-signal reflections inferred to be from or very near the ground surface, referred to as ground returns. In areas of forest canopy, most of the laser signals are reflected from vegetation, so the spatial density of ground returns varies spatially over the area of a DTM depending on canopy density and to some extent on ground steepness (Petras et al., 2023; Su and Bork, 2006). For DTMs from the Lidar Portal, elevations between ground returns have been interpolated using a triangulated irregular network (TIN), in which planar triangular facets connect the ground-return points and elevations at the grid-point locations are measured on the facets. Where ground returns are spaced further apart than the DTM grid spacing, these triangular facets are visible in shaded relief images. The length scale over which ground features can be resolved thus depends on the ground-return spacing in conjunction with the spacing between DTM grid points. Although a 3-foot DTM grid spacing should be sufficient to resolve features down to 10 to 15 feet in length, the actual detail that can be resolved depends on the ground-return spacing. This is not an issue for mapping ~~bedrock-hollow~~bedrock hollow RILs, which typically span lengths of 30 feet or more, but may be a constraint for resolving inner gorges, particularly because canopy density tends to be high along riparian zones.

Topographic attributes used to identify and delineate RILs can be estimated at each DTM grid point (Maxwell and Shobe, 2022). These estimates are made by fitting a smooth surface to a set of surrounding grid points and finding the gradient, aspect, and curvatures of that surface at the center point (Florinsky, 2016). The degree to which these estimates match measurements taken on the ground depend on many factors: the length scale and method of making the measurements on the ground, the degree of surface roughness, the length scale over which that smooth surface was fit on the DTM, the type of surface used, the algorithm used, the methods used to implement that algorithm, and the degree to which changes in the elevations of the surrounding points on the DTM match corresponding changes in elevation on the ground, which depend on the ground-return spacing of the original lidar data. This is a long list of potential things to go wrong, so it is important for anyone using DTMs and DTM derivatives to be aware of these issues. With care, lidar DTMs can provide detailed measures of topographic characteristics consistent with ground observations and provide the ability to map these characteristics over areas far greater than feasible solely with ground surveys.

In rough terrain, such as that prone to landslides, metrics based on measures of elevation change over some specified distance, such as gradient and curvature, can vary with the distance over

which those changes in elevation are measured. A practitioner ~~on the ground~~in the field measuring gradient along a hollow axis or side slope will ignore tree-fall pits and down trees and measure the slope over a length scale sufficient to average out those small-scale variations. This is true on the ground and for a DTM. With a DTM, the algorithms for fitting a surface need to span at least two DTM cells. With lidar DTMs, this length should extend far enough to span several ground returns. Likewise, these measurements should be made over length scales pertinent to the processes and physical factors that drive soil failure, a minimum for which might be inferred from the extent of typical shallow landslide scars – on the order of 10m (30ft).

Only recently have algorithms for measuring gradient and curvature over arbitrary lengths become available in widely available software (e.g., Ilich et al., 2023). For the virtual-watershed (Section 3.1.3) analyses and overlay comparisons of mapper's RIL polygons (Section 4) presented in this report, we used an algorithm implemented in the Netstream suite of programs (Miller, 2003) that measures gradient over a specified length looking either upslope or downslope from each DTM grid point. At a grid point, this algorithm fits a polynomial surface to nine points: the grid point and eight points on the circumference of a circle of specified radius centered over the grid point (Shi et al., 2007). Elevations of the points on the circumference of the circle are determined using bilinear interpolation from the four corner points of the DTM cell in which the point falls. To focus solely on the upslope or downslope directions, the elevations on the circle are mirrored on either the downslope or upslope direction and the surface fit again. This algorithm provides measures of gradient and curvature over arbitrary lengths (longer than two DTM cell widths) with no increase in processing time with increasing length. The coordinate systems used for these DTMs reference the DTM grid points on a horizontal plane, so lengths between grid points and the specified radius of the circle represent horizontal distances. These may be rotated at each grid point to measure gradient over slope distance, which was done for the analyses here to be consistent with the way that gradient is generally measured on the ground using a hand-held clinometer. Software to implement this algorithm is available in the TerrainWorksUtils R package.

Several other landscape features useful for geomorphic interpretation of a landscape and delineation of landforms are derived from a DTM. These include tracing of a channel network and delineation of the valley floor. Methods for tracing a channel network from a DTM are well developed (e.g., ESRI's Arc Hydro), but also continually evolving. The methods and algorithms used for this sub-project are described in Miller et al. (2015) with software updates since 2015 to be consistent with USGS specifications for development of hydrography datasets for the 3DHP program. The valley floor is delineated by detrending valley elevations relative to channel profiles. This is done for each DTM grid point within a specified radius of the traced channel centerlines. The channel elevation associated with a grid point is estimated using the distance-weighted average of all channel grid points over a larger specified radius. Subtracting the averaged channel

elevation from the grid-point elevation gives height above the channel in meters. This value is then divided by the distance-weighted average channel depth to give the elevation above the channel in terms of channel depths. Channel depths are estimated using regional regressions to contributing area and mean annual precipitation (Kresch, 1998a; Kresch, 1998b; Magirl and Olsen, 2009). Measures of height above the channel in terms of channel depths provides a measure independent of channel size that may correlate with flood recurrence (see this blog post). We used a height above the channel of five channel depths to delineate the potentially active floodplain. This may roughly correspond to the valley-floor area inundated during a 100-year flood (Clarke et al., 2008).

Commented [PJ277]: If this is important, can the regressions be included in a table?

The unchannelized swales draining to the upslope-most extent of channels can be primary sites of upslope shallow landsliding if they are steep enough to promote soil failure. These swales are referred to as bedrock hollows or zero-order basins (Dietrich et al., 1986). Not all zero-order basins are prone to landsliding, but by our definitions, all bedrock hollow RILs lie within zero-order basins. Once a channel network has been defined, zero-order basins can be delineated on a DTM as the area draining to the channel head. This is described in greater detail in Section 3.1.3.

Manual mapping of bedrock hollows and inner gorges~~RIL~~s using DTM basemaps relies on interpretation of topographic attributes measured from the DTMs or inferred from DTM derivatives. As described in Section 3.1.1, mappers used shaded-relief images, slope classes, and contour lines derived from DTMs interpolated from lidar point-cloud ground returns. Shaded-relief images provide a view of the terrain from which basic landform features and their locations relative to each other are visible: valleys can be distinguished from ridges, and convergent slopes from planar and divergent slopes. To map ~~bedrock hollow~~bedrock hollow and inner -gorge landforms~~RIL~~s, a mapper must decide precisely where the landform~~RIL~~ criteria are met and draw a line delineating those areas inside and outside of each landform~~RIL~~. Each type of RIL landform is found in specific landscape locations. For example, hollows occur on steep hillslopes and are separated from the valley floor by steep-sided small channels that serve as debris-flow corridors. Inner gorges occur at the base of steep slopes adjacent to channels. These general locations can be determined from the shaded relief (Figure E.1). Mappers then use variations in the shaded relief image, the slope classes, and the contour lines to identify where slope thresholds are met, where slope breaks occur, and to determine the degree of hillslope convergence. Their choices are guided and constrained by the RIL landform descriptions provided in the Board Manual. Nevertheless, each mapper relies on their own unique field observations and experience mapping RILs in making those choices. Hence, each mapper will have a unique understanding and interpretation of the data available from a DTM, which can translate to differences in the RIL polygons drawn by different mappers. Additionally, each mapper brings their own degree of meticulousness, patience, and GIS (digitizing) skills to this task, which can also contribute to

differences in RIL polygons drawn by different mappers, or even by the same mapper at different times.

