Unstable Slope Criteria Project: Object-Based Landform Mapping with High Resolution Topography

Elise Freeman (Northwest Indian Fisheries Commission)
Dan Miller (M2 Environmental Services)
Julie Dieu (Washinton Farm Forestry Association)
Jeff Keck (DNR)
Tiffany Justice (Weyerhaeuser)

Cooperative Monitoring
Evaluation & Research

CMER 2026.01.27

This Page intentionally left blank

Washington State
Cooperative Monitoring, Evaluation, and Research Committee (CMER) Document

Unstable Slope Criteria Project: Object-Based Landform Mapping with High Resolution Topography

Prepared by:
Elise Freeman (Northwest Indian Fisheries Commission)
Dan Miller (M2 Environmental Services)
Julie Dieu (Washinton Farm Forestry Association)
Jeff Keck (DNR)
Tiffany Justice (Weyerhaeuser)

Project Manager:
Theryn Henkel

Prepared for the
Upslope Processes Scientific Advisory Group (UPSAG)
of the
Cooperative Monitoring, Evaluation, and Research Committee (CMER)
of the

Washington State Forest Practices Board
Adaptive Management Program
Washington State Department of Natural Resources
Olympia, Washington

CMER 2026.01.27

Washington State Forest Practices Adaptive Management Program

The Washington Forest Practices Board (FPB) has adopted an adaptive management program in concurrence with the Forests and Fish Report (FFR) and subsequent legislation. The purpose of this program is to:

> Provide science-based recommendations and technical information to assist the board in determining if and when it is necessary or advisable to adjust rules and guidance for aquatic resources to achieve resource goals and objectives. (Forest Practices Rules, WAC 222-12-045)

To provide the science needed to support adaptive management, the FPB made the Cooperative Monitoring, Evaluation and Research Committee (CMER) a participant in the program. The FPB empowered CMER to conduct research, effectiveness monitoring, and validation monitoring in accordance with guidelines recommended in the FFR.

Document Type and Disclaimer

This Final Report contains scientific information from research or monitoring studies that are designed to evaluate the effectiveness of the forest practices rules in achieving one or more of the Forest and Fish performance goals, resource objectives, and/or performance targets. The document was prepared for the Cooperative Monitoring, Evaluation and Research Committee (CMER) and was intended to inform and support the Forest and Fish Adaptive Management program. The project is part of the Mass Wasting Effectiveness Monitoring Program and was conducted under the oversight of the Upslope Processes Scientific Advisory Group (UPSAG).

This Final Report was reviewed and approved by CMER and was reviewed and approved through the Adaptive Management Program's independent scientific peer review process (ISPR)[i]. As a CMER document, CMER is in consensus on the scientific merit, conclusions, and interpretations of the document. Recommendations contained within this document are those of the authors and may not reflect the views of any or all CMER members.

The Forest Practices Board, CMER, and all the participants in the Forest Practices Adaptive Management Program hereby expressly disclaim all warranties of accuracy or fitness for any use of this document other than for the Adaptive Management Program. Reliance on the contents of this document by any persons or entities outside of the Adaptive Management Program established by WAC 222-12-045 is solely at the risk of the user.

Proprietary Statement

This work was developed with public funding, as such it is within the public use domain. However, the concept of this work originated with the Washington State Forest Practices Adaptive

[i] a process used to "determine if the scientific studies that address program issues are scientifically sound and technically reliable; and provide advice on the scientific basis or reliability of CMER's reports" (Board Manual Section 22.4.1).

Management Program and the authors. As a public resource document, this work should be given proper attribution and be properly cited.

Full Reference

Freeman, E., D. Miller, J, Dieu, J. Keck, and T. Justice. 2026. Unstable Slope Criteria Project: Object-Based Landform Mapping with High Resolution Topography. Cooperative Monitoring Evaluation and Research. 01-27-2026. Washington Department of Natural Resources, Olympia, WA.

Author Contact Information

Elise Freeman
Northwest Indian Fisheries Commission
6730 Martin Way E.
Olympia, WA 98516
efreeman@nwifc.org

Acknowledgements

We would like to thank previous project team members Ted Turner, Susan Shaw, Jenelle Black, and Greg Stewart for their valuable contributions to this study and final report. We also thank UPSAG, CMER, and ISPR reviewers for their thorough and thoughtful comments which resulted in a much-improved report.

UNSTABLE SLOPE CRITERIA PROJECT: OBJECT-BASED LANDFORM MAPPING WITH HIGH RESOLUTION TOPOGRAPHY

Final Report

January 2026

Authors:

Elise Freeman (Northwest Indian Fisheries Commission)
Dan Miller (M2 Environmental Services)
Julie Dieu (Washington Farm Forestry Association)
Jeff Keck (DNR)
Tiffany Justice (Weyerhaeuser)

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 large woody debris) but also adversely impact ecosystem functions and increase threat to public safety. It is well recognized that forest practices can elevate landslide rates, so a key goal of the HCP is to minimize those negative effects.

To meet these goals, the WAC defines landforms that are likely more susceptible to landslides. In these areas, timber harvest and associated activities are prohibited without review by Qualified Experts and approval by Washington Department of Natural Resources. Landslide-prone areas that are potentially sensitive to forest-practice activities, known as Rule-Identified Landforms (RILs), are identified by narrative descriptions of landforms. 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.

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,*
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 Digital Elevation Models (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 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 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 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 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)).

We developed two automated approaches, both of which use Digital Terrain Models (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 two phases. Phase 1 focused on spatial comparisons of computer-based map products and remotely drawn map products 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 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 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 unsatisfactory, and efforts to increase the overlap did not produce useful results.

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 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, thirty-nine percent of bedrock hollow polygons and 62% of inner gorge polygons were designated to be correctly drawn based on lidar interpretation. Inner gorge polygons consistently received higher "correctly drawn" ratings than bedrock hollows, and VW model polygons tended to perform slightly better than the OBIA polygons on average for both reasonable screen and correctly drawn criteria.

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 delineate 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 5.4 describes the ways this project did and did not meet the objectives of 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 new, lidar change detection (LCD) derived, 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 evaluating landforms used in landslide hazard assessments and forest practice regulations.

Table of Contents

List of Acronyms

Abbreviation	Definition
BH	Bedrock Hollow
CMER	Coopertive Monotoring, Evaluation, and Research Committee
DEM	Digital Elevation Model
DTM	Digital Terrain Model
DTN	Distance-to-Neighbor
EP	Experienced Practitioners
FN	False Negative
FP	False Positive
GEOBIA	Geographic Object-Based Image Analysis
HCP	Habitat Conservation Plan
IG	Inner Gorge
ISPR	Independent Scientific Peer Review
LHZ	Landslide Hazard Zonation
MWMU	Mass Wasting Map Units
OBIA	Object-Based Image Analysis
OBM	Object-Based Mapping
RILs	Rule-Identified-Landforms
SEPA	State Environmental Policy Act
TN	True Negative
TP	True Positive
USC	Unstable Slopes Criteria
VW	Virtual Watershed
WAC	Washington Administrative Code
WAU	Watershed Administrative Unit

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) definitions (described in Table 1). The project involves a sequence of sub-projects. We address the first USC sub-project in this report, Object-Based Mapping (OBM) with High-Resolution Topography .

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. The USC Research Alternatives (USC TWIG 2017) and OBM study design (Dieu et al., 2018) documents provide detailed background information for the USC Project, 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 Report and adopted as Appendix L in the HCP, 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*." Acknowledging that actual measurement of this target is intractable, the practical interpretation on the ground is to avoid the occurrence of landslides with obvious forest practices triggers such as removing rooting strength from naturally unstable landforms and draining ditch water onto steep slopes.

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..."

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 potentially unstable slopes and landforms with associated geomorphic criteria that may require forest practices prescriptions if deemed high risk upon further assessment as 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.

RILs, as defined by Section 16 of the Forest Practices Board Manual (Board Manual henceforth), 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 Board Manual, Section 16, Part 4.1 provides detailed descriptions of these features using quantitative slope and slope-break 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 directly lend themselves to automated modeling techniques compared to 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).

Bedrock hollows are described as unchanneled, concave areas where colluvium accumulates by creep processes and where shallow groundwater is focused during storm events (Dietrich et al. 1986). They are also identified as zero-order basins which contribute to the channelized hydrologic network. They fail by landslide initiation on the order of once in thousands of years, such that some of colluvium evacuates and the feature may be channelized until refilled with colluvium. For the practical purposes of forest practices, all concave features which are of 70% gradient or greater and with a path into the channelized network are protected as the bedrock hollow landform; some of these are formed in unconsolidated materials such as glacial deposits, where bedrock hollows as technically defined (e.g., Dietrich et al. 1986) do not actually exist but where landslides in response to rooting strength loss are probable.

In the Board Manual, inner gorges are described as channelized valleys where an erosional process creates steep side walls that are subject to shallow landsliding. The 70% gradient, as with bedrock hollows, is used by forest practices, but there is recognition that process trumps threshold such that inner gorges are recognized below the 70% threshold if shallow landslide scars are observed. Demarcation of inner gorges occurs along the upper edges of the side walls where a break-in-slope shows the upper extent of failure activity. In forest practices, inner gorges occur in any lithology and include small, side hill features caused by repeated debris flow passage (i.e., those below a bedrock hollow) or by fluvial erosion, and large, valley-bottom features caused by the downcutting and undercutting of rivers.

Bedrock hollows and 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 RILs (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% gradient, they often exceed 70% and also meet the inner gorge landform 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 specified 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 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. *locations susceptible to impacts from upslope landslides such that an adverse impact to public resources or a threat to public safety is possible?*

Ultimately, the objective of the USC Project is to evaluate RIL criteria and recommend modifications to improve RIL mapping accuracy and consistency. 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 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 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

[1] "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.

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); and
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 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. As part of the first sub-project, we also describe a set of manually-mapped landforms that we produced and used to evaluate if the OBM methods could be tuned or parameterized to match observed landform locations. As is described in this report, the OMB methods need to be tunable to complete later steps of the project. Why we decided to compare OBM model results to a new, manually mapped dataset versus simply using the LHZ Mass Wasting Map Units is described in the Discussion section. In the next sections, we provide background information necessary to understand the objectives and tasks of the first project.

1.2 Background: Landform Mapping and OBIA

Landform mapping includes 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 manual and computational 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 DTM products to delineate and classify landforms based on geomorphometric variables like surface elevation, gradient, curvature, and roughness or texture (e.g., Clubb et al., 2014; Dikau, 1989; Dragut and Blaschke, 2006; Passalacqua et al., 2010; Pelletier, 2013; Wood, 2009; Jenness, 2013).

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 and Wicaksono, 2021), landslides (Van den Eeckhaut et al., 2012; Li, et al., 2015), channel networks (Sofia et al., 2011), channel knickpoints (Gailleton, et al., 2019), floodplains and terraces (Clubb et al., 2017), zero-order basins (hollows, Grieve et al., 2018), 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 and Haryono, 2013), loess plateaus (Lin et al., 2022), and volcanic flows (Feizizadeh et al., 2021).

In the Pacific Northwest, semi-automated methods (i.e., analysis of terrain elements such as slope and curvature at the spatial scale described by a DEM) with forest management applications have been developed to detect and classify areas 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. 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. More recently, researchers have used eCognition or comparable programming software to apply OBIA methods on an expanding array of land surface features.

This report evaluates two OBIA-based landform mapping approaches: the Virtual Watershed (VW) automated model for landform detection developed by Terrainworks, 2015 (described in section 3.3) and an automated model for landform detection and delineation by Shaw et al., (2017) (Weyerhaeuser model) using OBIA methods powered by eCognition software (described in section 3.2). The USC scoping document (USC TWIG 2017) and the OBM study design (Dieu et al., 2018) specified OBIA implemented using eCognition as the method to use for automated landform mapping. An OBIA-based method was chosen both because it is an established method for landform mapping, as indicated by the citations above, and because former and current Project Team members had extensive experience using OBIA for landform mapping (Shaw et al., 2017). Over the course of this study, the VW approach was added to provide an open-source alternative to compare with OBIA as implemented in eCognition. The VW approach was chosen because it also has an established track record (e.g., Barquin et al., 2015) and one of the Project Team members has been involved in the conceptual development and software implementation of the methods (Benda et al., 2016).

The 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 documented tools for geospatial, spectral, and temporal (e.g., change-detection) analyses. The eCognition users develop “rulesets” (i.e., computer scripts) that take advantage of software capabilities in analyzing data at multiple scales and hierarchical levels, processing multiple data types simultaneously and efficiently handling massive volumes of high-resolution data via a batch server application. The 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), this process is described in detail in Appendix A.

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

The Virtual Watershed takes an alternative approach, starting with identification of the source areas and transport pathways for movement of water, sediment, and organic materials through a watershed. The landscape can then be delineated into zones of differing process types (e.g., source, transport corridor, sink) and these zones can be further sub-divided by supply and transport rate. Process (e.g., shallow landslide source areas), supply rate (e.g., relative landslide density), and transport (e.g., debris-flow corridors) are inferred from terrain characteristics, including topography, substrate, landcover, and climate. Landforms can then be delineated based on inferred or postulated topographic and substrate controls on the processes and rates of interest, such as, channel networks (Mazzeo et al., 2024; Rodríguez-Castillo et al., 2018), riparian zones (Acker et al., 2023), flood plains (Benda et al., 2011), shallow landslide source areas (Miller and Burnett, 2007), and debris-flow corridors (Miller and Burnett, 2008). Landform mapping critically depends on scale for extracting and delineating landform features based on terrain elements 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 (e.g., Roering et al., 2010). 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.

1.3 Objectives and Deliverables

Section 1.2 of the 2018 Study Design for this sub-project, (OBM) with High-Resolution Topography , 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.

As described in this report, we sought to meet these objectives through development of two OBIA-based automated modeling methods that rely on geomorphic characterization of terrain

features. One method uses image analysis with classic feature extraction techniques and a second method uses the virtual watershed approach.

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.

2.0 DATA

2.1 STUDY SITES

This study focused on sites in four areas (Figure 1): the North Fork Calawah Watershed Administrative Unit (WAU) (Sites 1A and 1B), the Willapa Hills study area (Site 2) in the Stillman Creek WAU, the Howard 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.

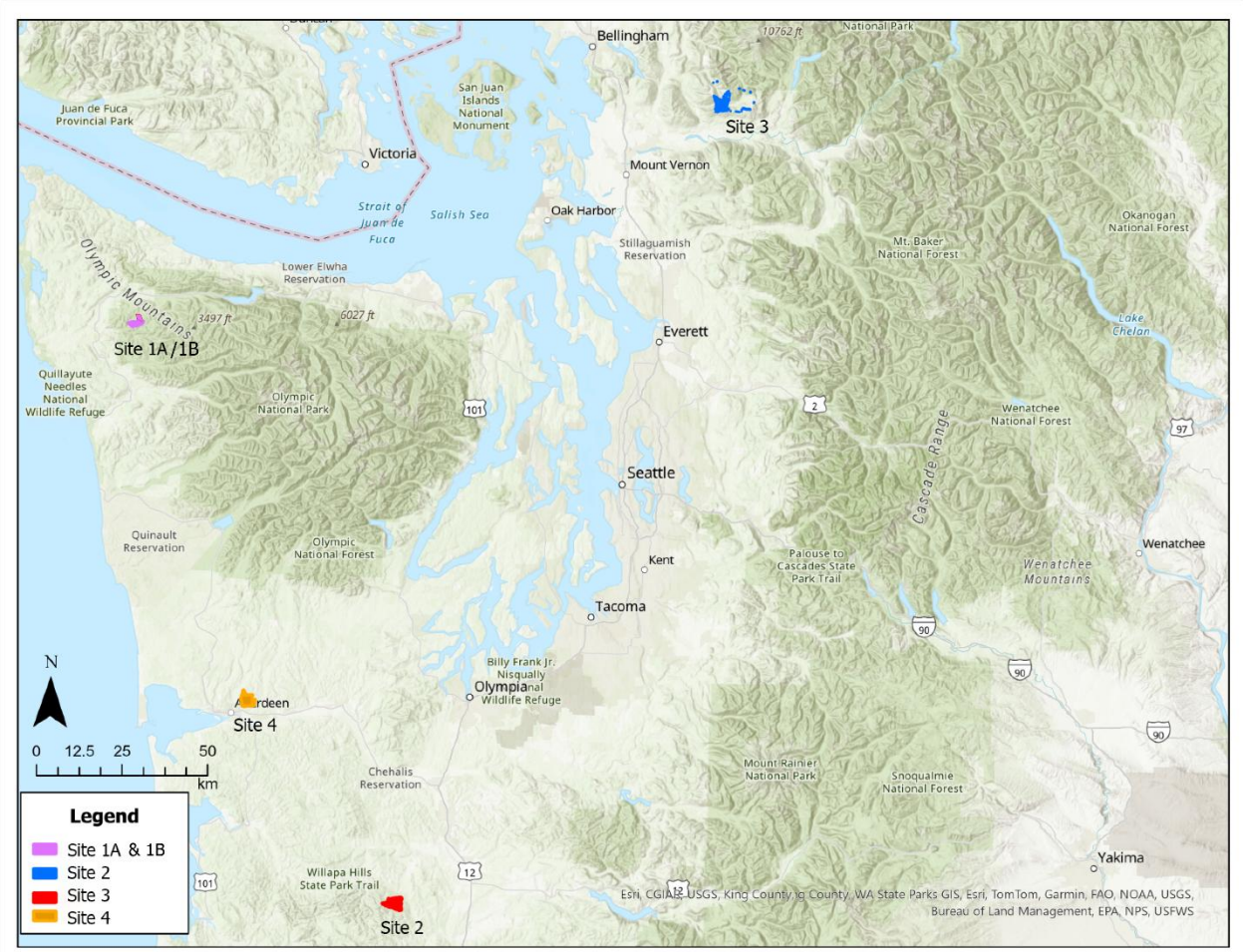


Figure 1: Study area locations in western Washington. Figures 2-5 in this section show individual sites at larger scales.

The Project Team chose the North Fork Calawah WAU (Sites 1A and 1B) 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.