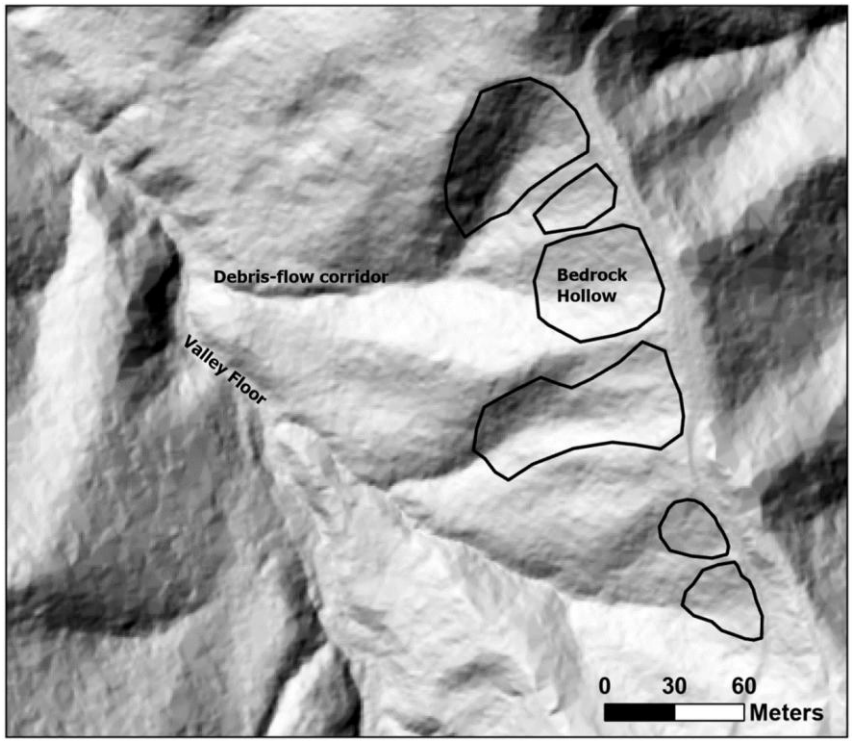


Figure E.1. A shaded relief image from Site 1A showing several landscape features. The valley floor extends diagonally from the southeast to northwest across the figure. A series of mapped ~~bedrock hollow~~bedrock hollow RILs lie on the upper portion of the hillslope.

~~Confusion matrices and their derivatives show how the areas identified as inside and outside of RILs are different and alike among different mappers. Another way of comparing RILs is to look~~Landform polygons can be compared by looking at differences and similarities in the range of topographic-attribute values contained within the RIL polygons. For example, thresholds in slope gradient are an important component in determining the presence and extent of a~~n~~ landform~~RIL~~, but one mapper might extend their landform~~RIL~~ boundaries into lower-gradient areas than another mapper. We can quantify the difference using statistics of the frequency distributions of slope values within each mapper's polygons, which together with visual comparisons of those distributions using histograms, box-and-whisker plots, and cumulative frequency distributions, provide an additional measure of the similarities and differences in the way different mappers resolve landforms~~RIL~~s using DTM base maps. These distributions are not

Commented [PJ278]: This paragraph could use a clearer introductory sentence to help the reader understand what to expect. I was expecting the paragraph to be about confusion matrices. But I think the paragraph is about multiple ways to quantify the differences between mappers (not just confusion matrices).

indicators of the degree to which any set of RIL polygons delineate potentially unstable slopes; that assessment is a task for the next sub-project. Rather, they simply show differences in the way that different mappers (or the same mapper at different times) use DTM derivatives to infer the location and extent of RILs. We illustrate these concepts using a small subset of the EP mapped ~~bedrock hollow~~bedrock hollow RIL polygons at Site 1B. In subsequent sections, we will use these methods to characterize the distribution of topographic attributes found over the entire set of EP mapped and modeled polygons.

Numerous indices and derivatives can be calculated with a DTM to characterize aspects of the topography. We focus on gradient and curvature, both of which enter explicitly into the narrative descriptions of RILs provided by the Board Manual. Gradient and curvature can vary continuously with position and both can be measured from any point on the ground surface. We can also measure gradient and curvature at each DTM grid point. Each measurement applies to the area associated with the DTM point, referred to as a DTM "cell" or pixel. The distribution of DTM-cell values measured within an RIL polygon then indicates the proportion of RIL-polygon area falling within any increment of gradient or curvature. We can look at how that distribution differs between polygons and across sets of polygons.

For example, Figure E.2 shows the spatial distribution of gradients measured for DTM cells that are 0.9144 meters (3-feet[6]) on an edge (measured horizontally). The gradients are divided into slope classes. Five-foot contour lines are also shown. The mappers use the shaded relief image to provide the positional context – valley floor, ridge top, divergent and convergent slopes – and use the slope classes and contour lines to gauge how steep and how convergent different portions of the hillslopes are. Mappers must also gauge the spatial scale over which to mentally integrate variability in slope and curvature. In Figure E.2, you can see changes in the gradient classes over length scales of a few meters, with isolated zones of very steep gradient (greater than 90%), separated by zones of lower gradient. Mappers must determine over what length to average out this variability. A typical bedrock hollow is larger than the length over which these DTM-measured gradients vary, so the delineated hollow polygons encompass a range of gradients. Each mapper must decide what range is appropriate and where to place the polygon boundary to encompass that range.

[6] We use metric units for all analyses. However, Washington DNR distributes data using United States customary units with length given in U.S. Survey Feet. However, this unit has been depreciated (https://www.nist.gov/pml/us-surveyfoot) in favor of the international foot, which may apply to newer datasets.

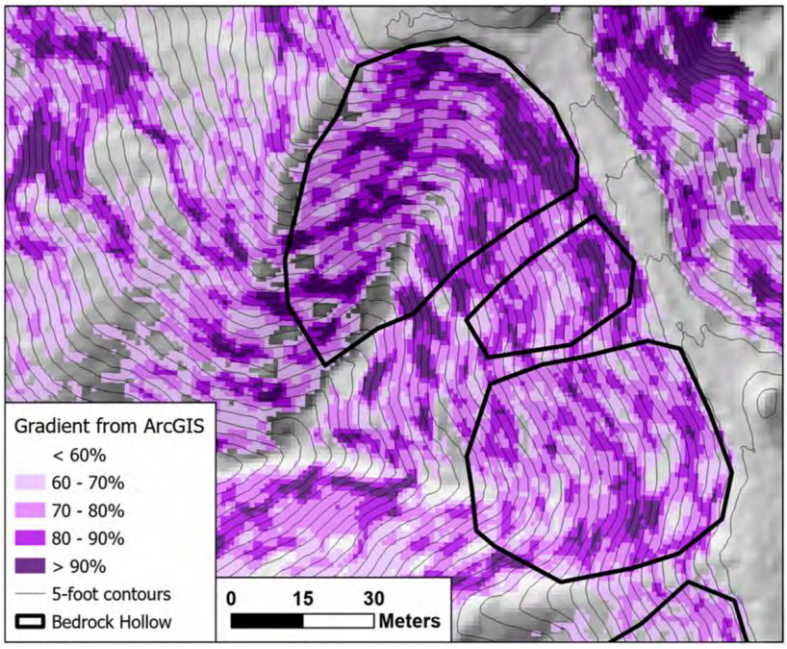


Figure E.2. Gradient classes, contour lines, and one mapper's bedrock hollow RIL polygons overlain on the shaded-relief image.

Different mappers make different decisions, even when all are working from the same base RIL definitions, using the same data, and working at the same map scale (1:500). Figure E.3 shows ~~bedrock-hollow~~bedrock hollow RIL polygons from three other mappers (yellow outlines) overlain on the original three (black lines). Mappers place polygons in the same general landscape positions, but different mappers can end up with different numbers of polygons and each mapper's polygons have unique shapes and sizes, even for the same bedrock hollow. Figure E.4 shows a bar chart with the mean gradient within each polygon. The means differ for each individual mapper's polygons and the mean for specific hollows varies a bit between mappers. The range in mean values is not large, generally within 10%, and there is general consistency in the ranking.

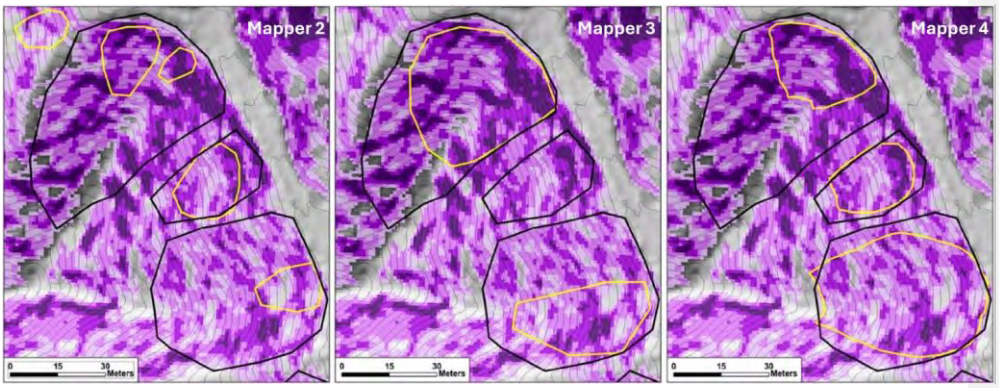


Figure E.3. Three other mappers' polygons (yellow lines) overlain on the original three polygons.

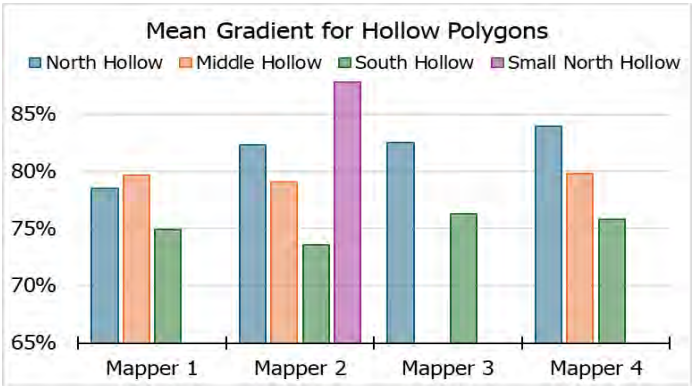


Figure E.4. Bar chart of mean gradient within each polygon.

Commented [PJ279]: It would be helpful if the colors in the bar chart matched the colors on the map

Another way to examine the distribution of gradient values within mapped polygons is to look at the proportion of the polygon that exceeds a specified gradient value. The Board Manual indicates 70% as the lower threshold for hollow and inner gorge gradient; not that the entire RIL must exceed this threshold, but that 70% is the lowest gradient at which shallow landslides are typically observed. In Figure E.5 we plot a bar chart that shows the proportion of each mapped polygon that exceeds 70% gradient. Again, we see variability in this value for different hollow polygons drawn by a single mapper and for polygons from the same hollow drawn by different mappers.

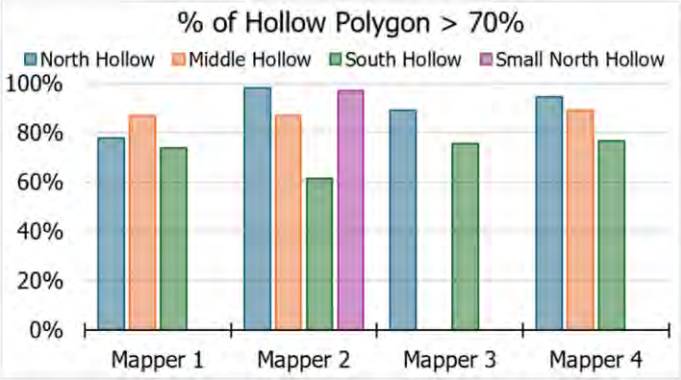


Figure E.5. Proportion of each mapped hollow with gradient exceeding 70%.

The issue of what length scale to use applies also to the measurement of topographic attributes directly. Small features like tree-throw pits might be resolved if gradient is measured over 3 meters, but not if gradient is measured over 30 meters. A practitioner on the ground may not be interested in those tree-throw pits and will measure gradient over length scales commensurate with the size of the features of interest, probably in the range of ten to thirty meters for bedrock hollows and inner gorges. Figure E.6 shows gradient classes for a measurement length of 15 meters. The spatial variability in measured gradient is less than that visible in Figure E.2 above, where gradient was measured over about 3 meters. Over 15 meters, smaller-scale features are smoothed out[7]; the patches within a single gradient class are larger and more contiguous. Fewer small zones of steep gradient are resolved, so the mean gradient, when measured over this length, is less than when measured over a smaller length, as shown in the bar chart of Figure E.6.

[7] Note that some of the small-scale features indicated by variability in gradient measured over 3 meters may result from noise in the DTM, e.g., from laser reflections off of above-ground vegetation that are interpreted as reflections from the ground.

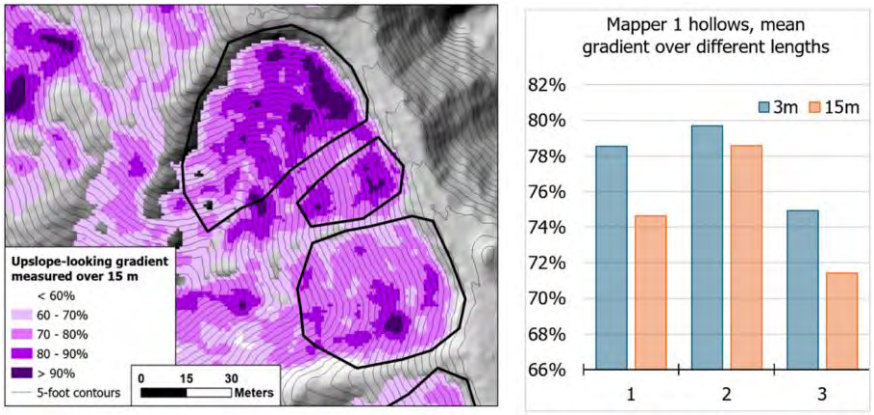


Figure E.6. Gradient looking upslope measured over a length of 15 meters. Bar chart compares the mean gradient for each hollow using gradient-measurement lengths of 3 meters and 5 meters.

Figure E.7 shows tangential curvature measured over a length of 15 meters with the four sets of mapped bedrock hollow RIL polygons. Tangential curvature provides a measure of how convergent or divergent the topography is. Planar slopes have zero curvature. As plotted here, negative values indicate divergent topography, positive values indicate convergent, with larger values indicating greater convergence. The mappers did not use plots of curvature for mapping RILs, but rather relied on the shaded-relief image and the contour lines to assess the degree of curvature. The ~~bedrock hollow~~bedrock hollow polygons tend to straddle areas of high curvature, but placement with respect to those high-curvature zones varies between mappers. Mean curvature for each polygon is shown in Figure E.8. There is a variability between different hollow polygons, with mean values varying over a factor of two, but there is general consistency as to which hollows have greater or lesser curvature.