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, 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 (OEm) and quaternary alluvium (Qa) 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 extends through the middle of the project area and a syncline extends along the west side, both northwest-southeast trending (Schuster, 2005). The study area is bisected by Fahnestock Creek, which flows east-to-west and 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.

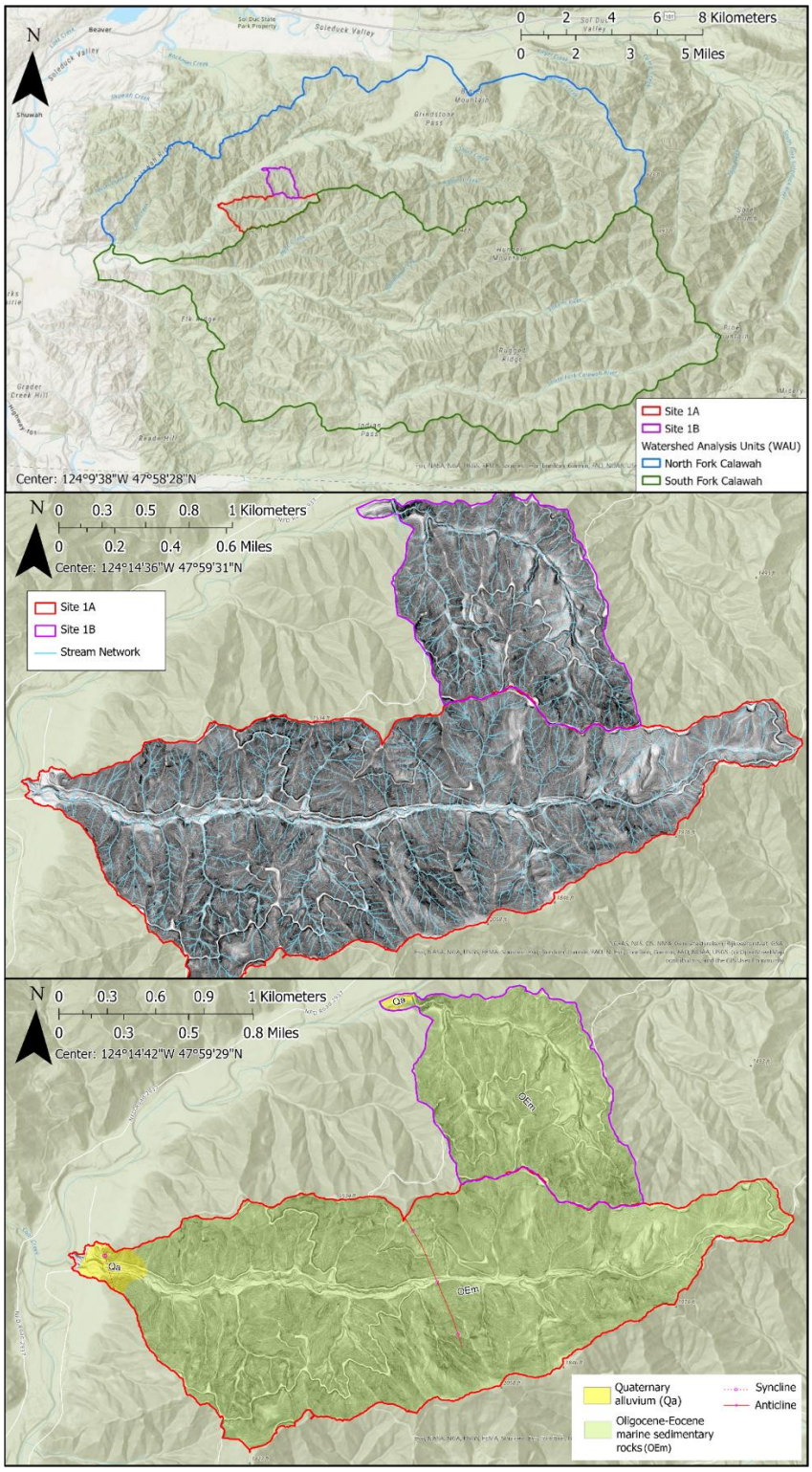


Figure 2: Location map of Calawah study area, with the North Fork Calawah and South Fork Calawah WAUs indicated in blue and green, Site 1A indicated in red and Site 1B in purple (top). Middle figure shows slopeshade and stream network over the study area. Bottom figure shows geology and structures mapped in the area.

Site 2 – Willapa Hills

The Willapa Hills 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 chose this area due to the inventory of field-based landslide locations available (Stewart et al., 2013).

The eastern Willapa Hills are comprised of numerous, broad, convex surfaces of moderate gradient. Channel denisty (i.e., stream length per unit area) 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 where the water table intersects the surface. Collectively, these broad terrain characteristics of numerous, broad convex surfaces, low channel density and deep soils/regolith suggest that the eastern Willapa Hills are tectonically stable and not experiencing uplift like Site 1.

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 consisting of fine-grained basalt-lapilli tuff, basaltic sandstone and conglomerate, interbedded with submarine basalt flows, locally pillow lava (Walsh et al., 1987). 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). Both the Crescent and Cowlitz formations have been mapped in the western portion of the site where the units are highly fractured due to faulting on plateaus and benches in the western portion of the site (Laprade, 1994). 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.

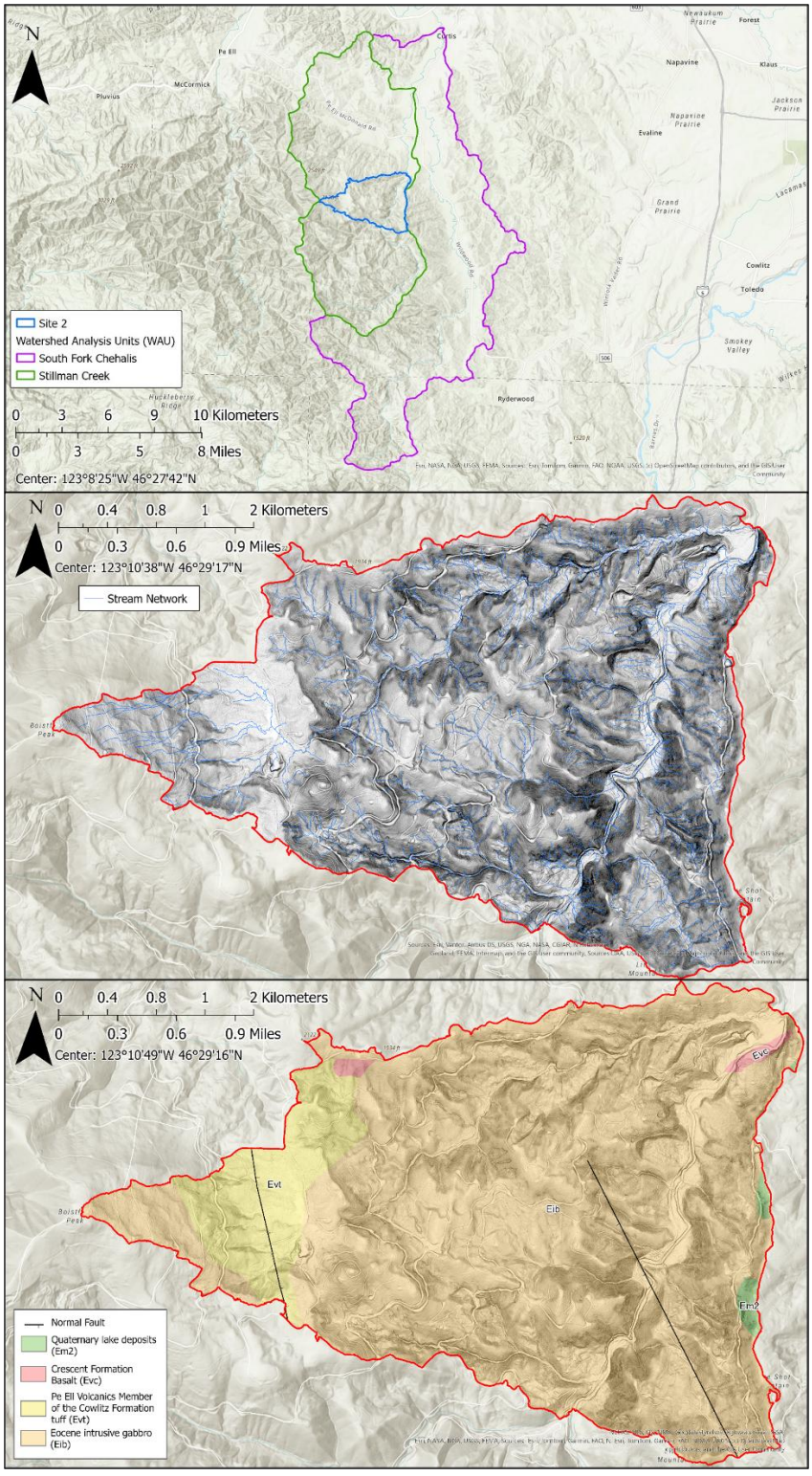


Figure 3: Location map of Willapa Hills study area, with the South Fork Chehalis and Stillman Creek WAUs indicated in purple and green and Site 2 indicated in blue (top). Middle figure shows slopeshade and stream network over the study area. Bottom figure shows geology and structures mapped in the area.

Site 3 - Howard Creek

The Howard Creek 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 (Jmv) and phyllite of Mt Josephine (Jph). This bedrock is exposed at higher elevations where hillslopes of moderate to steep gradients exist. At lower elevations, the bedrock is overlain by landslide deposits (Qls) and Pleistocene continental glacial drift (Qgd) 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 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.

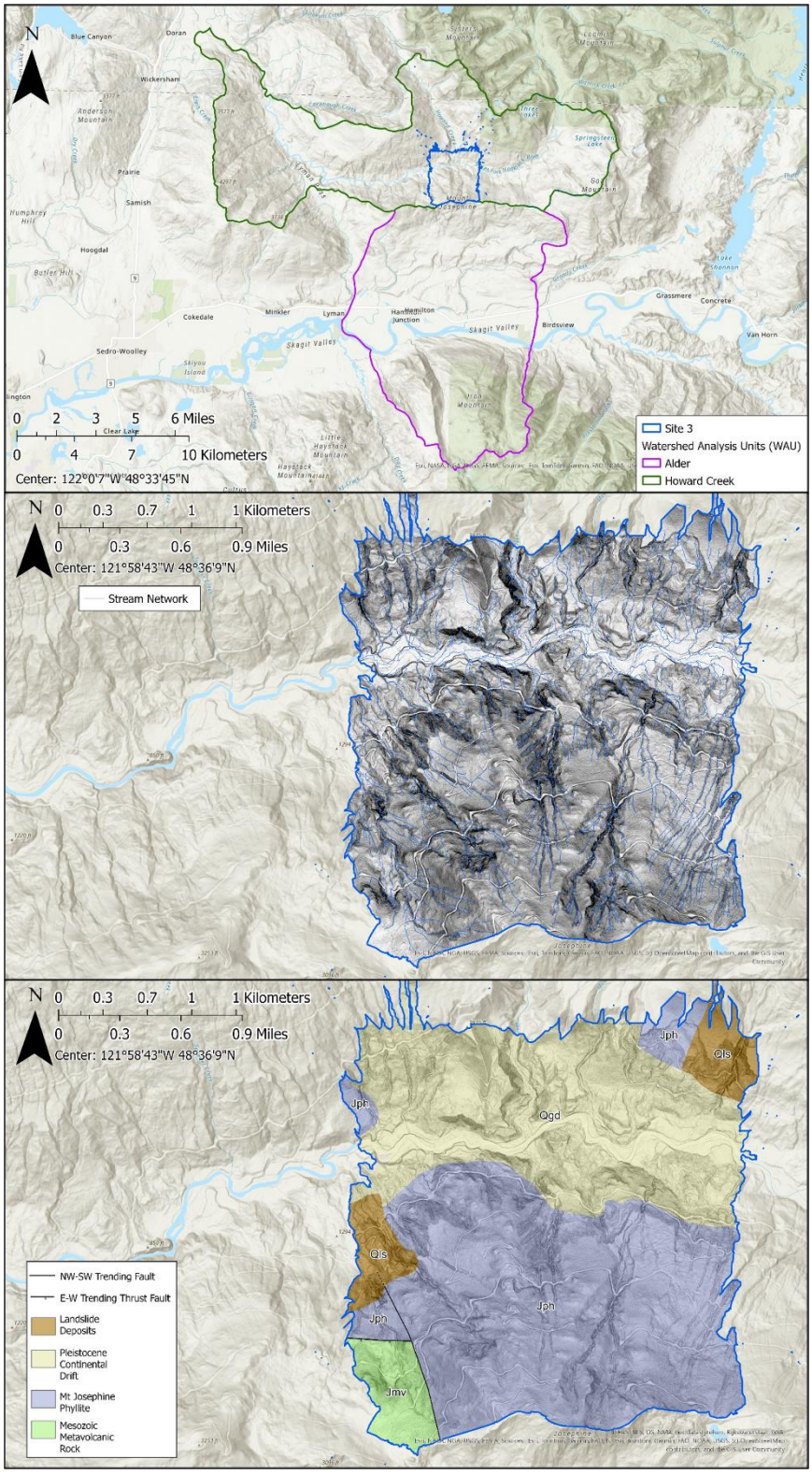


Figure 4: Location map of Howard Creek study area, with the Alder and Howard Creek WAUs indicated in purple and green and Site 3 indicated in blue (top). Middle figure shows slopeshade and stream network over the study area. Bottom figure shows geology and structures mapped in the area.

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 Quaternary old alluvium (Qoa) (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 ($Qapwo_1$) composed of deeply weathered sand and pebble gravel (Logan, 2003). Quaternary alluvium (Qa) 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 (Logan, 2003) 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 thealpine outwasht. 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.

The Tertiary marine sediments are composed of silty, friable, feldspathic Astoria formation sandstone (Mm1) and lithofeldspathic and feldspatholithic Montesano formation sandstone (Mm2s) & siltstone (Mm2t)(Rau, 1967). 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).

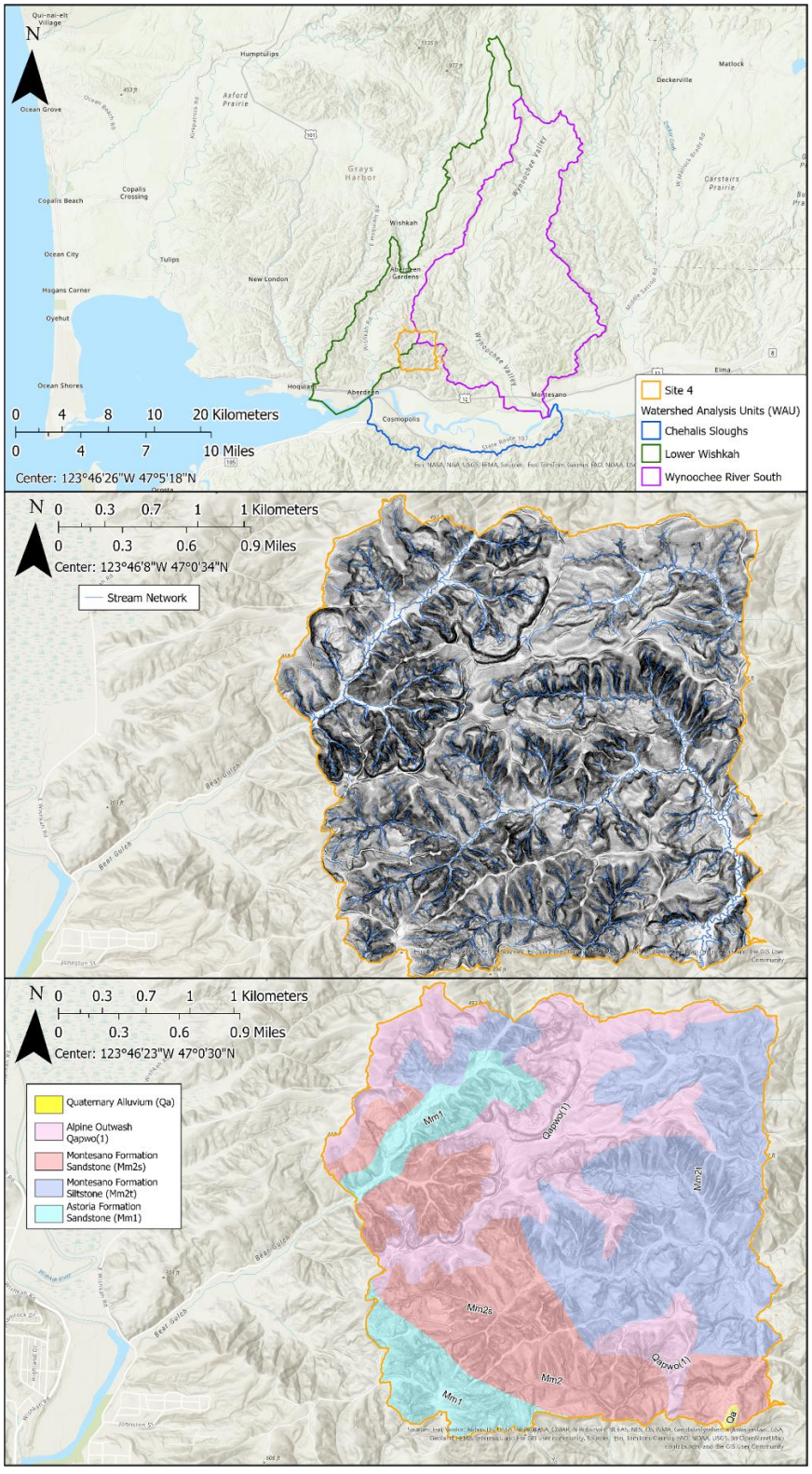


Figure 5: Location map of Wishkah area, with the Chehalis Sloughs, Lower Wishkah and Wynoochee River South WAUs indicated in blue, purple and green and Site 3 indicated in yellow (top). Bottom figure shows bare earth slopeshade layer over the study area.

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 in western hemlock forest zone, transitioning to silver fir and mountain hemlock in the higher elevations (Franklin & Dyrness, 1973).

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 topographic attributes measured from the DTM, four of which are shown by the bar charts in Figure 6. These reveal a rather striking difference between Site 1A and the other three sites: Site 1A has the largest channel density (channel length per unit basin area)[2], the steepest mean slope, and by far, the highest density of zero-order basins,[3] 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 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 inner gorge (IG) and bedrock hollow (BH) landforms between the sites. Steeper slopes together with greater channel length and number of zero-order basins per unit basin area suggest greater potential for existence of inner gorge and bedrock hollow RILs. Detailed descriptions of the topographic indices are included in Appendix E.

[2] Channel density was estimated using channel networks traced from the DTMs. Traced channel extent was dependent on exceeding thresholds for flow accumulation times gradient squared, tangential curvature, and deviation from mean elevation over a minimum flow length (30m). The thresholds varied with gradient to reflect differing channel-forming processes in steeper and flatter terrain (see Miller et al., 2015).

[3] 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.

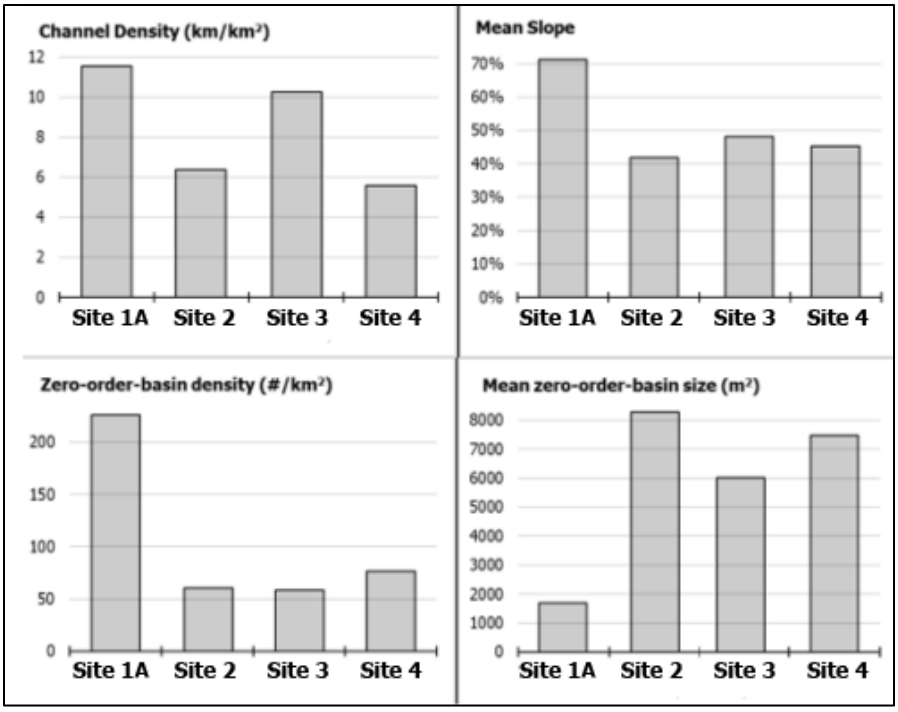


Figure 6: Topographic characteristics of the four study sites (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

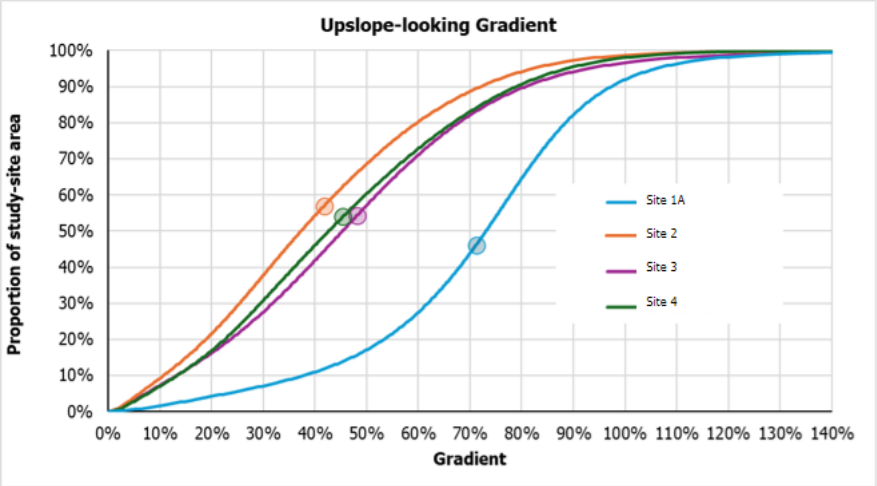


Figure 7: Frequency distributions of surface gradient for the four study sites (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah). Circles indicate the mean.

3.0 MAPPING METHODS

We used manual and automated mapping methods to produce landform maps based on RIL definitions and Board Manual guidance for inner gorge and bedrock hollow landforms. Four to five Project Team members, all experienced practitioners (EP), independently conducted the manual mapping. Two of the Project Team members developed automated mapping methods based on input provided by the entire Project Team. The EPs have a broad range of experience and backgrounds in modeling, field mapping, and remote mapping, which created sample diversity in these exercises. 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 OBIA approach described in Section 3.1.2 and Appendix A and a 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.

3.1 MANUAL MAPPING METHODS

Four mappers manually delineated potential bedrock hollow and inner gorge RILs based on WA Forest Practice definitions (WAC 222-16-050) in Sites 1A, 2, 3, and 4. 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 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 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. 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 (Weyerhaeuser 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 OBIA; 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 visually identify and delineate landforms 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 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 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 OBIA model generates an independently derived suite of landforms that subsequently can be compared geospatially with landslide distributions, mapped landforms, and other types of thematic data.

The 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, 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 were extracted first in order to isolate the RIL landforms considered for this sub-project (bedrock hollows and inner gorges).

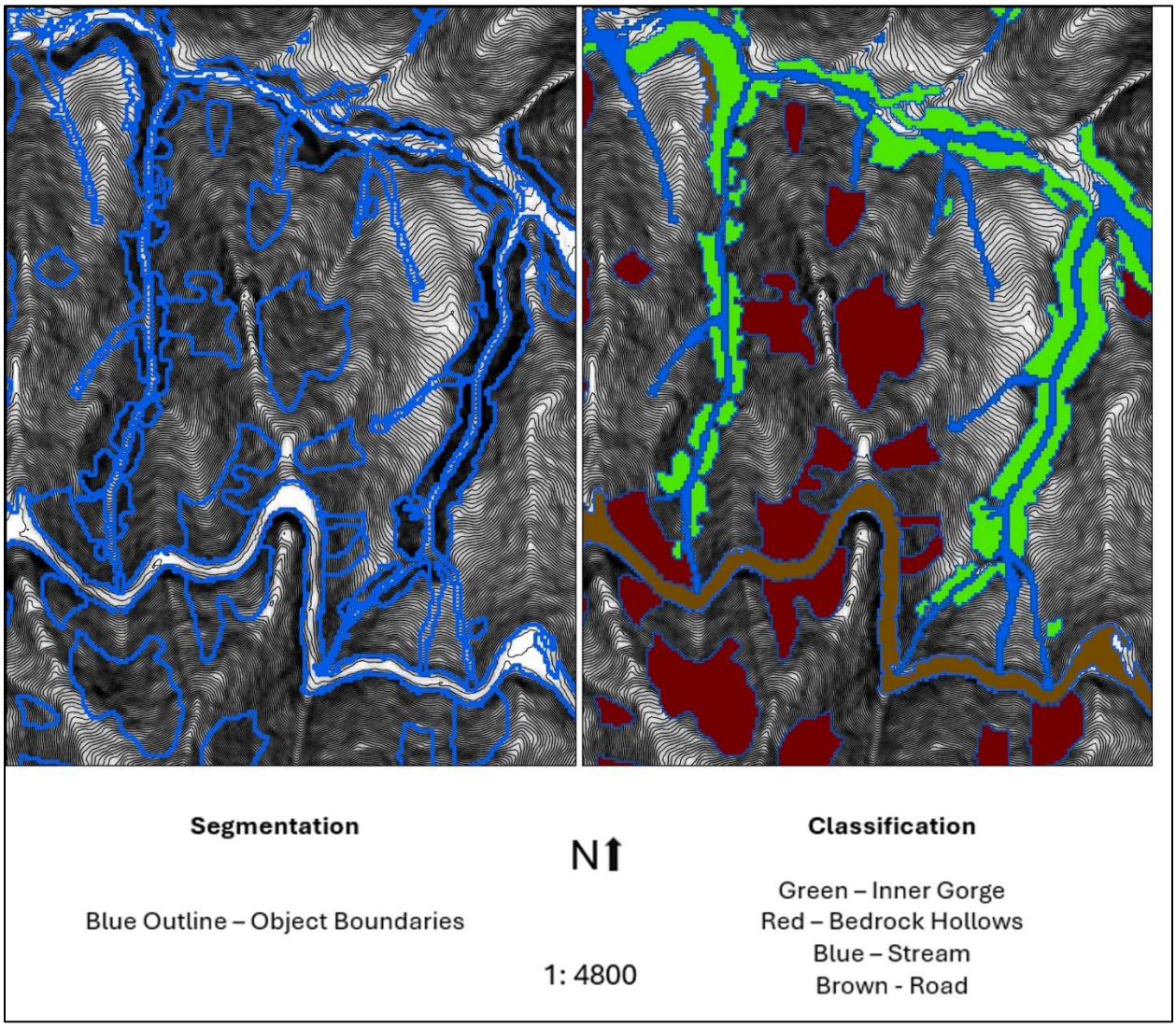


Figure 8: Fundamental image analysis processes; segmentation and classification at Site 1B. Hillshade shown as background.

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 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 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 – an expansion of a forest road to create additional level areas necessary for decking logs and loading trucks). We analyze data on hierarchical levels to evaluate membership to features extracted at different scales; for example, we can model whether 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 like 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 an extensive collection of 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 (Atwood et al., 2016; Saha & Landeghem, 2021; Witharana, 2014; Zangana et al., 2025). 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 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 in ArcGIS using NetMap tools (Benda et al., 2007)[4]. A virtual watershed enables geospatial simulation of watershed processes and human interactions (Figure 9) by enumerating watershed landforms linked explicitly to a DTM-traced channel network (Miller et al., 2015). A virtual watershed is characterized by five analytical capabilities:

1. Landform characterization: Every cell in a DTM is characterized by its geomorphometry (e.g., gradient, curvatures) and its topography (e.g., topographic position (Weiss, 2001), 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.

[4] The data structure and analysis tools for a virtual watershed are implemented in an open-source Fortran library (https://danmillerm2.github.io/NetStream_Doc/). Programs from this library are used in R scripts developed as part of a research compendium for the USC Project (https://github.com/DanMillerM2/UnstableSlopes).

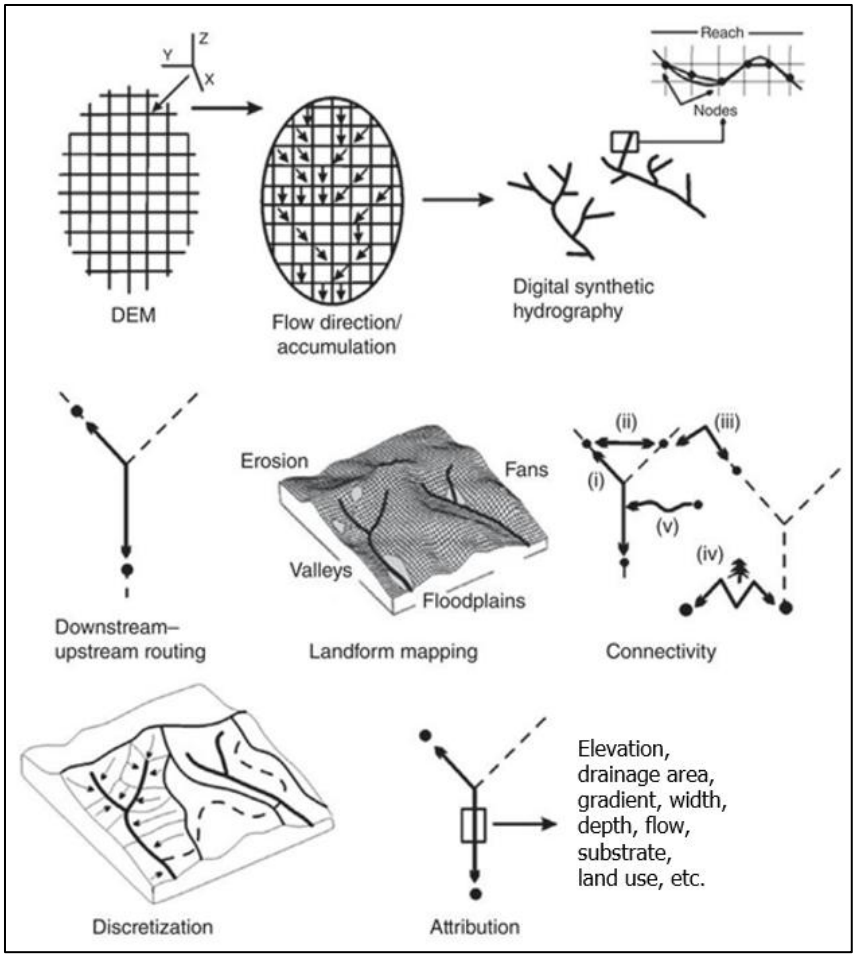


Figure 9: Steps in construction of a dataset for a virtual watershed. These data structures utilize multiple connectivity pathways (discussed in Appendix B), including: (i) river connected, (ii) Euclidean distance, (iii) slope distance, (iv) gravity-driven flow paths, and (v) modified slope distance. From Figure 3 in Barquín et al., 2015.

We used the NetStream suite of tools (https://danmillerm2.github.io/NetStream_Doc/) for creating and using a virtual watershed, with which we can delineate bedrock hollow and inner gorge landforms using the RIL definitions in WAC 220-16-050 and the descriptions in the 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.

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 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 flowing directly into unchannelized flow paths (the "drainage wings" that delineate the hillslope contributing area to a point in the traced channel network, Appendix B Figure B2), the traced network needs to extend beyond likely channel-head locations and into unchannelized swales. These unchannelized swales are referred to as zero-order basins[5].

With high-resolution lidar DTMs, channels can often be detected directly. To ensure that only a single flow path is traced upslope of the likely channel head, we used topographic indicators to differentiate erosional "process domains" (e.g., Dietrich et al., 2003; 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 et al., 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.

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). 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 can, however, result in "feathering" of the traced channel network where multiple channels are traced through a single hollow. To preclude feathering, we first identify all potential channel-initiation points for a specified threshold value,

[5] 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 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.

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 (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; Lashermes et al., 2007; Li et al., 2020; Metes et al., 2022; Passalacqua et al., 2010; Pelletier, 2013; 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. 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 adjusted their values to best match manually drawn landform polygons delineated based on the criteria derived from RIL definitions. We 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 automated methods, the ability to calibrate model outputs to better match observations will be used in the subsequent sub-projects. Here we adjusted model parameters to better match the range of topographic attributes that characterized landform polygons drawn by experienced mapping practitioners.

With the next sub-project, Susceptibility and Runout, we can adjust model parameters to better match the range of topographic attributes that characterize landslide locations.

4.0 Evaluation of Modeled Map Products

For this project, the modeled landform maps aim to reasonably represent EP would produce, but without field verification. The Project Team initially intended 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 accuracy. The Project Team evaluated the landform maps produced by automated mapping methods in two phases. In Phase 1, we compared the automated maps to landform maps manually produced by EP and to the maps of one EP 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 automated methods. The Project Team also evaluated the variability among 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.

4.1.1.1 Comparison of EP mapped landforms

All four mappers produced bedrock hollow and 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 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 polygon in the same area. Because we wanted sites that could be delineated as bedrock hollows by the mappers, we randomly selected 23 potential bedrock hollow polygons generated by the virtual watershed model and 24 potential 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 areas in Site 1A. In the exercise, each mapper voted "yes" or "no" for each of the manually drawn 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 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 RIL for each site. 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 way the selected EP's tools and methods varied from the rest of the team. Using 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 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 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 or not. The initial voting exercise was done in Site 1A, and using this site again allowed us to compare the agreement among manual mappers to the agreement between the manual mappers and the computer-model results.

4.1.1.2 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,[6] described in Table 3 below. We refer to one 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 constructed using the four values in Table 4. 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.

[6] 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 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.

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:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{Matched\ RIL + Matched\ NonRIL}{Total\ area\ (RIL + NonRIL)}$$

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)), 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 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:

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

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

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 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. 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 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 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 and inner gorge RIL maps for an additional study area, Site 1B.

In the voting exercise, mappers voted on the accuracy of each other's (and their own) bedrock hollow polygons for a randomly selected set of 47 potential bedrock hollow sites. These sites were identified by randomly selecting modeled potential bedrock hollow polygons, from each computer model, at Site 1A (Figure 10). In these polygons, there could be one bedrock hollow, many bedrock hollows, or none. 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.

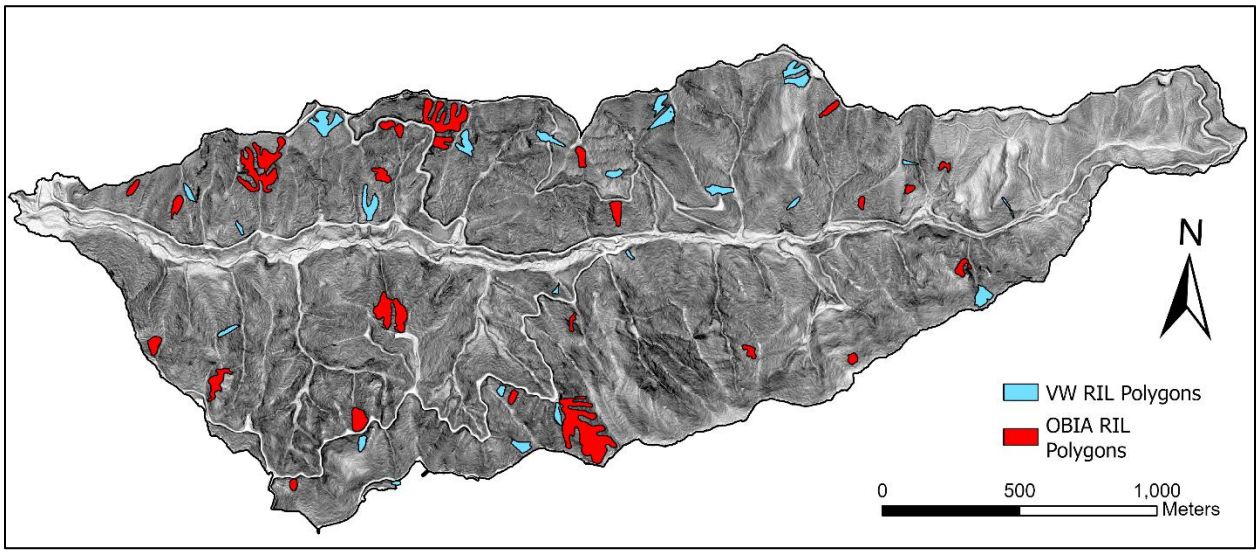


Figure 10: Location of randomly selected bedrock hollow areas mapped by OBIA (red) and VW (blue). See Figure 2.1A in Appendix A for site location (Site 1A or North Fork Calawah).