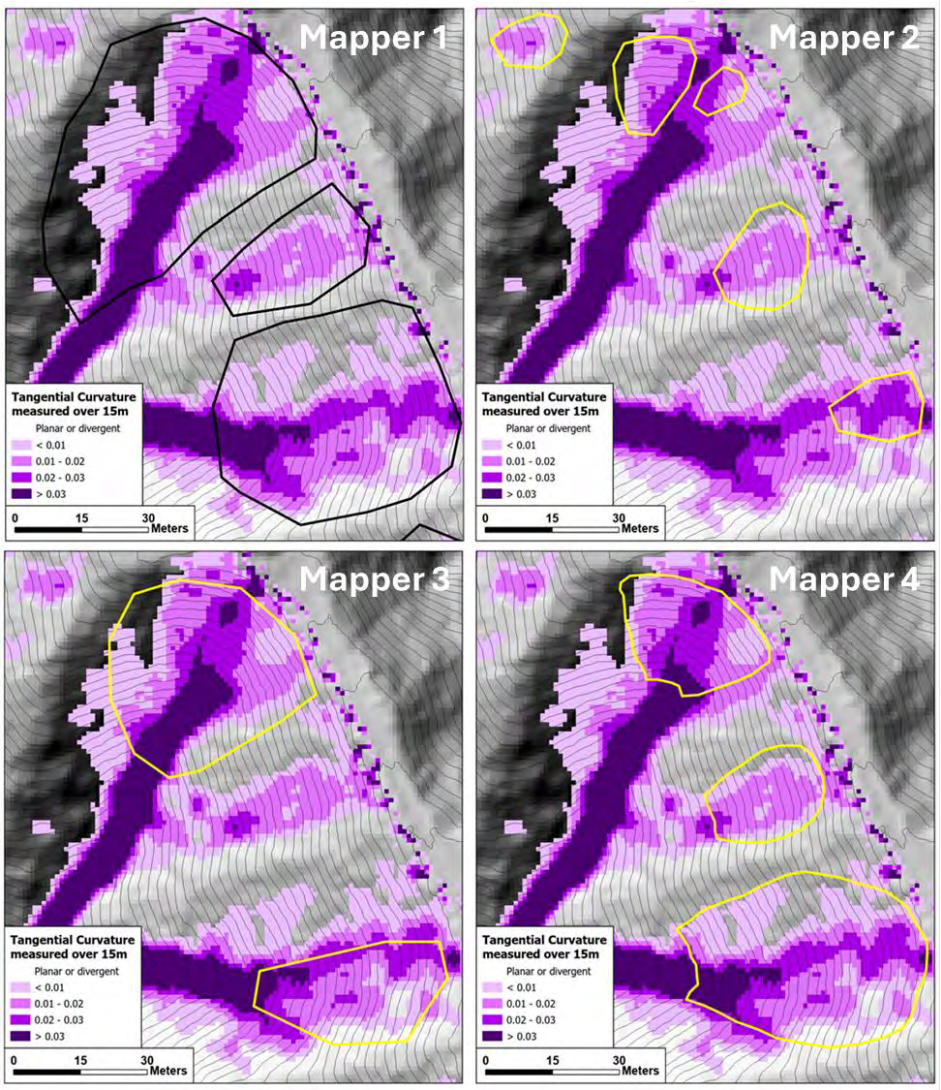


Figure E.7. Tangential curvature measured over a 15-meter length with mapped bedrock hollow polygons overlain.

Commented [PJ280]: Label the mapper for each window. Also, rephrase the caption to be a complete sentence. Same comment for E.10

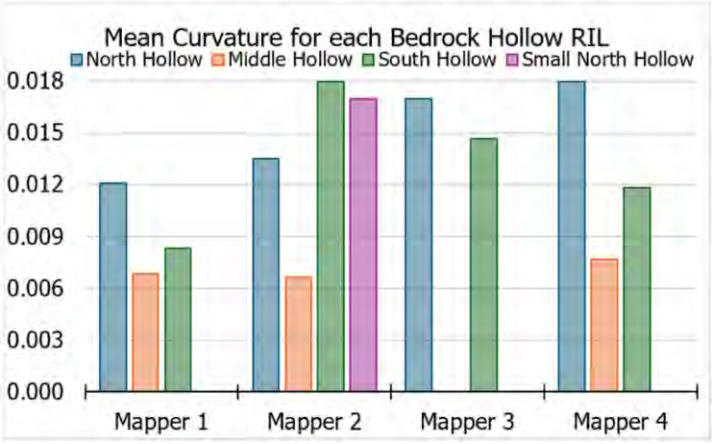


Figure E.8. Mean curvature for each mapped ~~bedrock hollow~~bedrock hollow RIL polygon.

We can also look at the distribution of values across all mapped RIL polygons, as shown with the histograms and corresponding cumulative distributions for gradient in Figure E.9. These types of plots provide both quantitative and visual measures of the degree of similarity and differences for the sets of polygons drawn by different mappers.

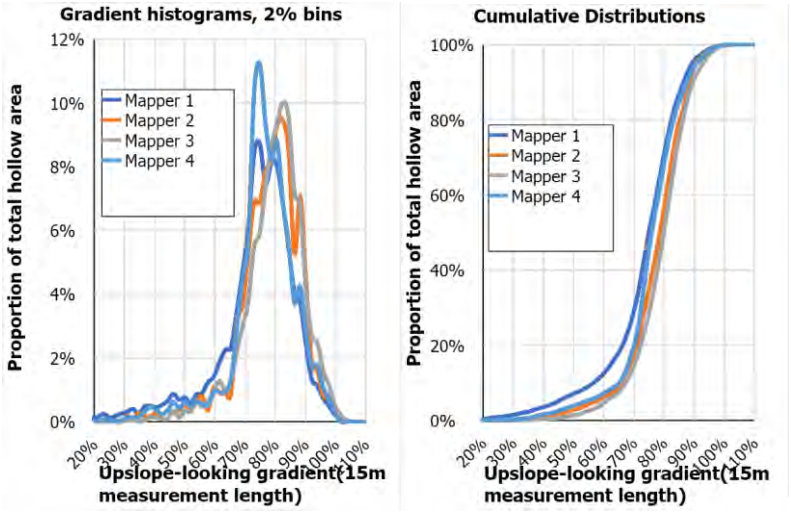


Figure E.9. Histograms and cumulative distributions of gradient over the area encompassed by the three hollows for each mapper.

Another important feature of the ~~bedrock-hollow~~bedrock hollow polygons is the distance they extend downslope. Looking at the previous figures, you can see considerable variability in how far downslope different mappers extend the polygons they draw. Conceptually, a hollow extends to the channel head where soil depth is limited by fluvial erosion and where soils are frequently saturated to the surface without failing. The location of a channel head is notoriously difficult to discern remotely. We can, however, compare the locations of potential channel heads identified by the different mappers by comparing the maximum contributing area encompassed within their ~~bedrock-hollow~~bedrock hollow polygons.

For these analyses, contributing area is estimated from the DTM using the D-infinity flow-direction algorithm (Tarboton, 1997), which allows downslope dispersion (Wilson et al., 2008). Within a bedrock hollow, water infiltrating the soil will tend to flow towards the axis of the hollow so that the surface area contributing flow increases downslope along the axis (Figure E.10). The rate of increase depends on the shape of the hollow, with wider hollows having larger rates.