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). Landform classes were modified from this rule-set to align with the Board Manual definitions. The VW model used parameter values based on the RIL descriptions in the Board Manual. 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 heuristic approach 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 the issue of multiple bedrock hollow polygons being mapped as a single feature and to refine edge extraction methods used for inner

gorges. We modified the VW parameter threshold values for gradient, up-to-down-slope-looking gradient ratio, tangential curvature, and the proportion of a polygon with gradient less than 70% based on visual inspection of the frequency distributions of the topographic attributes associated with the mapped RIL polygons at Site 1B, shown in Appendix G. We then used these updated models to produce modeled bedrock hollow and inner gorge RIL polygons for Sites 1A, 2, 3, and 4.

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 (Appendix G), 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.

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

4.1.2 PHASE 1 RESULTS

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. Figures C1 through C8 in Appendix C show all mapped RIL polygons of both models and all mappers.

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 and 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 or North Fork Calawah

Shaded-relief images, showing the potential 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 hollows in Figure 11. Comparing the sets, the difference in size, shape and number of polygons is displayed between mappers. Figure 12 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 13 quantifies these observations and shows differences in the number, sizes, and cumulative area of 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 RIL. However, there was some disagreement as to where bedrock hollows do exist (Figures 11 through 15).

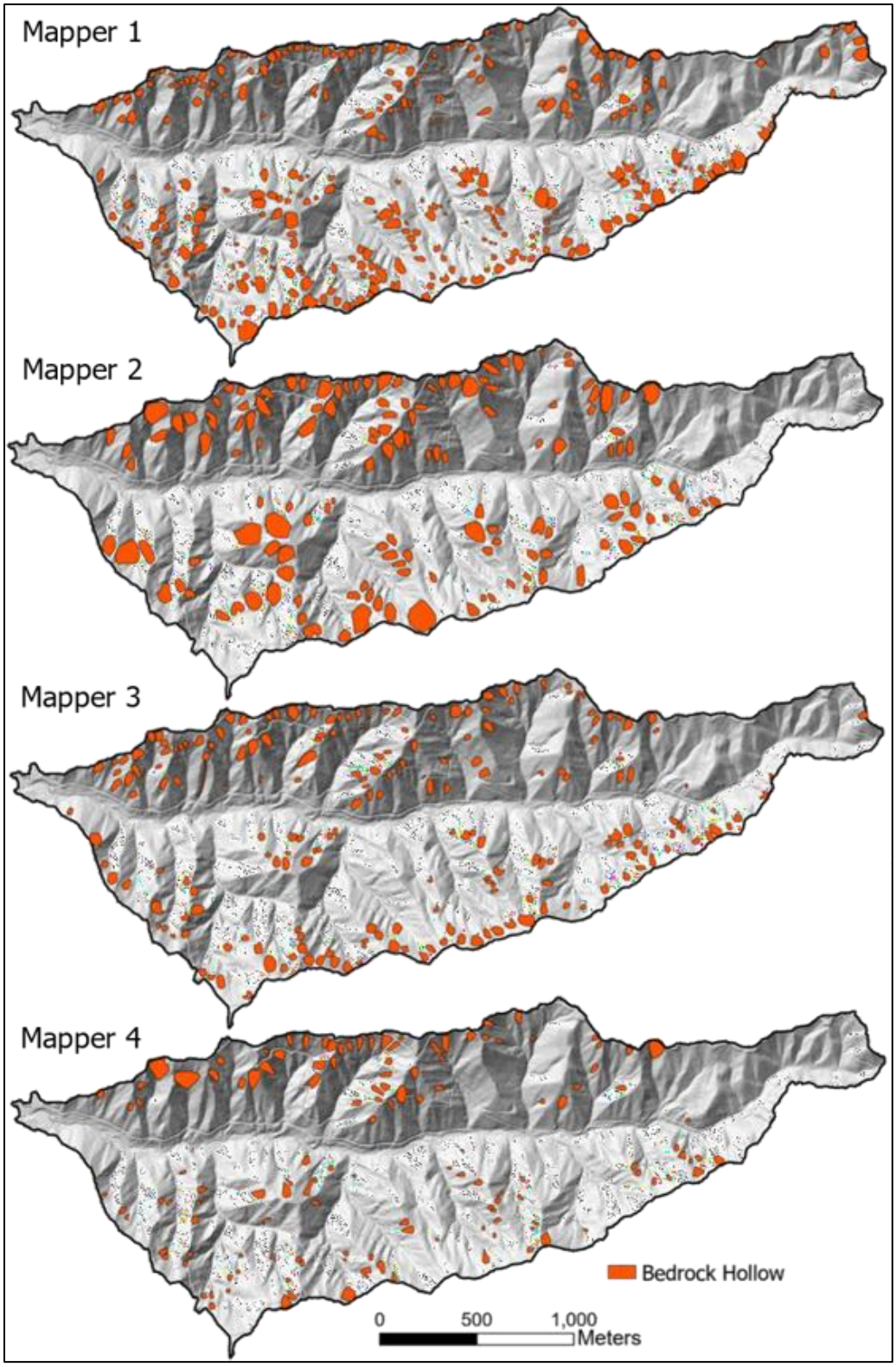


Figure 11: Each mappers' bedrock hollow RIL polygons overlain on a shaded relief image at site 1A or North Fork Calawah.

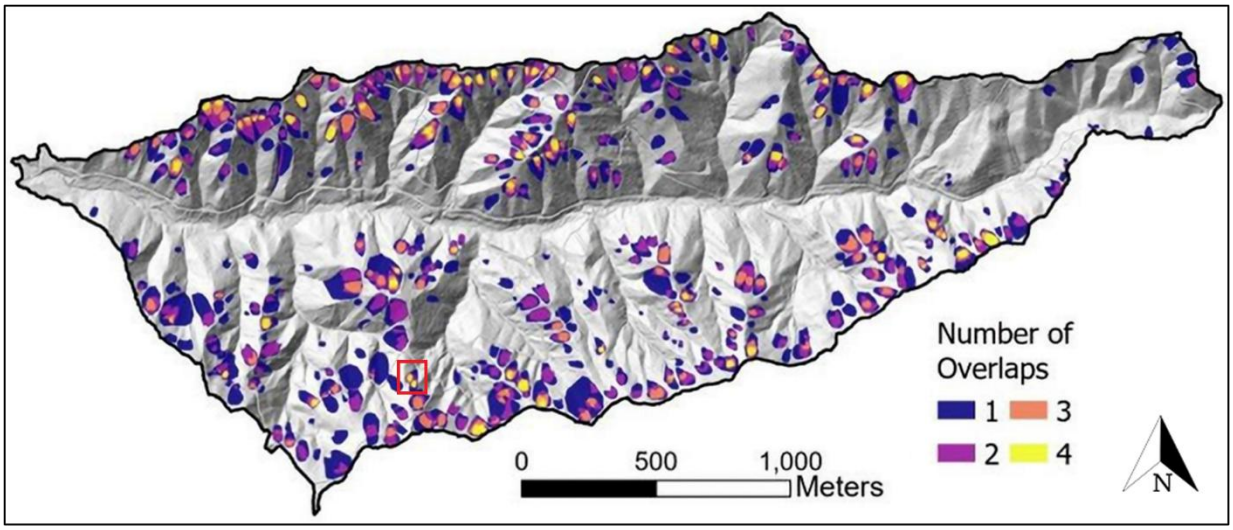


Figure 12: Number of overlaps among manual mappers at Site 1A or North Fork Calawah.

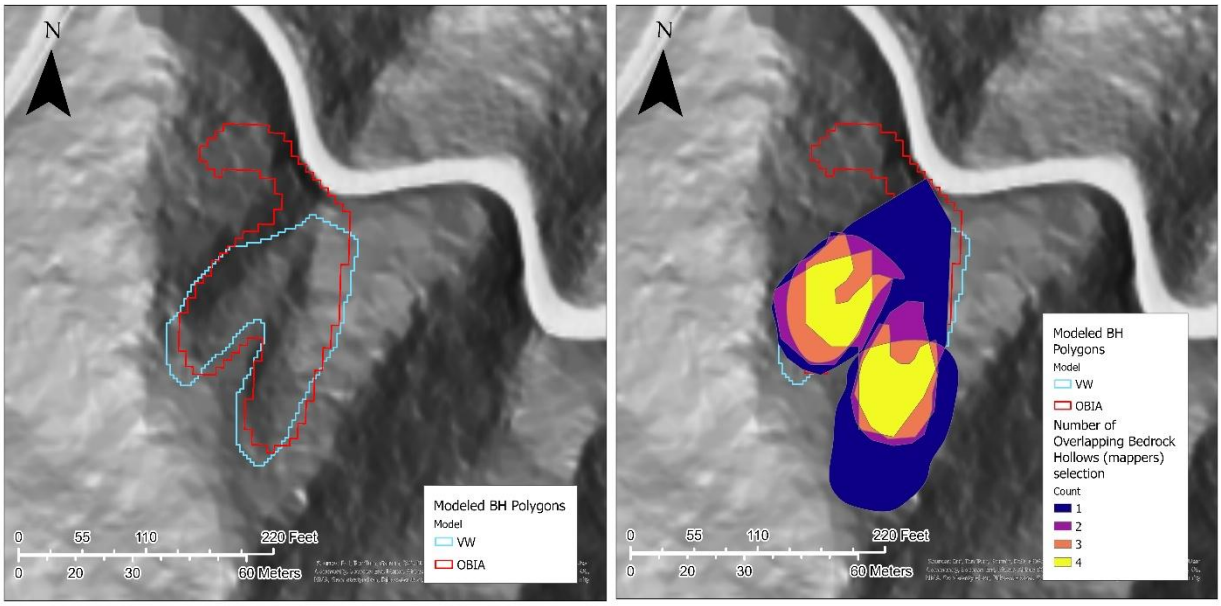


Figure 13: Left figure shows a bedrock hollow area with the polygons delineated by each of the models; the right figure shows the polygons delineated by the four mappers and the number of overlaps (right) location is indicated by the red square on Figure 12.

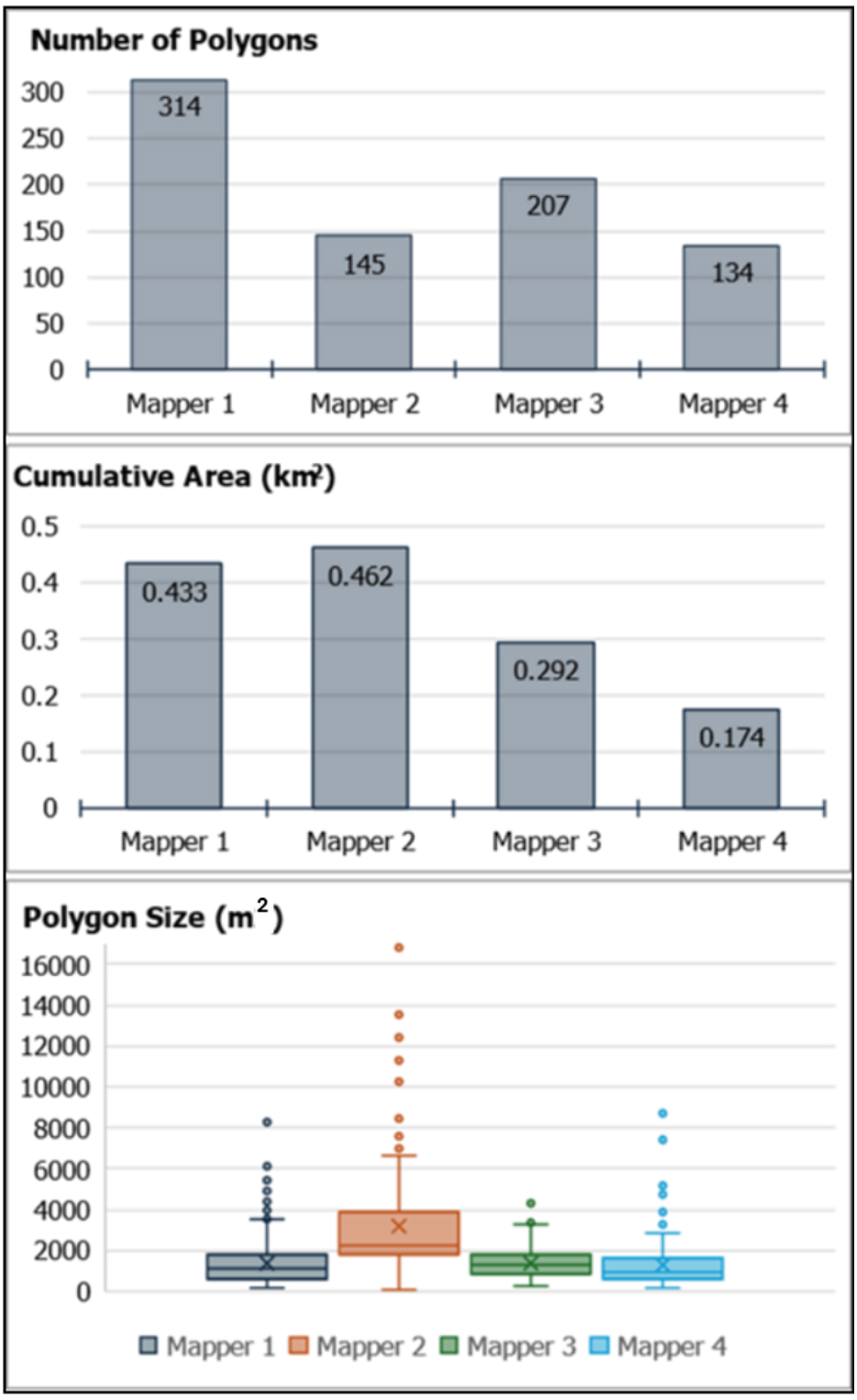


Figure 14: The number, cumulative area, and size distributions of bedrock hollow RIL polygons drawn by four experienced practitioners for Site 1A or North Fork Calawah.

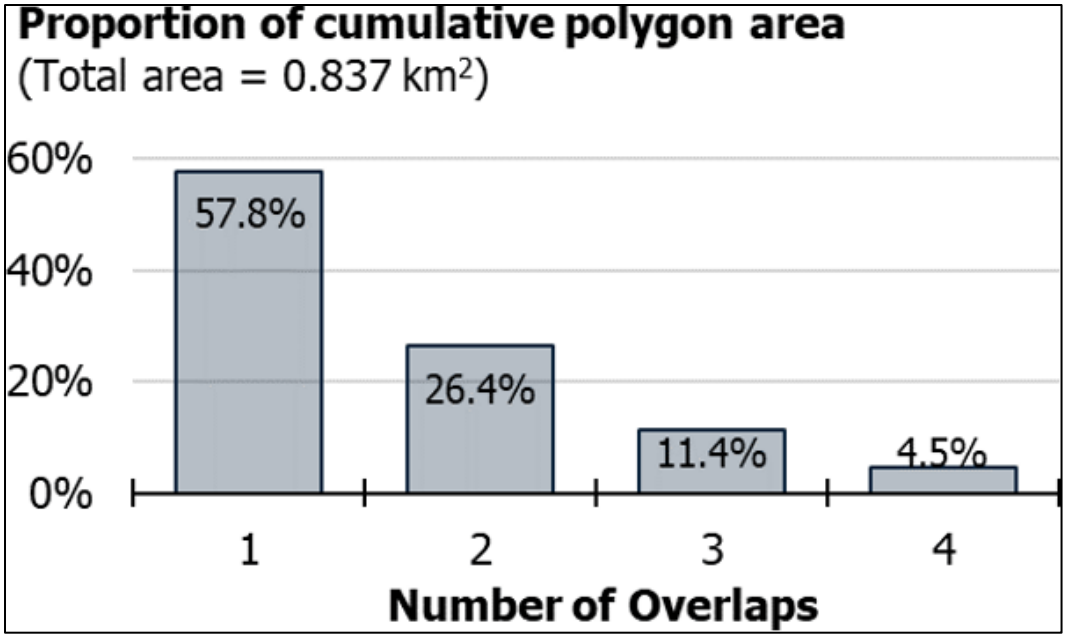


Figure 15: The agreement across the four mappers, based on the number of overlaps across all four mappers' bedrock hollow RIL polygons in Site 1A or North Fork Calawah. More than half (57.8%) of the total area encompassed by everyone's polygons included only one mapper; there was total agreement on less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

Error! Reference source not found.0Figure 16 shows shaded relief images with inner gorge polygons mapped by each mapper. Figure 17 overlays the four sets of polygons to show the number of overlaps in an area. Figure 18 shows bar charts for the length of channel with adjacent mapped 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 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.

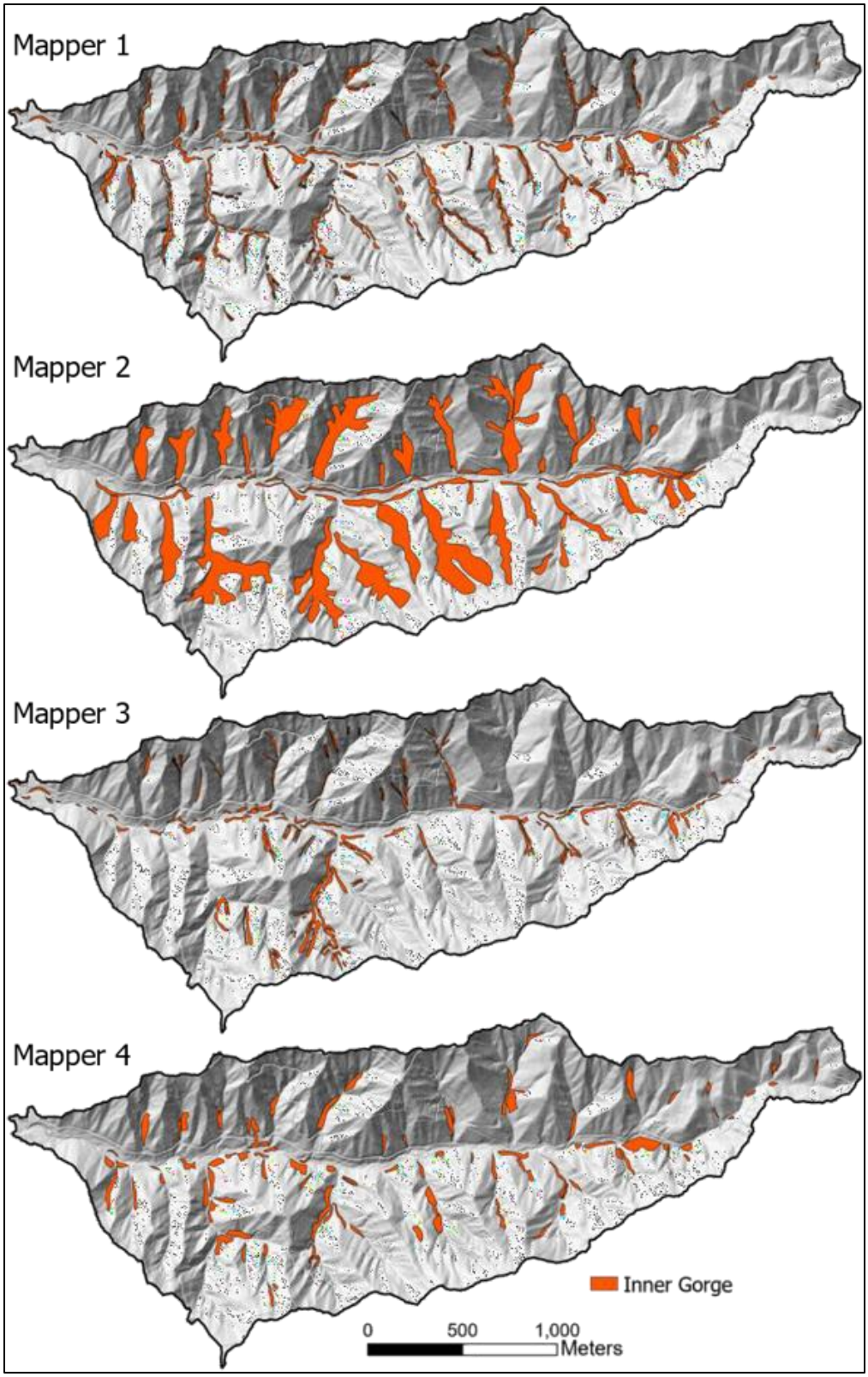


Figure 16: Inner gorge RIL polygons drawn by each of the four practitioners for the Site 1A or North Fork Calawah.

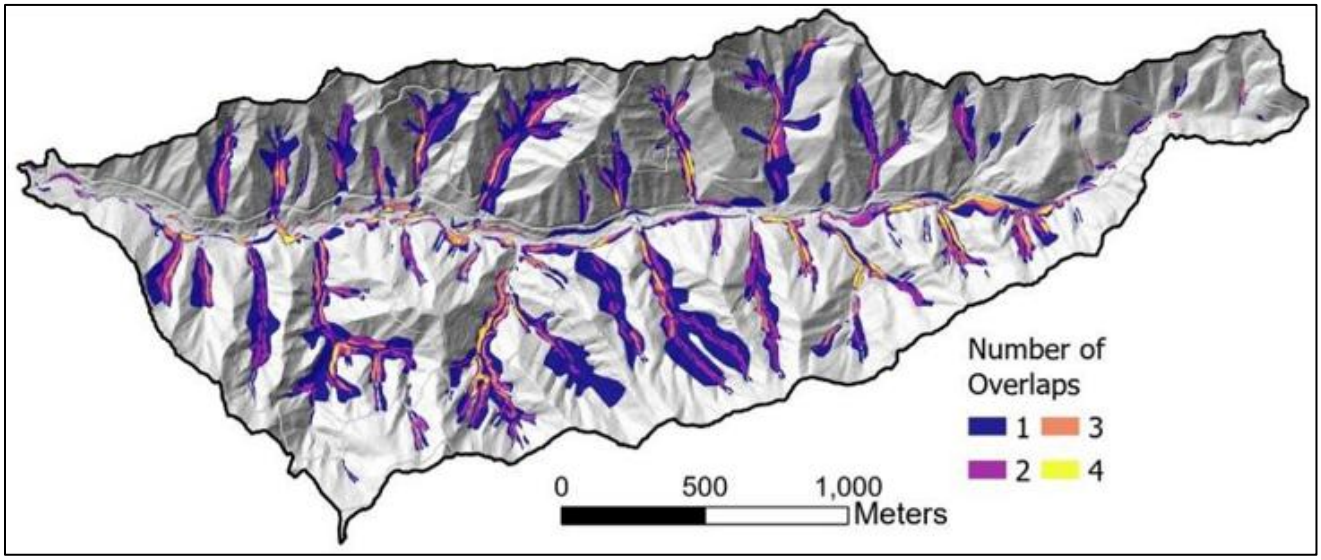


Figure 17: Inner gorge RIL polygons for all four mappers overlain at Site 1A or North Fork Calawah. Colors indicate number of mappers' polygons that overlap in any location.

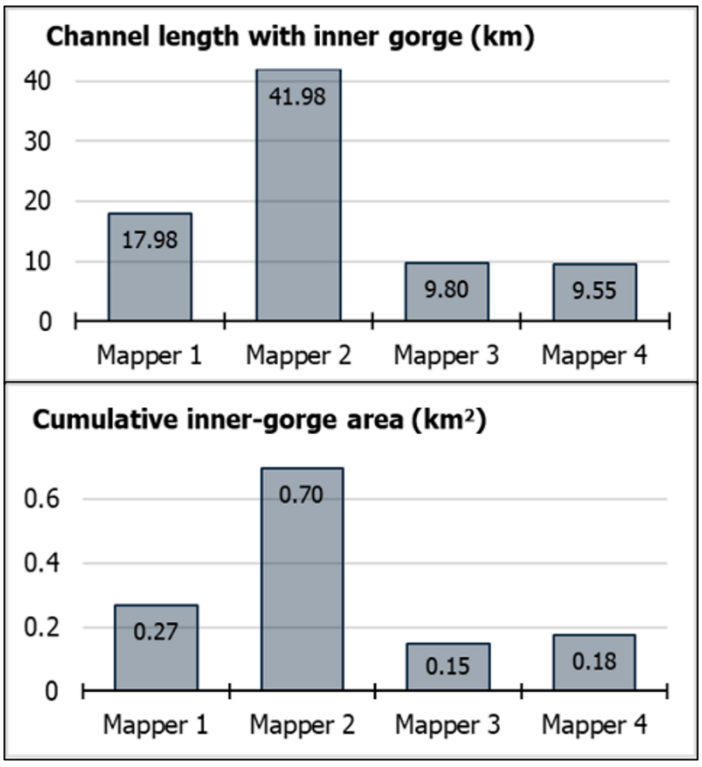


Figure 18: Cumulative channel length with mapped inner gorge and cumulative mapped inner gorge area for each mapper in Site 1A or North Fork Calawah. Channel length was measured for each side of a channel, so where inner gorges were mapped on both sides, that channel length counted twice.

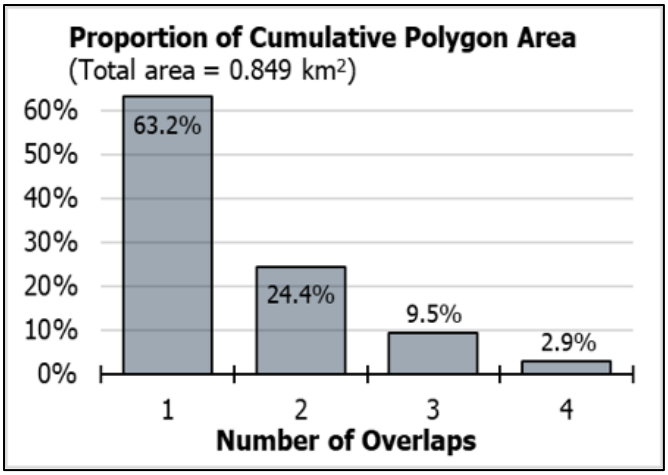


Figure 19: Consistency across all mappers, based on the number of overlaps across all mapper's inner gorge RIL polygons in Site 1A or North Fork Calawah.

4.1.2.2 Voting: Agreement among Mappers

The number of manually mapped polygon overlaps varied site by site. Five of the modeled bedrock hollow polygon sites had no coincident polygons drawn by any of the mappers. Some of the selected sites contained multiple bedrock hollow landforms delineated by the mappers. 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 5.

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 mappers 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 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 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, which shows how the proportion of each mapper's matching polygons (which varied among mappers, see Table 5) receiving "best polygon" votes varied depending on who was voting.

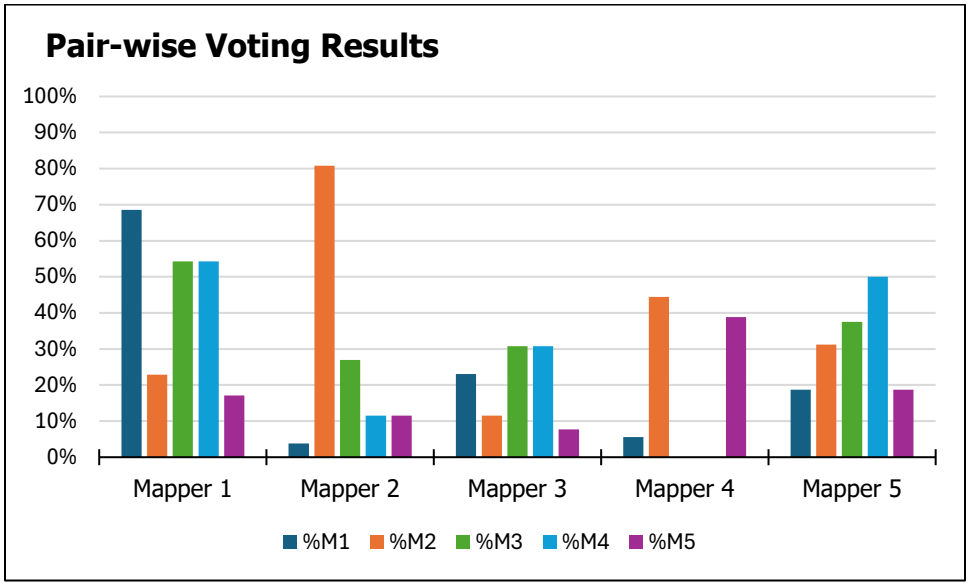


Figure 20: Pair-wise voting results for which mapper had the most accurate bedrock hollow RIL polygons. The x axis shows the mapper voting and the y axis shows what proportion of each mappers' matching polygons they gave a "best" vote. Ex. 23% of the 35 matching bedrock hollow polygons drawn by Mapper 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 differences in the mapping scales used and channel-head placements 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. 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 or Fahnestock Creek after mapper alignment

We repeated the manual mapping exercise, after identifying inconsistencies in mapping protocol, to see if mapping at the same scale using the same GIS data files and the same estimates of channel-head location resulted in greater consistency between mappers. To avoid bias, the

Project Team selected a new study site (Site 1B) to repeat the exercise in. Site 1B is the small basin just north of Fahnestock Creek Basin (Site 1A) shown in Figure 2 in Section 2.1. 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.

The fifth mapper who completed part of the exercise for Site 1A mapped bedrock hollows at this site; they 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 RIL polygons tended to be larger than everyone else's, reflecting what was deleted from a harvest proposal (highlighting the differences between remote mapping and actual field verification).

Figure 21 shows each mapper's bedrock hollow and inner gorge RIL polygons overlain on a shaded relief image, like those shown in Figure 10. Figure 22 shows the number of polygon overlaps, like those shown in Figure 11.

Figure 23 shows balanced accuracy, recall, and precision for a one-to-one comparison of each mapper's bedrock hollow RILs for both Site 1A and Site 1B. As described in Section 4.1.1.2, higher values indicate a greater degree of overlap. We expected generally higher values and more consistently equal values among mappers for Site 1B. The results of averaging these measures over all values for combinations of Mappers 1 through 4 (i.e., the mean of all pair-wise comparisons) at each site are shown in Table 6.

Table 6. Averaged balanced accuracy and recall and precision for mapped bedrock hollows at Site 1A or North Fork Calawah and Site 1B or Fahnestock Creek.

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

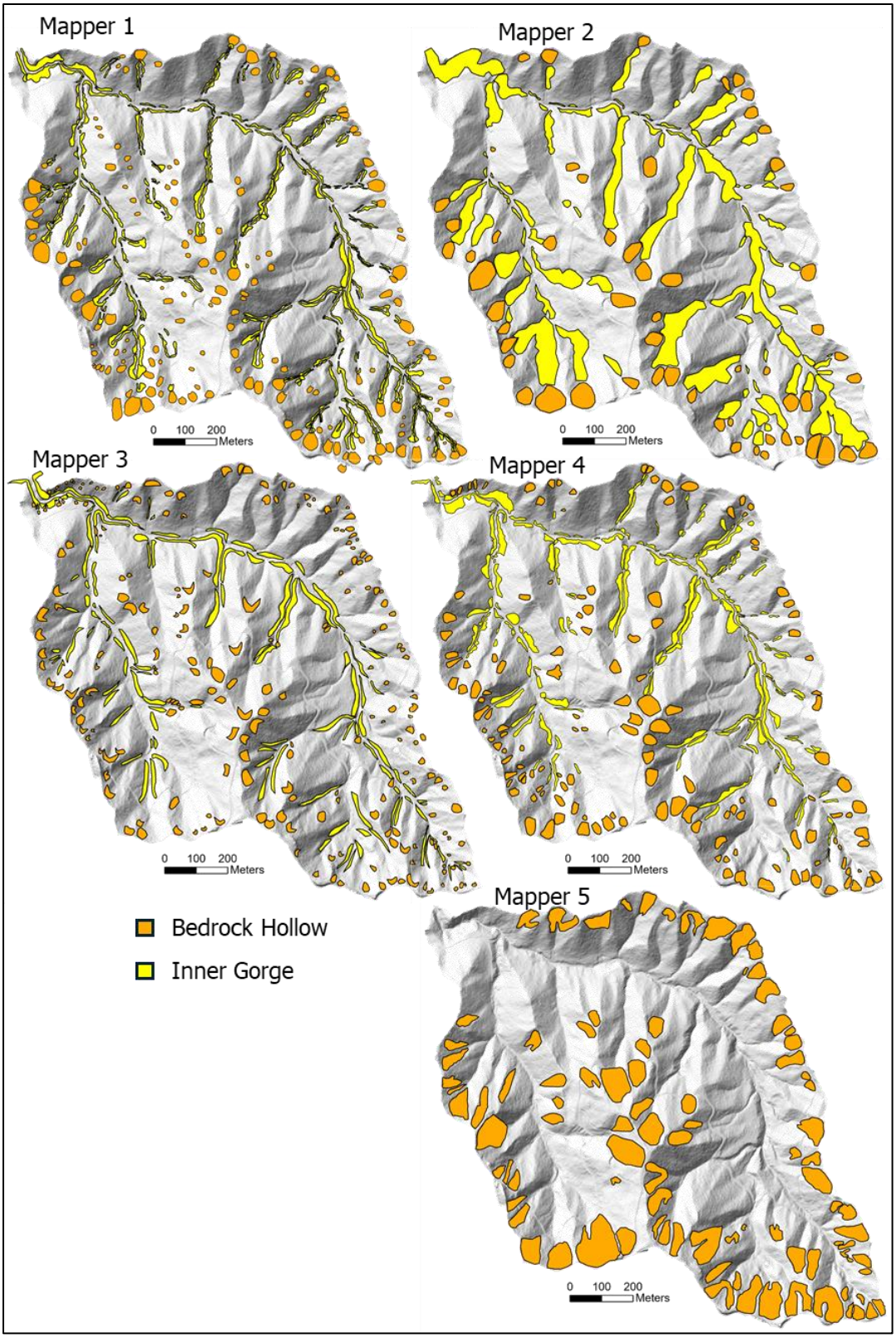


Figure 21: Each mappers' bedrock hollow and inner gorge RIL polygons for Site 1B or Fahnestock Creek. Mapper 5 did not map inner gorges.