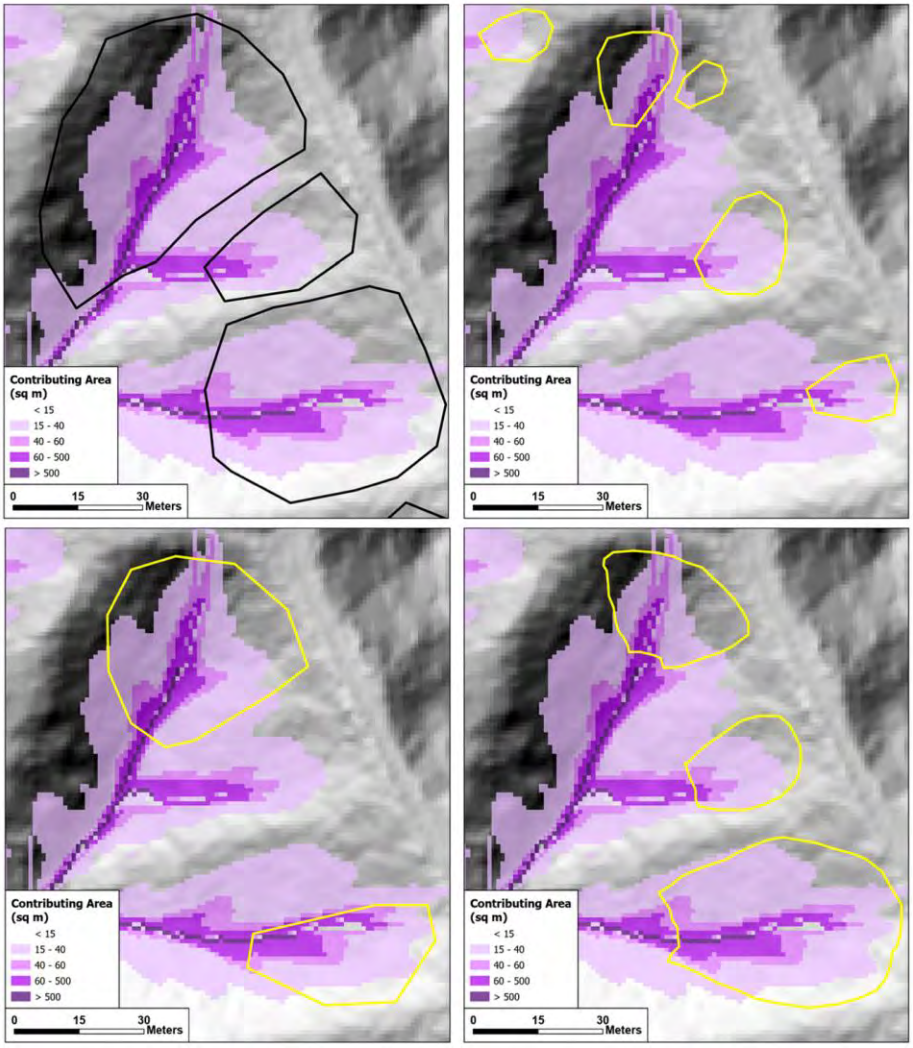


Figure E.10. Upslope contributing area with each mapper's polygons overlain.

Within a mapped hollow polygon, the maximum contributing area occurs where modeled flow lines exit the polygon at the downslope end and so provides a measure indicative of the downslope extent of a mapped polygon. Figure E.11 shows the maximum contributing area for each mapper's polygons. There are large differences between mappers for the north and south

hollows, less for the smaller middle hollow. These differences reflect the different downslope extent of the mapped hollows, as seen in Figure E1.0.

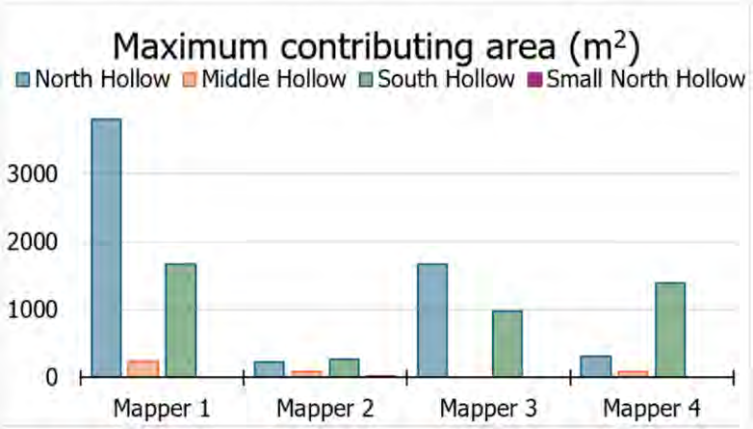


Figure E.11. Maximum contributing area at downslope exit point for each hollow.

Here we have looked at how different experienced mappers, all using the same criteria for identifying RILs and mapping at the same map scale, offer different interpretations of the same data in determining where RILs exist and where their boundaries lie. These differences in interpretation are manifest not only in differences in the locations, size, and shape of mapped RIL polygons, but also in the range and distribution of topographic attributes within the polygons. Confusion matrices and their derivatives provide a way to quantify the degree of agreement and disagreement between mappers in terms of the area mapped as inside and outside of RILs. The frequency distributions and statistics of those distributions provide a way to compare each mapper's polygons to the topographic criteria used to define an RIL and, in the next sequence of sub-projects, to the topographic attributes of landslide locations identified with the landslide inventory. In Section 4 we looked at confusion matrices and frequency distributions to compare results among mappers for the four study sites and to compare results between mappers and computer-generated RIL polygons.

APPENDIX F. FREQUENCY DISTRIBUTIONS

Here we provide cumulative frequency distributions of selected topographic indices (Figures F.1 through F.6) over the cumulative area of all RIL polygons drawn by each mapper and the two computer models at each site. We also provide box-and-whisker plots of the average value of those indices within each mapped polygon and discuss them in Section 4.14. The frequency distributions show the range of index values encompassed within all mapped polygons. For example, we can see what the minimum and maximum gradients the mappers and the computers included within the mapped RILs. The plots (Figures F7 & F.8) show differences in the distributions across mappers and across study sites. We have no definitive rules for interpreting or judging these; rather, our goal is to provide a visual comparison of how different mappers interpret the information available from the DTMs and of how the computer-generated results compare to those interpretations.

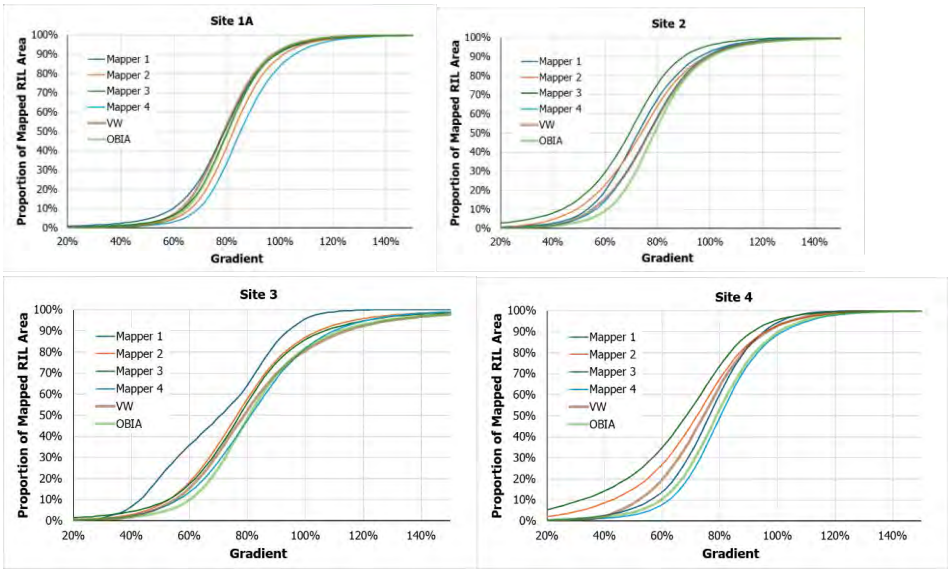


Figure F.1. Cumulative frequency distributions of upslope-looking gradient for ~~bedrock hollow~~bedrock hollow RILs for each of the four mappers and the two computer models at each of the four study sites.

Commented [PJ281]: These are too small to be able to see the differences between mappers and the computer. They need to be bigger. (same for all)

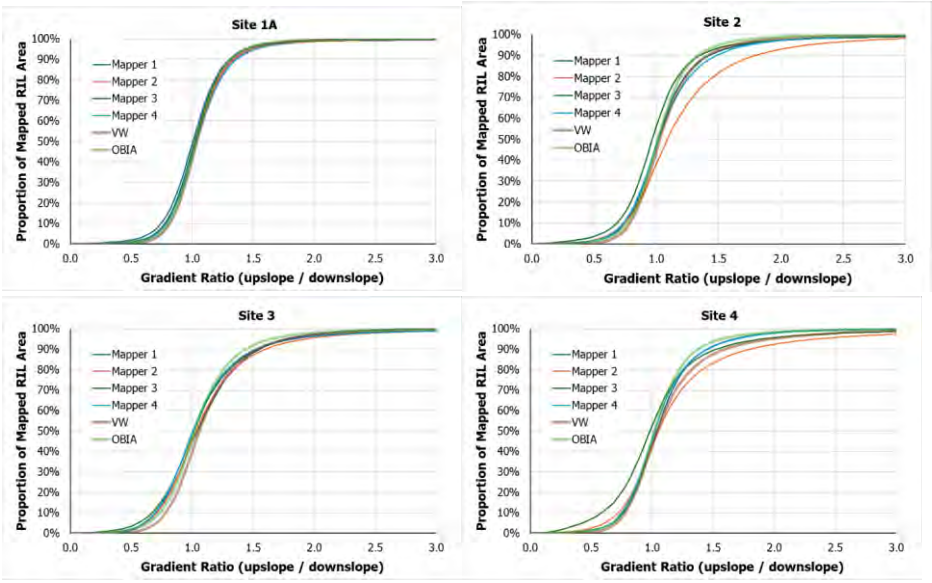


Figure F.2. Cumulative frequency distributions of the ratio of upslope-looking to downslope-looking gradient (profile curvature) for ~~bedrock hollow~~bedrock hollow RILs for all four mappers and the two computer models at the four study sites.