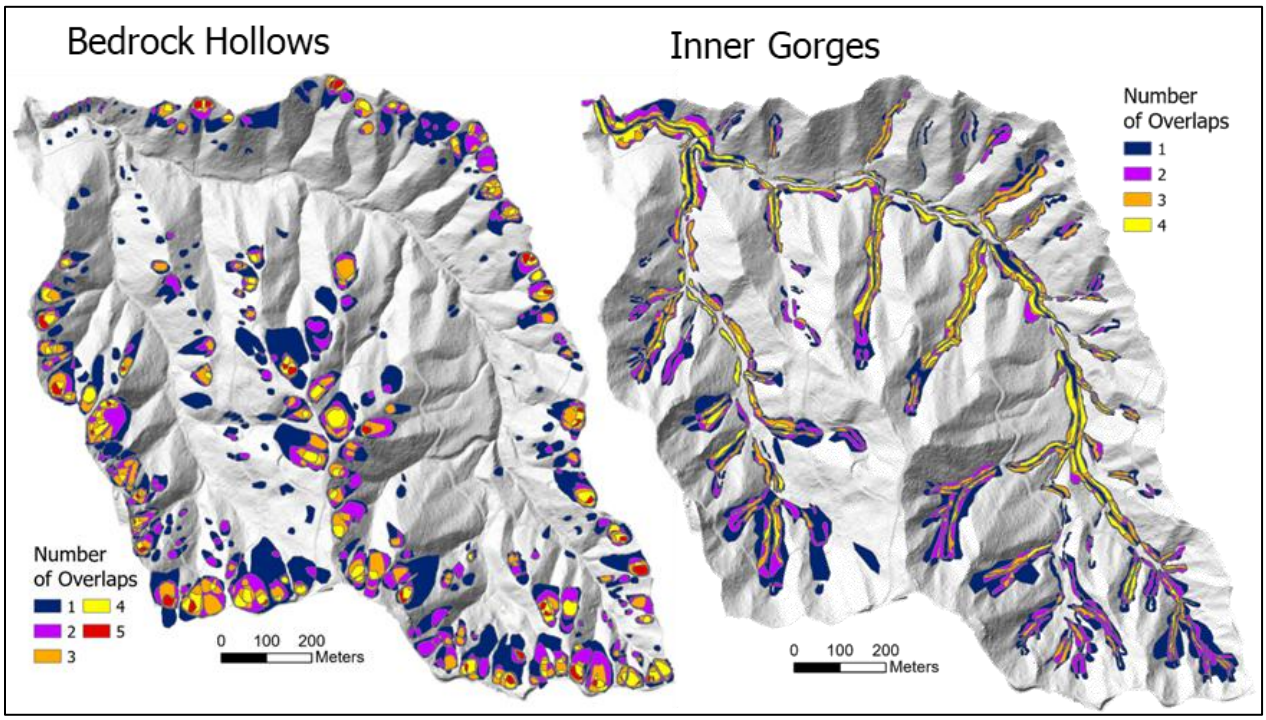


Figure 22: Overlap count of manually mapped polygons in Site 1B or Fahnestock Creek. Each color represents the number of polygons that overlap in that area.

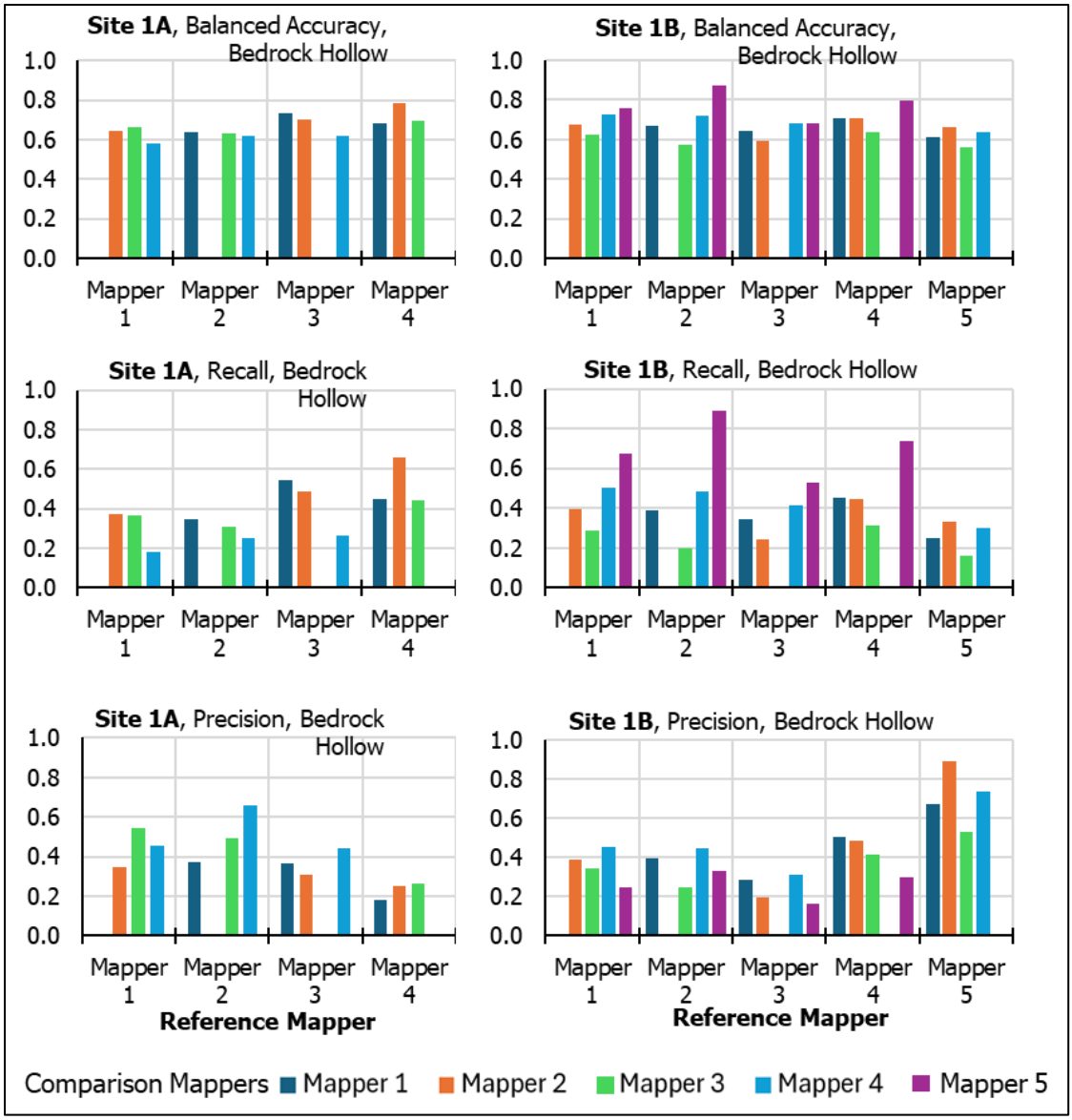


Figure 23: Balanced accuracy, recall, and precision for manually mapped bedrock hollows at Site 1A or North Fork Calawah and 1B or Fahnestock Creek.

Bar charts of balanced accuracy, recall, and precision for mapped inner gorge RILs at Sites 1A and 1B are shown in Figure 24. The averages over all combinations of mappers are shown in Table 7.

Table 7. Averaged balanced accuracy and recall and precision for mapped inner gorges at Sites 1A or North Fork Calawah and 1B or Fahnestock Creek.

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.

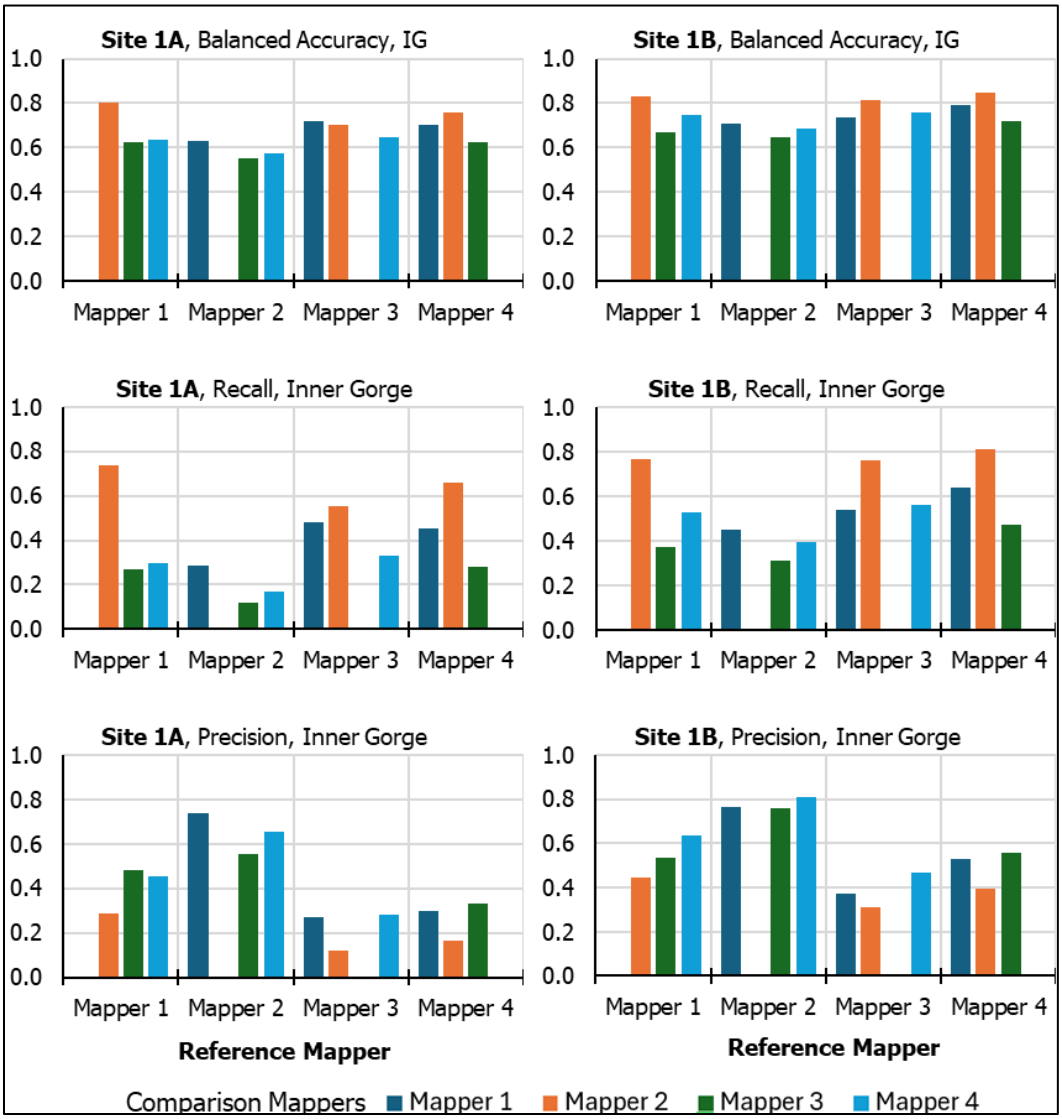


Figure 24: Balanced accuracy, recall, and precision for manually mapped inner gorges at Sites 1A or North Fork Calawah and 1B or Fahnestock Creek.

Sites 1A and 1B are adjacent to each other and have essentially the same terrain features. The little improvement in consistency after revising our mapping protocol 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.

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. We also evaluated agreement among mappers by seeing how often mappers both did or did not delineate a bedrock hollow polygon in that site. Figure 25 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 landforms at 47% of those sites.

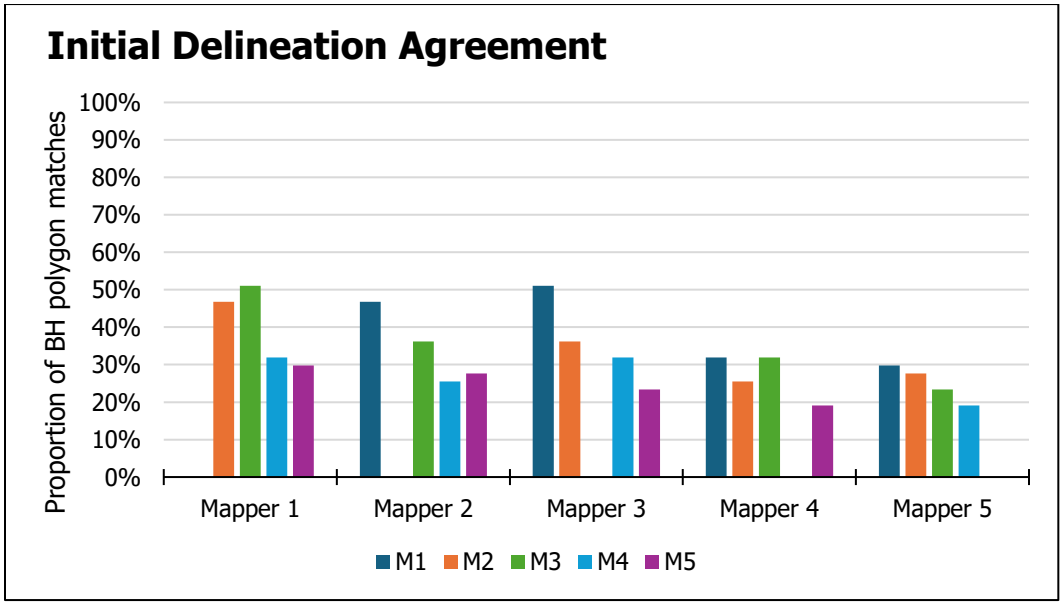


Figure 25: Proportion of agreement between mappers' initial bedrock hollow delineations at Site 1A. The height of the bars shows the percentage of bedrock hollow polygons drawn by each of the other mappers that matched the polygons drawn by the mapper on the x axis.

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 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 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). 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 8, 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.

We acknowledge that mapping ALL the RILs for a large area requires substantial effort, 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 initially drew a polygon	35	26	26	18	16
# of sites in which the mapper did not initially draw 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 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 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 or North Fork Calawah.

	# 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%

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 5).
2. Mappers generally chose their own polygon as "best," but not always (Figure 20).
3. Only 4 out of the 42 sites with matching mapper polygons got unanimous votes as to which mapper's polygons best represented a typical EP delineation. There was disagreement for the remaining 38.
4. In this project, 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).

From these we draw three conclusions:

1. The mappers in this project had inconsistent opinions regarding what a bedrock hollow RIL looks like using only remote data, as indicated by the variable bar heights in Figure 25.
2. Uncertain initial delineation of boundaries led 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 like 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 26 through Figure 30 and as cumulative distribution function (CDF) plots in Appendices F and G.

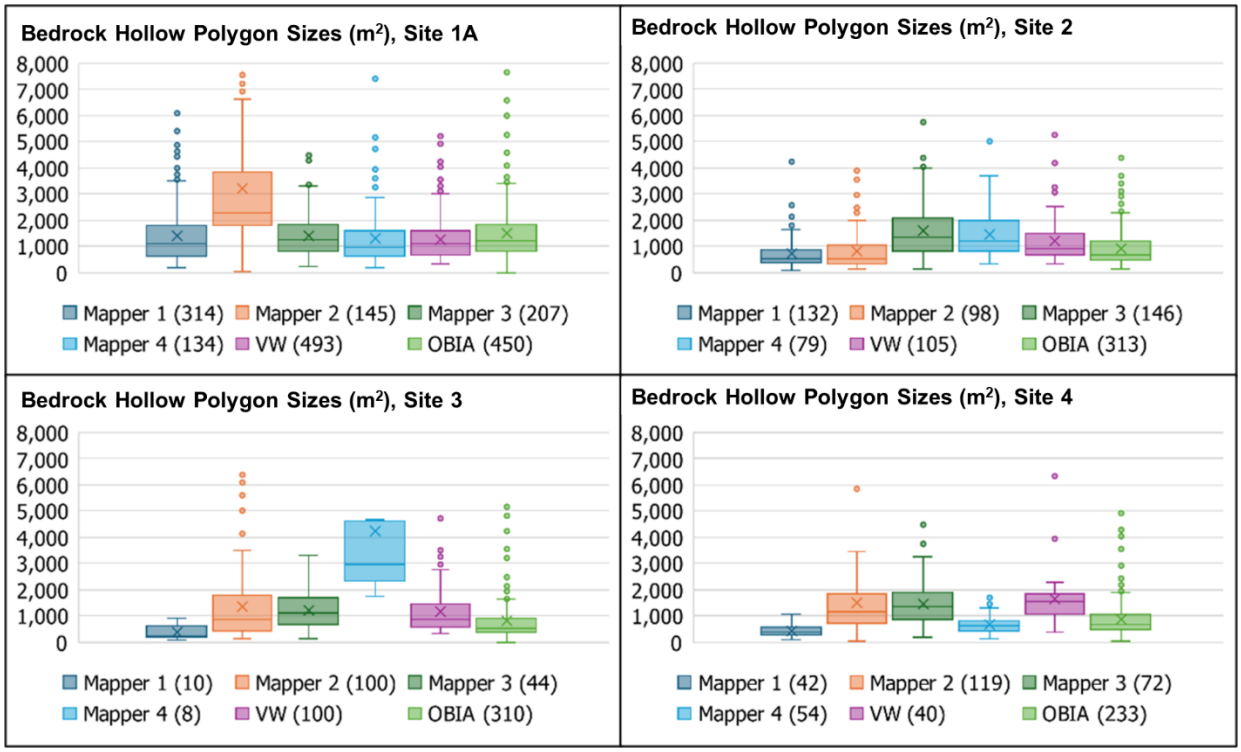


Figure 26: Box and whisker plots of bedrock hollow RIL polygon size in square meters. The boxes include 50% of all polygon values, extending from the 25th to the 75th percentiles. The horizontal line inside the box indicates the median; the X indicates the mean. The size of the box indicates the "inter-quartile 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 (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

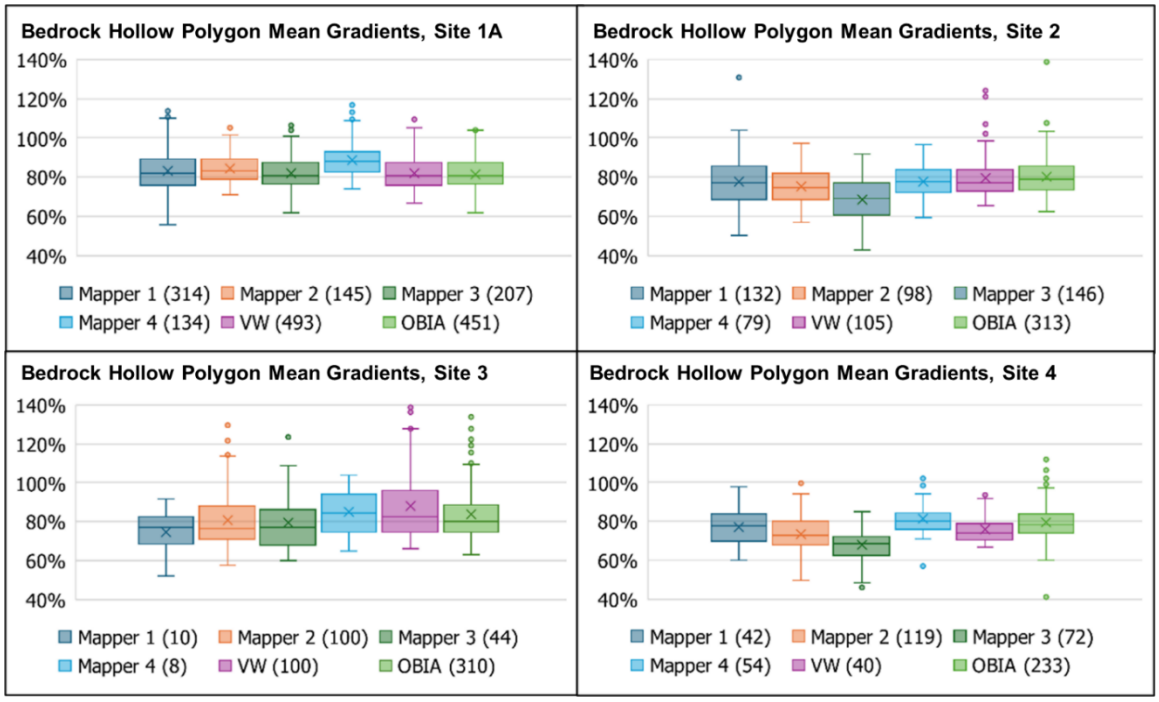


Figure 27: Box-and-whisker plots of mean gradient for bedrock hollow RIL polygons for each of the mappers and the two computer models for each of the four study sites (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

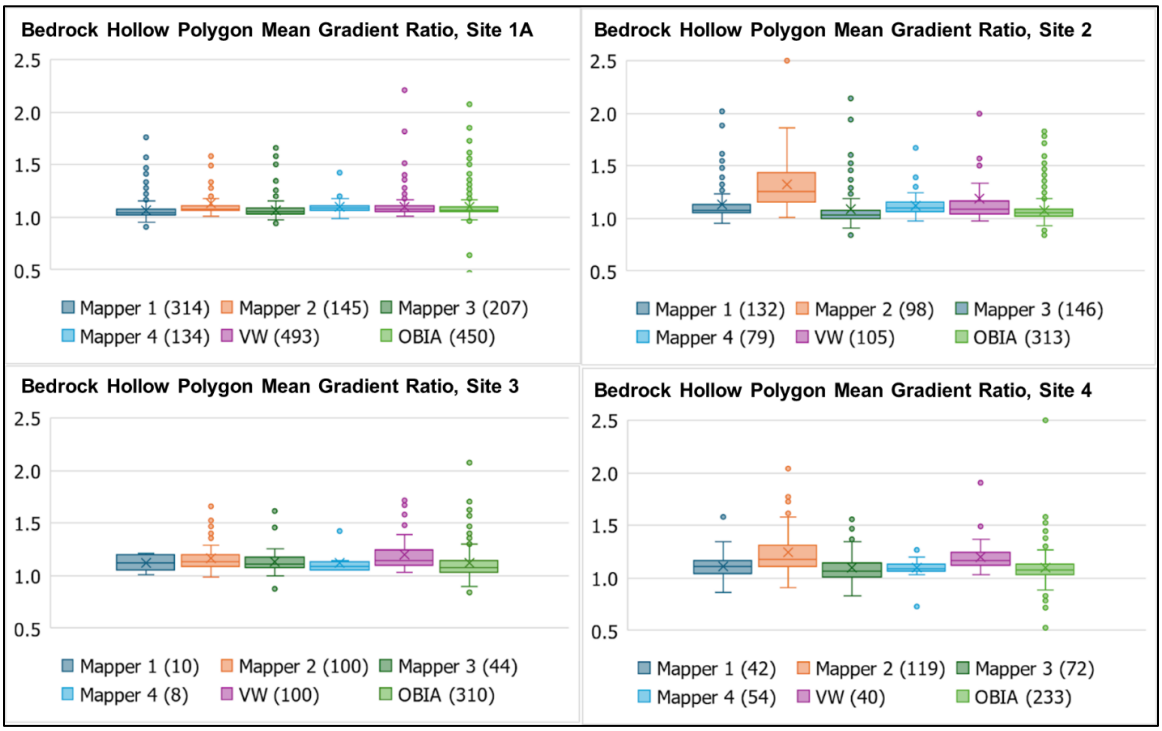


Figure 28: Box and whisker plots of the mean gradient ratio (upslope-looking / downslope-looking) for bedrock hollow RIL polygons (Site 1A = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

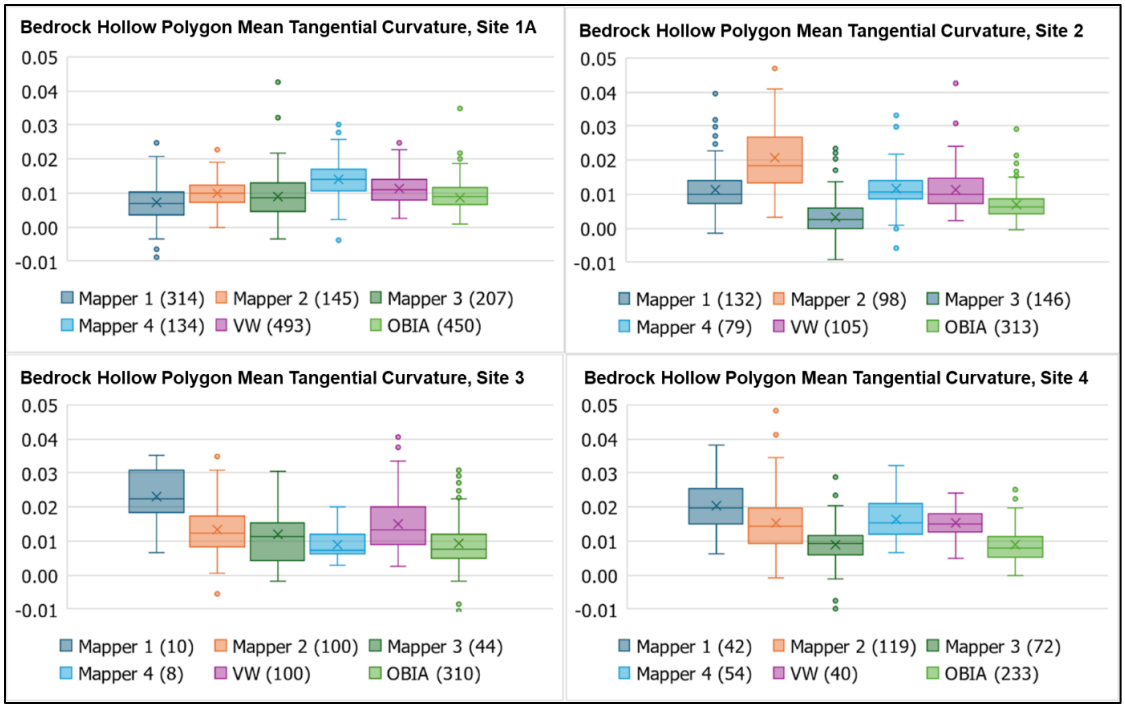


Figure 29: Box and whisker plots of the mean tangential curvature for bedrock hollow RIL polygons for all mappers, including the two computer models, at all four study sites (Site 1A = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

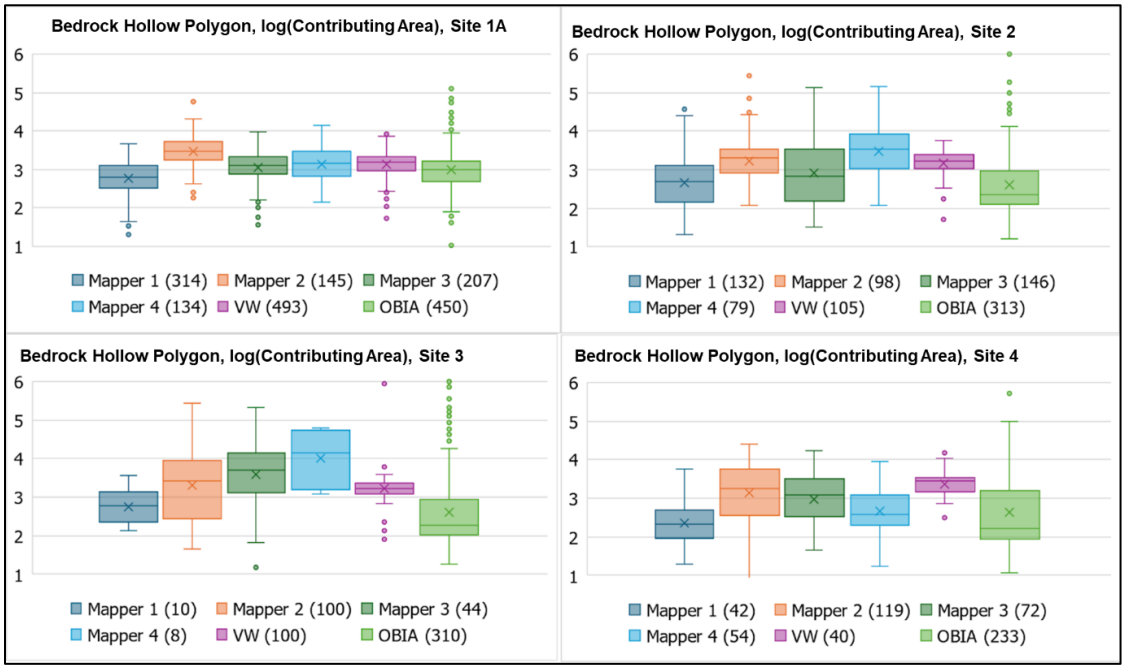


Figure 30: Box and whisker plots of the logarithm (base 10) of the contributing area (m^2) at the outlet of the bedrock hollow RIL polygons for each mapper, including the two computer models, at each of the four study sites (Site 1A = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

4.1.2.6 Voting: Agreement between Mappers and Models at Site 1A or North Fork Calawah

Evaluation of the computer mapped RILs was completed using a voting exercise described as follows: The 47 randomly selected 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 adequately identified the presence of an actual RIL or "no" if it did not. Results of these votes are shown in Table 10 and Figure 31.

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%

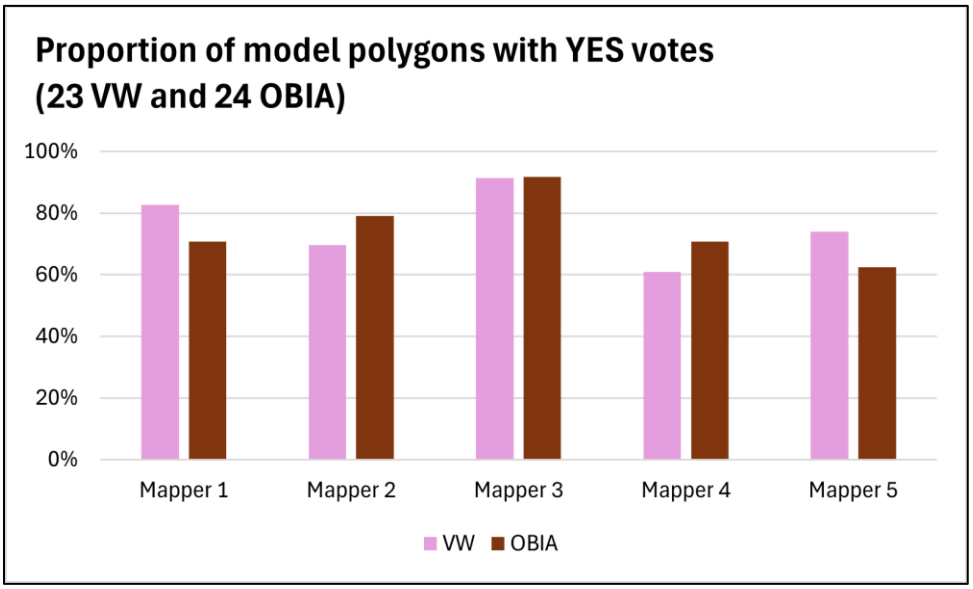


Figure 31: Mappers' agreement with computer models identification of potential bedrock hollow sites.

Table 11 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 polygon site delineated by one of the models, that mapper had also drawn a polygon. Because Mapper 5 only mapped a portion of the study area, they are excluded from these results. The number of initial matches each mapper had with the models varied substantially, ranging from

Mapper 1, who had 35 matches out of the 47 (74%) modeled polygons, to Mapper 4, who matched only 18 out of the 47 (38%) model polygons (Table 10). This is consistent with the relative number of polygons each of them mapped: Mapper 1 identified more bedrock hollow RIL polygons (314) at Site 1A than Mapper 4 (134). All 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 identify the presence of a bedrock hollow RIL, averaging 75% yes-votes for the OBIA polygons and 76% for the VW polygons.

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%)

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 32).

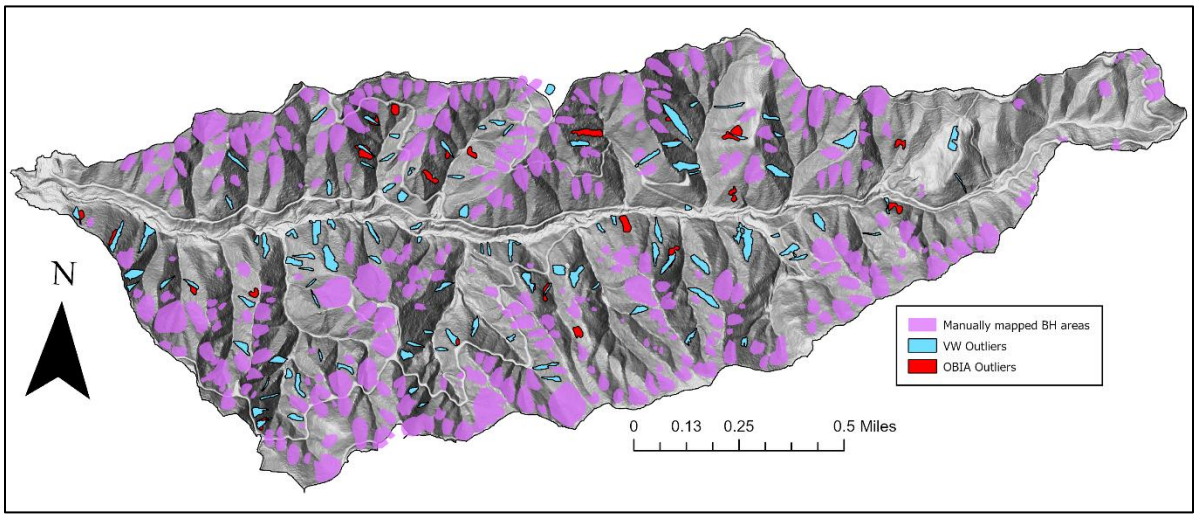


Figure 32: Site 1A or North Fork Calawah with manually mapped bedrock hollow areas highlighted in purple, OBIA outliers highlighted in red and VW outliers highlighted in blue.

Of these “outlier” model polygons (those with no matching manually mapped polygons) we randomly chose 28 from each model set (Figure 33). The mapping team individually voted on these 56 outlier polygons without team discussion, 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 34 for each mapper.

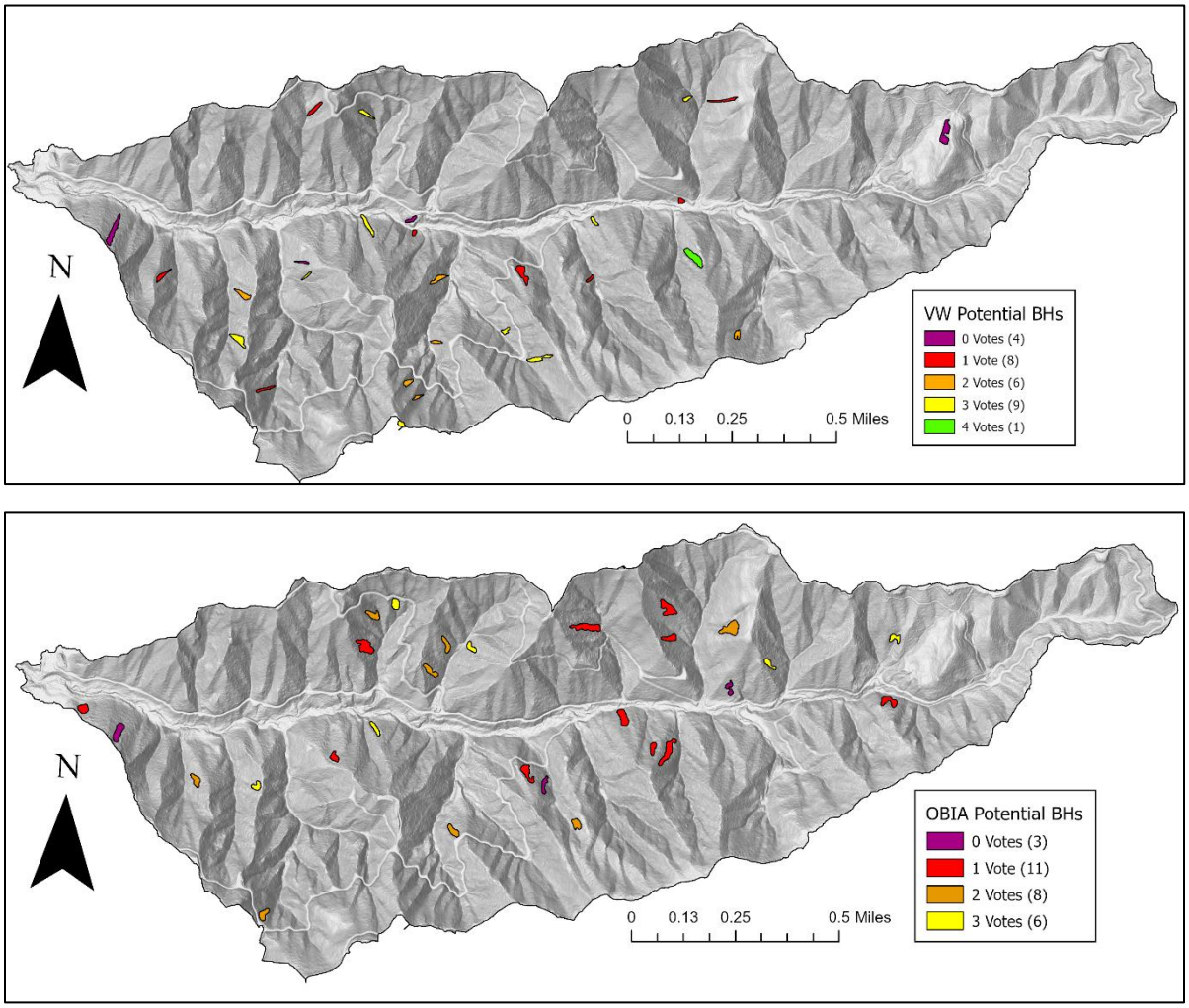


Figure 33: Modeled polygon outliers with colors representing the number of “yes” votes at Site 1A or North Fork Calawah.