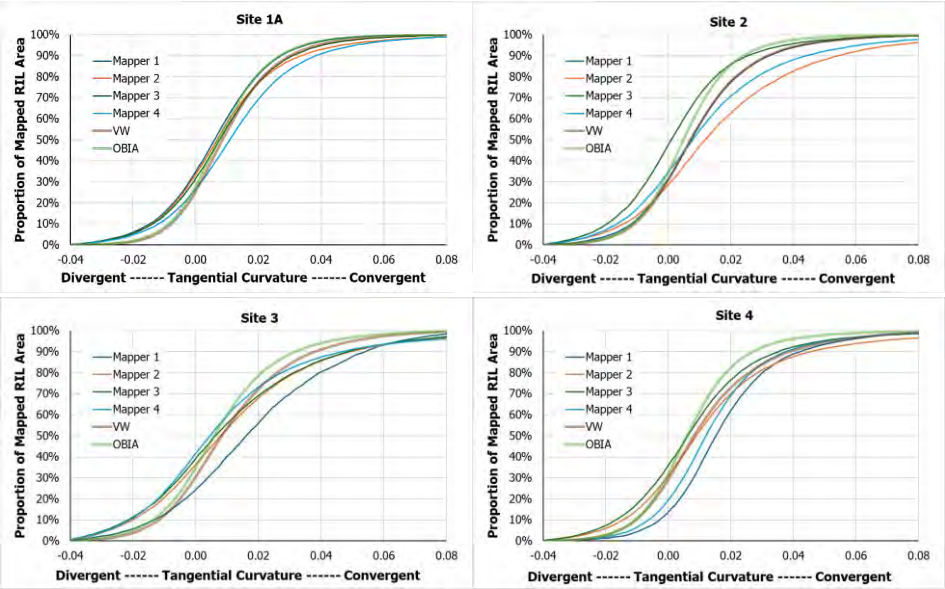


Figure F.3. Cumulative frequency distributions of tangential curvature for ~~bedrock hollow~~bedrock hollow RILs for each of the four mappers and the two computer models at each of the four study areas.

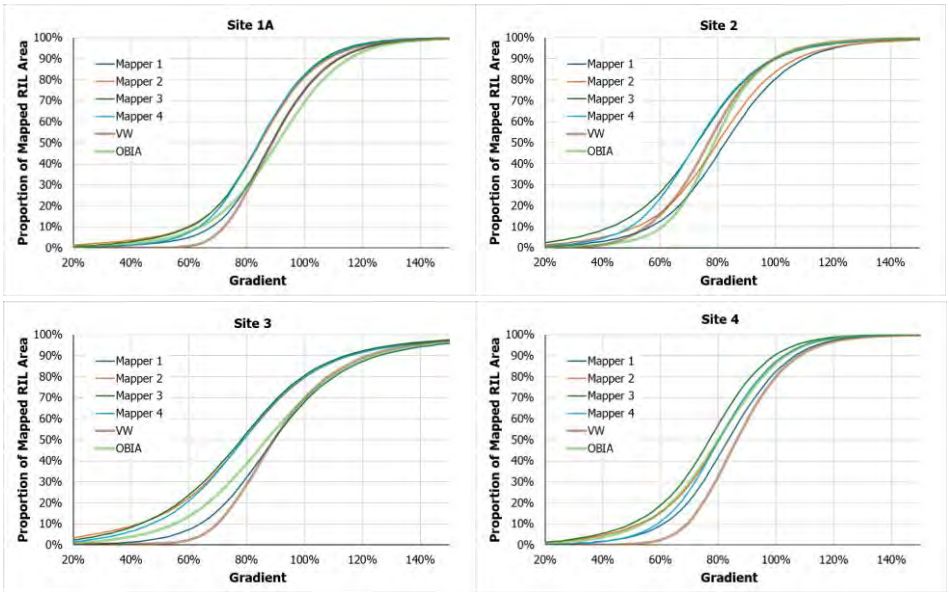


Figure F.4. Cumulative frequency distributions of upslope-looking gradient for inner_-gorge RILs for all four mappers and the two computer models at each of the four study sites.

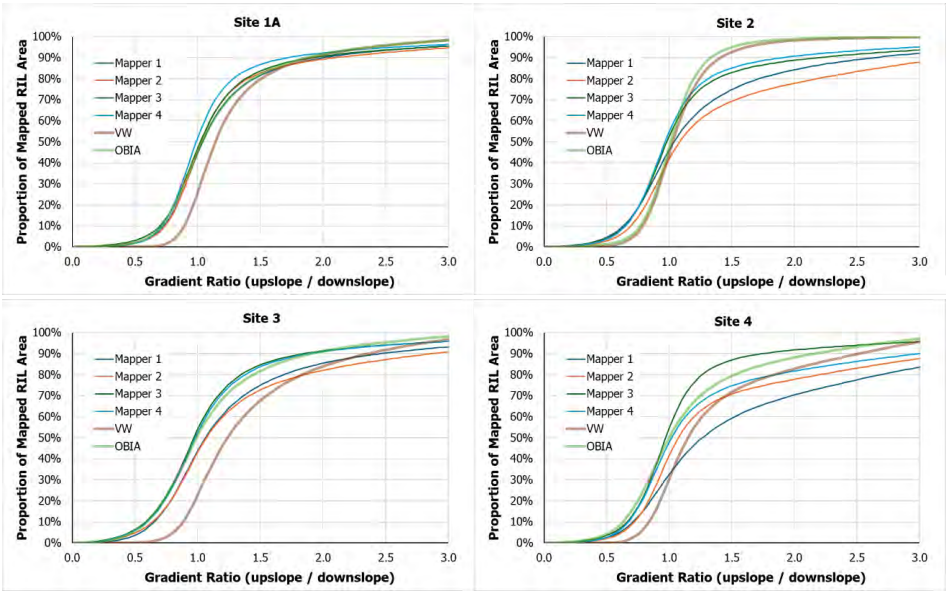


Figure F.5. Cumulative frequency distributions of the ratio of upslope-looking to downslope-looking gradient of inner gorge RILs by all four mappers and the two computer models for all four study sites.

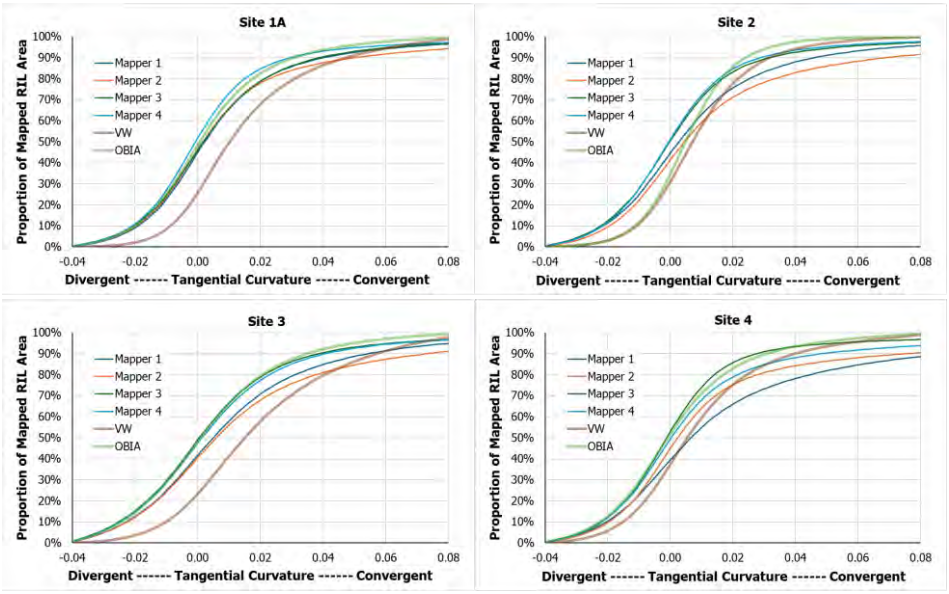


Figure F.6. Cumulative frequency distributions of tangential curvature in inner gorge RILs for the four mappers and the two computer models at all four study sites.

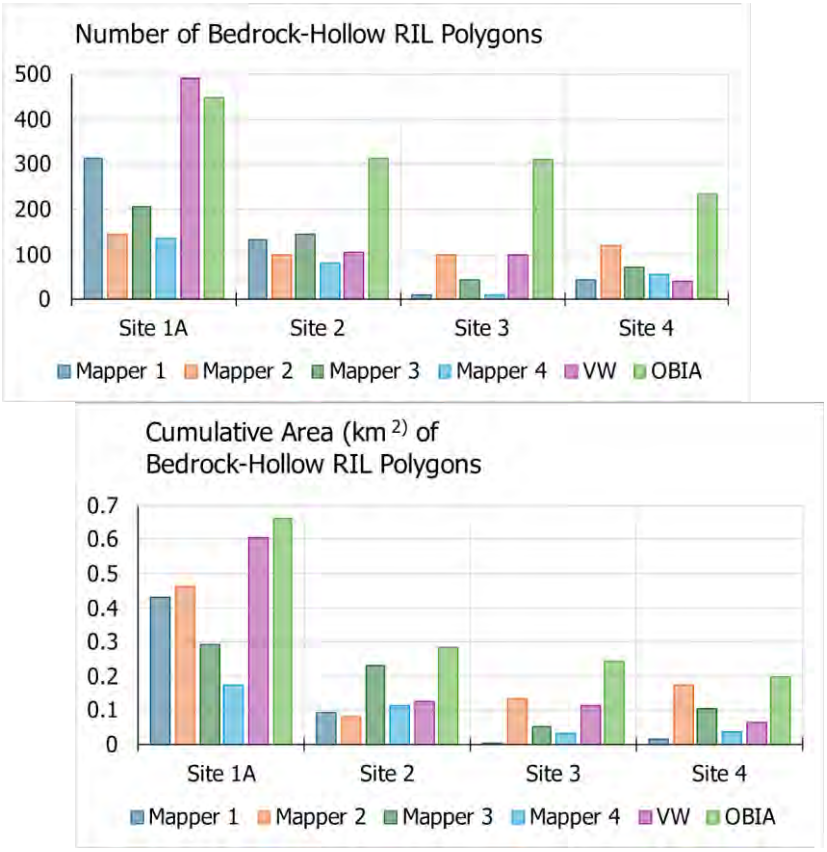


Figure F.7. Bar charts comparing the number and cumulative area of ~~bedrock-hollow~~bedrock hollow RIL polygons for each mapper and the two computer models at each study site. Study-area size and zero-order-basin density varied from site to site, so the primary point of these charts is not the change in magnitude from site to site, but the variability between mappers and between mappers and computers at a site. OBIA identifies more RILs and greater cumulative RIL area than any of the mappers and, except for Site 1A, than the VW model.

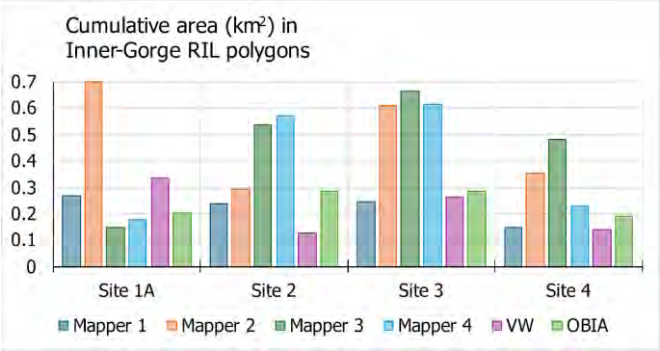


Figure F.8. Bar chart of cumulative area of inner gorges mapped by each mapper and the two computer models at each study site.

APPENDIX G. CUMULATIVE FREQUENCY DISTRIBUTIONS FOR SITE 1B – ELEVATION DERIVATIVES

To briefly review, in a virtual watershed, locations for hollows and inner gorges are initially identified based on threshold values of gradient and of the ratio of upslope-looking to downslope-looking gradient. Locations that meet these thresholds provide seed locations from which the delineated landforms can then grow outward. Seed locations for hollows are restricted to zero-order basins – the contributing area to the modeled channel head – and, for inner gorges, to zones immediately adjacent to the active floodplain or channel. The landform polygons can grow outward from the seed locations to include zones meeting second thresholds for gradient and gradient ratios. We can adjust the number, locations, and sizes of hollows and inner gorges based on the primary and secondary thresholds specified for gradient and gradient ratio. For hollows, location of the channel head is also a primary controlling factor affecting the number, location, and size of delineated hollows.

Channel-head locations can be set using a threshold value of the product of specific contributing area times channel gradient squared (AS2, see Appendix B). Figure G.1 shows cumulative distributions of AS2 values at the point of maximum contributing area at the downslope end of each mapped polygon for each of the five mappers. There is still sufficient variability across mappers to make the choice of a threshold uncertain. Mapper 5 (light blue line) had extensive experience mapping landslides in this region, so their polygons probably provided the best indicator of where channel heads are found. We set the AS^2 threshold value to the median distribution for Mapper 5 (1,000 m), so that most of the virtual-watershed modeled hollows will extend only downslope to that value (Figure G.1). That means that about half of Mapper 5's, about a third of Mapper 2's, and less than 10% of Mapper 1's, 3's and 4's hollows will extend further downslope than the modeled hollows.[8]

[8] In hindsight, we probably should have calibrated the VW model specifically to Mapper 1's polygons, because we used Mapper 1's maps as the reference to compare with the computer-model outputs.

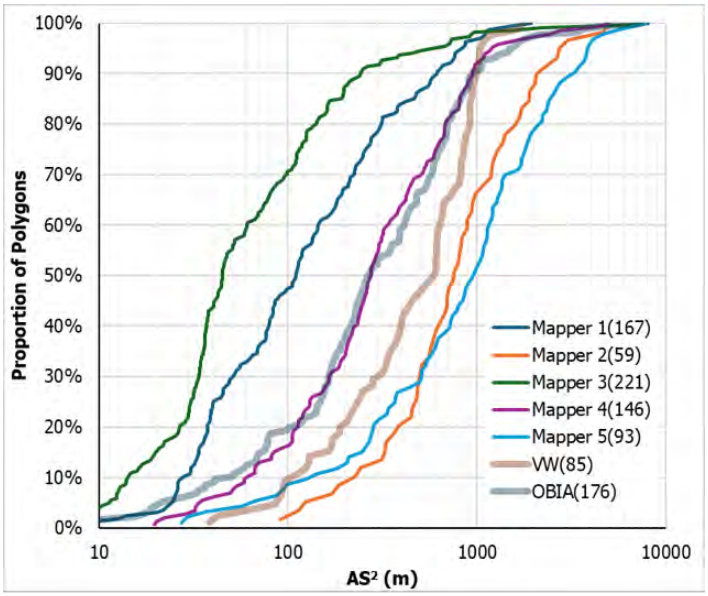


Figure G.1. Cumulative frequency distributions for the product of specific contributing area and channel gradient squared (AS2) at the point of maximum contributing area for each mapped ~~bedrock hollow~~bedrock hollow RIL polygon for Site 1B. The number of polygons drawn by each mapper are indicated in parentheses in the legend. The computer-modeled results are shown by the thicker, slightly transparent lines.

Figures G.2 and G.3 show cumulative frequency distributions of gradient and the ratio of upslope-looking to downslope-looking gradient for the five mappers. Based on these graphs, a primary hollow gradient and gradient ratio of 80% and 1.0 were chosen with secondary values of 40% and 0.4. As seen in these figures, the VW-modeled RILs still included a portion of their area below the threshold values. This is because filling in of holes and embayments extended the polygons into lower-value regions. Figure G.4 shows cumulative distributions of tangential curvature from which primary and secondary threshold values of -0.005 were chosen.

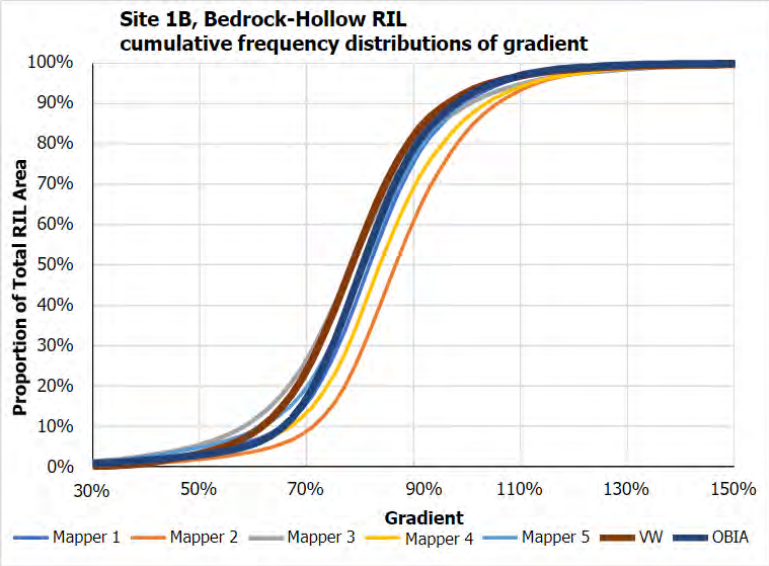


Figure G.2. Cumulative frequency distributions of gradient for mapping-team and computer-generated ~~bedrock-hollow~~bedrock hollow RILs for Site 1B.

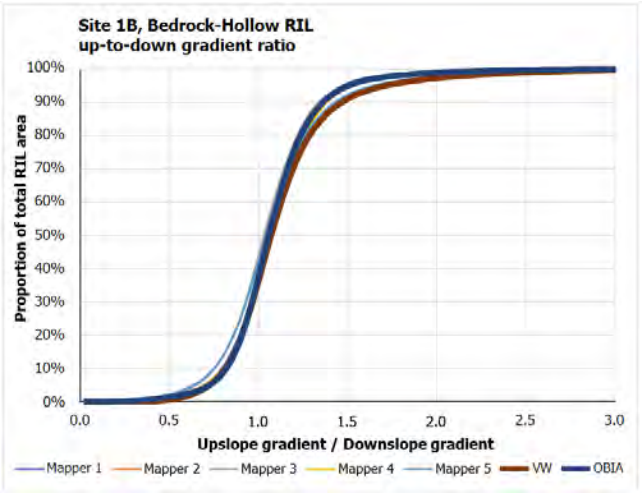


Figure G.3. Cumulative distributions for the ratio of upslope-looking to downslope-looking gradient for ~~bedrock-hollow~~bedrock hollow RILs in the Site 1B.

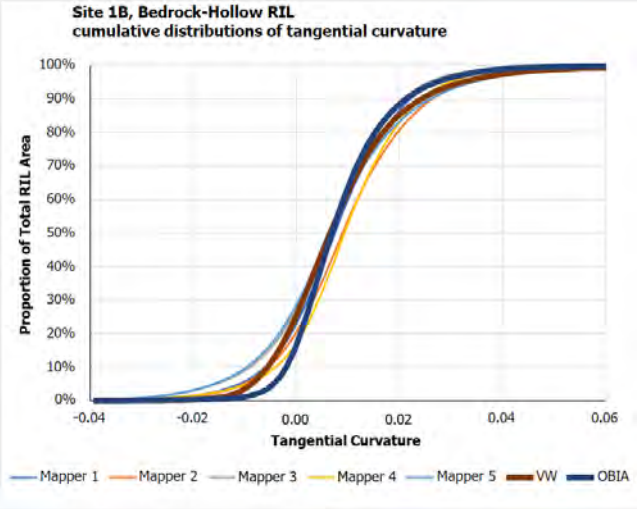


Figure G.4. Cumulative distributions of tangential curvature for ~~bedrock-hollow~~bedrock hollow RILs, Site 1B.

Figures G.5 and G.6 show cumulative frequency distributions of gradient and the up-to-downslope-looking gradient ratio for mapped and modeled inner -gorge RILs. The primary thresholds for the VW inner gorges were set to 70% for gradient and 0.95 for the gradient ratio; secondary thresholds were 65% and 1.1. Based on the cumulative frequency distributions, these thresholds were set too high. These curves provide clear guidance on how to adjust the parameter values.

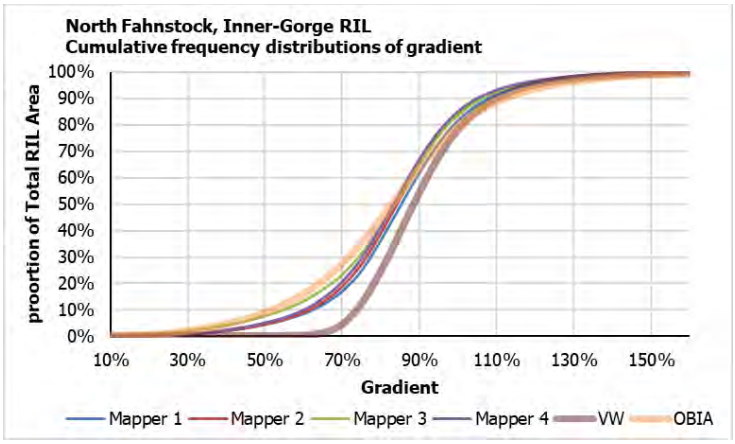


Figure G.5. Cumulative distributions of gradient for mapped and modeled inner -gorge RILs in the Site 1B.

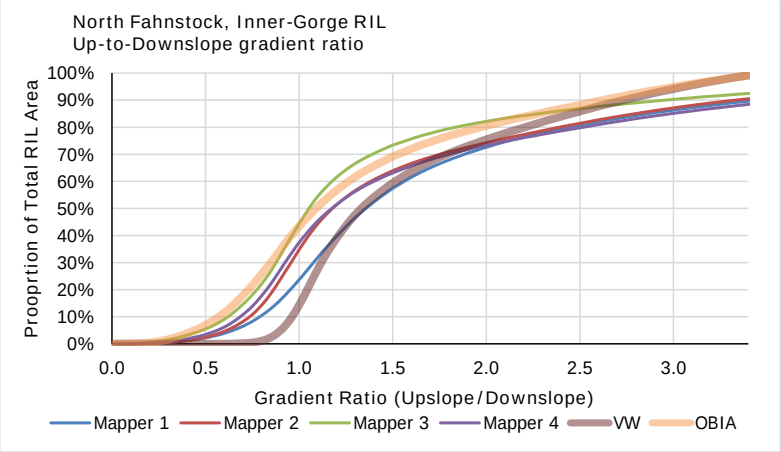


Figure G.6. Cumulative distributions of the gradient ratio for the Site 1B.

The cumulative frequency distributions of topographic attributes shown in Figures G.1 through G.7 included curves for each of the mappers and for each of the models at Site 1B. For bedrock hollows, the model curves fall within or very close to the envelope of mapper curves. For inner gorges, the VW model as currently calibrated excludes areas of lower gradient that the mappers and OBIA model included. We did not iteratively continue to adjust the model parameters; this exercise simply demonstrated that we can.