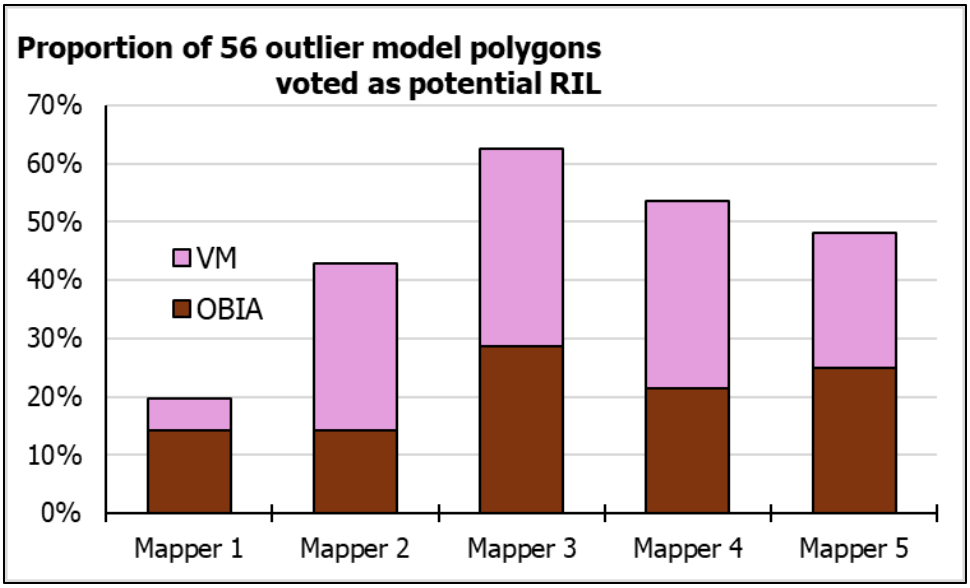


Figure 34: Proportion of "yes" votes awarded by each mapper for the outliers in each modeled polygon sets in Site 1A or North Fork Calawah.

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 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.

4.1.3 PHASE 1 CONCLUSIONS

Phase 1 revealed 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 unsatisfactory, and efforts to align the overlap did not produce useful results.

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 recognized bedrock hollows in model-generated polygons that they had not initially delineated themselves. 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 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? 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 adjust the model's delineations of defined landform class boundaries to better meet the EP's expectations.
3. Was deliverability considered for this polygon? If so, could the landform deliver sediment to a public resource (e.g., a stream channel)? A yes-vote indicated that the model appeared to consider deliverability, and that the EP agreed the landform could deliver sediment. Answers in this category helped determine if the model focuses on those landforms that meet this critical part of rule criteria or if it included features that the EPs felt would not deliver.
4. Was convergence considered for this polygon (did the EP consider convergence when voting yes-or-no)? While the Board Manual criteria for convergence are 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 deep seated landslide)?

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.

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. The number of modeled landform polygons within any of the study sites is variable, but on the order of hundreds, not thousands. The appropriate sample size for smaller populations, such as our modeled landform polygons, can be reduced slightly from that indicated by Cochran's formula above (Divakar, 2021). A population with N members, may be adequately represented with a sample n given by:

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

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 35 shows the percentage of potential bedrock hollow polygons to receive yes-votes for prompts 1 & 2 across all sites for both models. The "reasonable screen yes-vote percentage" represents the proportion of the total polygon set that received a yes-vote for prompt 1. The "correctly drawn 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 reasonable screen-percentage at 91%, but the third-lowest correctly drawn-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 reasonable screen and 41% for correctly drawn. Of the OBIA polygons, Sites 1A, 2 & 4 had average reasonable screen yes-vote percentages within 1% of each other and an average of 71%, though the average correctly drawn yes-vote percentages varied more than reasonable screen and had an average of 46%.

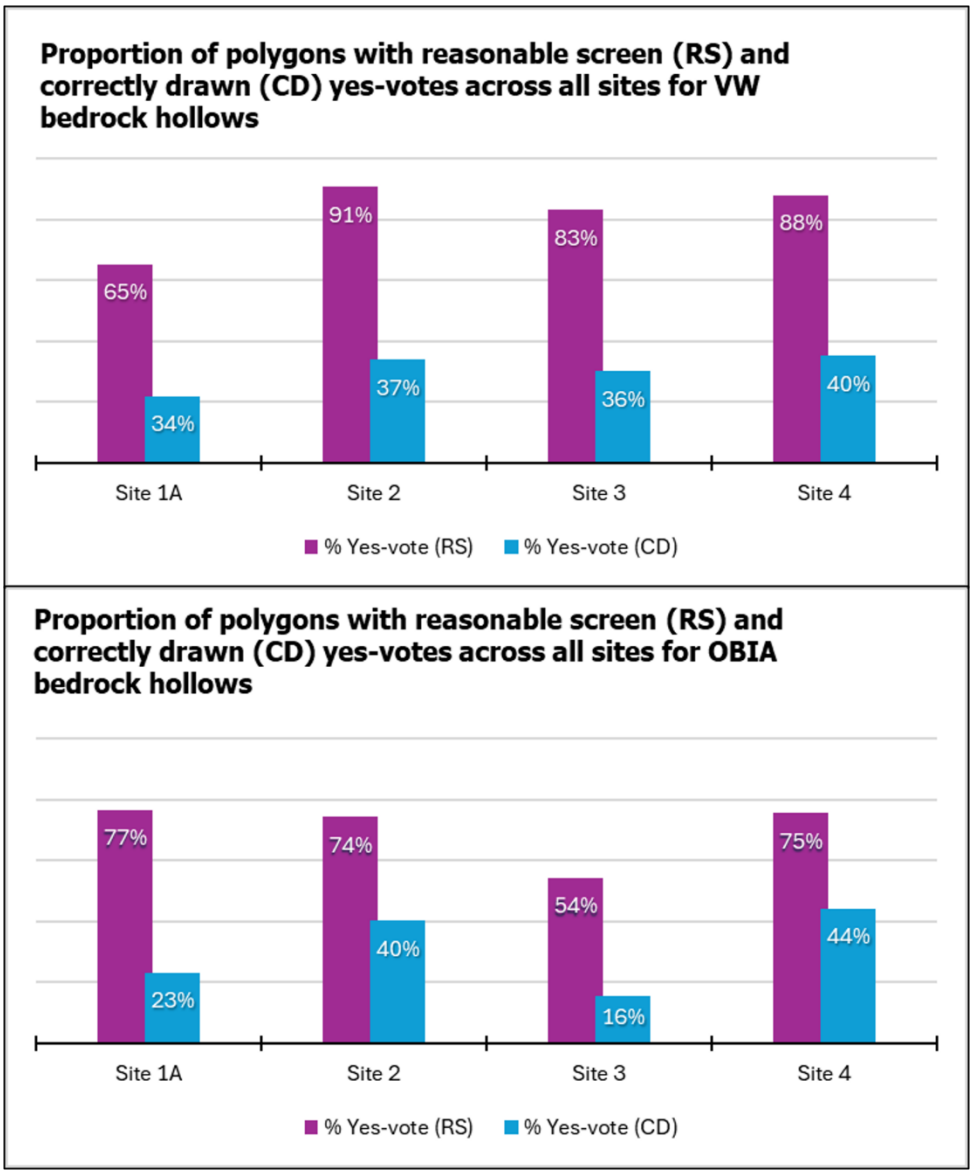


Figure 35: Bedrock hollow reasonable screen (RS) and correctly drawn (CD) yes-vote percentages across all sites for the total VW (top) and OBIA (bottom) modeled polygon sets. For CD, the percentages shown represent the proportion of polygons, out of the subset of polygons that received yes-votes for RS, also received yes-votes for CD (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

Inner Gorges

Figure 36 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 reasonable screen yes-vote

percentages in bedrock hollows ranged between 12-29%, whereas the difference for inner gorges ranged between 0-7%. Additionally, the correctly drawn yes-vote percentages were higher overall for inner gorges than the bedrock hollows.

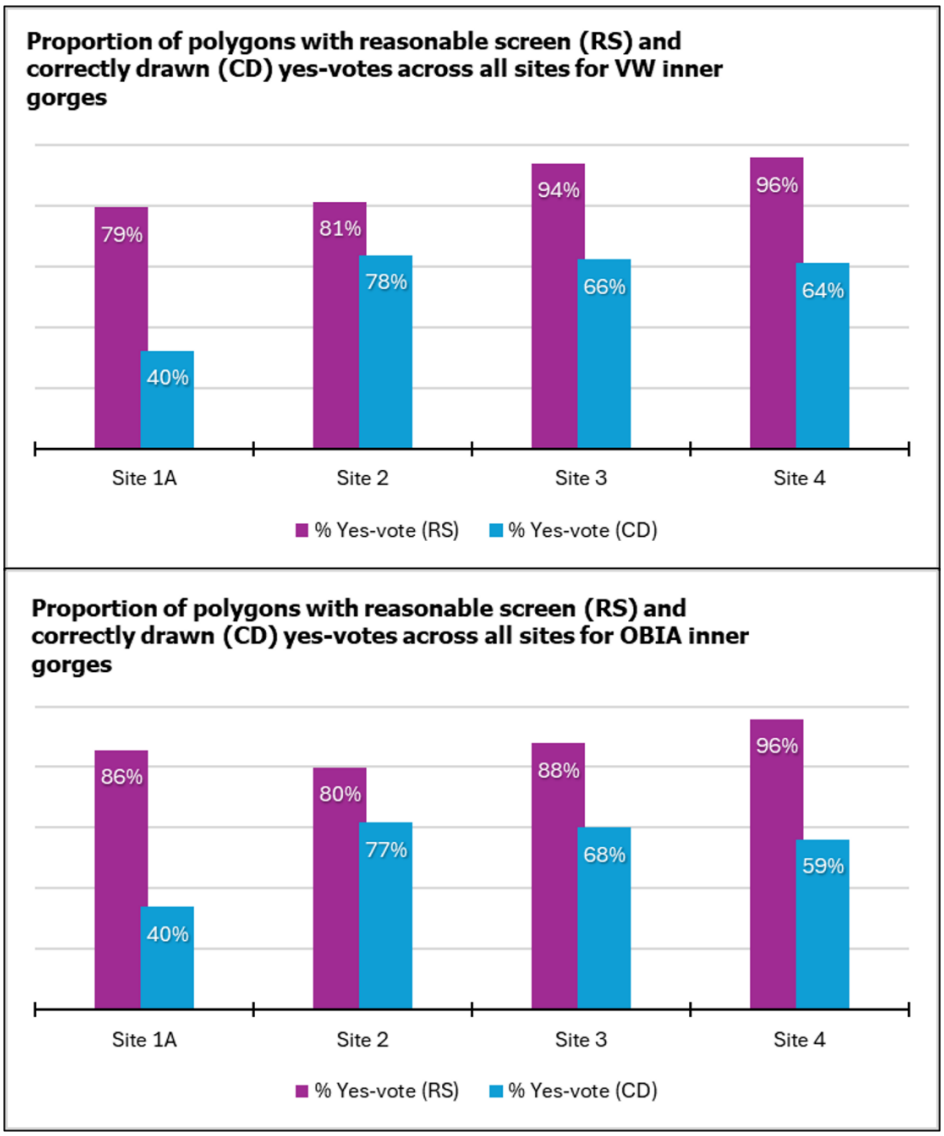


Figure 36: Inner gorge reasonable screen (RS) and correctly drawn (CD) yes-vote percentages across all sites for the total VW (top) and OBIA (bottom) modeled polygon sets. For CD, the percentages shown represent the proportion of yes-vote RS polygons that received CD yes-votes (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

Key observations from the Phase 2 exercise:

- For the VW model, polygons at Site 1A had the lowest average reasonable screen and correctly drawn 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.

- 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 or North Fork Calawah

Bedrock Hollows

The voting results for Site 1A potential bedrock hollow polygons for each model are shown in Tables 12 & 13. 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%, and of those, they considered 35% to be correctly drawn.

Table 12: Voting results for Site 1A or North Fork Calawah VW potential bedrock hollow polygons (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/81)	%Yes (CD/RS)
M1	49	0	60%	0%
M2	51	15	63%	29%
M3	45	35	56%	78%
M4	66	21	81%	32%
Average			65%	34%

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, 30% of which they considered to be correctly drawn.

Table 13: Voting results for Site 1A or North Fork Calawah OBIA potential bedrock hollow polygons (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/79)	%Yes (CD/RS)
M1	69	0	87%	0%
M2	71	28	90%	39%
M3	50	34	63%	68%
M4	52	11	66%	21%
Average			77%	30%

The mappers voted unanimously on 70 (44%) of the 160 total potential bedrock hollow polygons produced by both models. Figure 37 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 38 shows the distribution of mapper agreement on reasonable-screen designation.

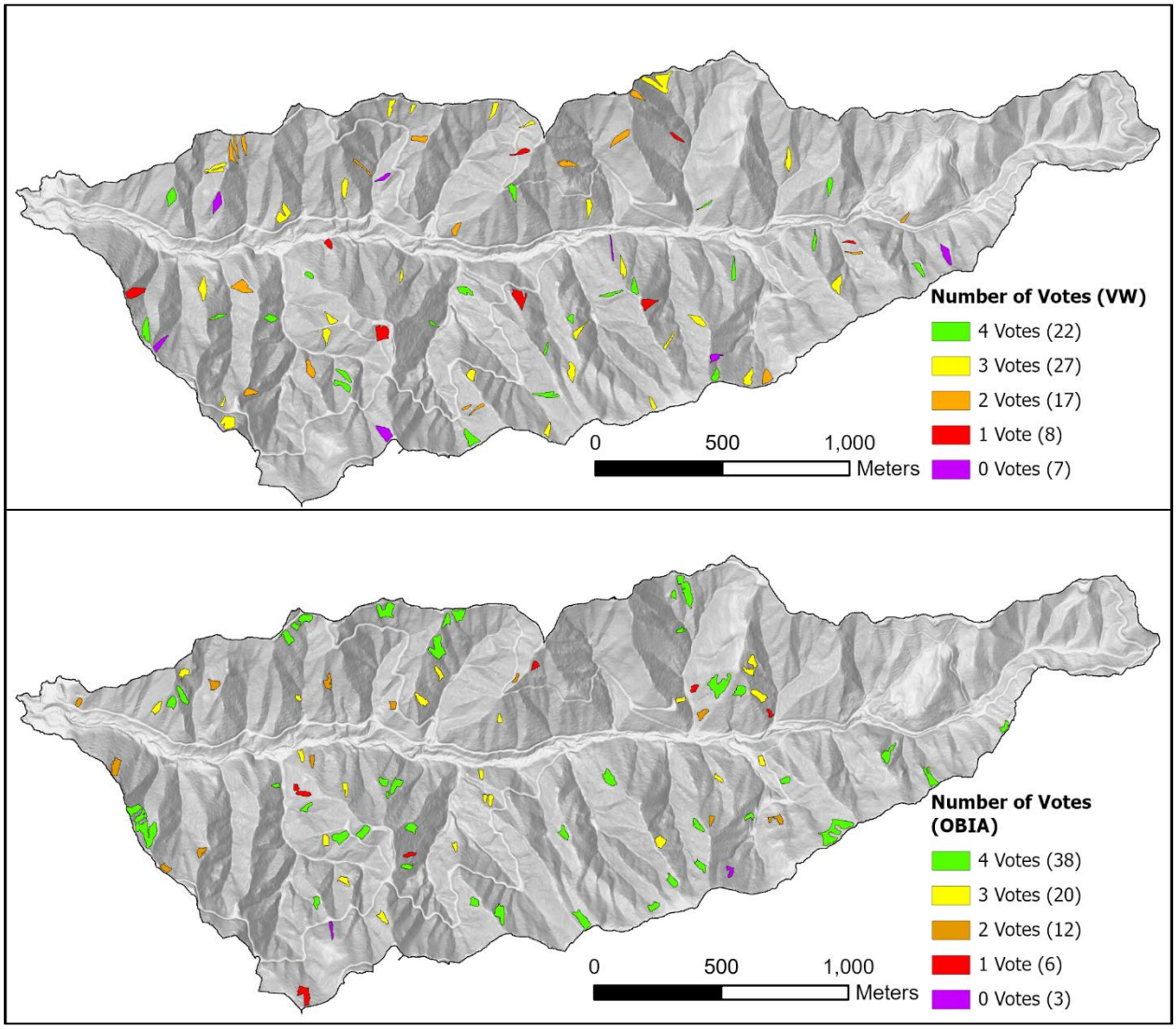


Figure 37: Number of reasonable screen yes-votes received for each polygon at Site 1A or North Fork Calawah in the potential bedrock hollow sets (VW top and OBIA bottom) and how many polygons received that number of votes.

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 14 & 15. 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.

Table 14: Voting results for VW potential inner gorge polygons at Site 1A or North Fork Calawah (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/81)	%Yes (CD/RS)
M1	78	18	96%	23%
M2	74	46	91%	62%
M3	72	38	89%	53%
M4	33	2	41%	6%
Average			79%	40%

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 or North Fork Calawah (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/76)	%Yes (CD/RS)
M1	72	8	95%	11%
M2	71	35	93%	49%
M3	68	43	89%	63%
M4	49	17	64%	35%
Average			86%	40%

The mappers voted "yes" unanimously on 79 (50%) of the 157 total potential inner gorge polygons produced by both models. Figure 38 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 38 shows the distribution of mapper agreement on RS designation.

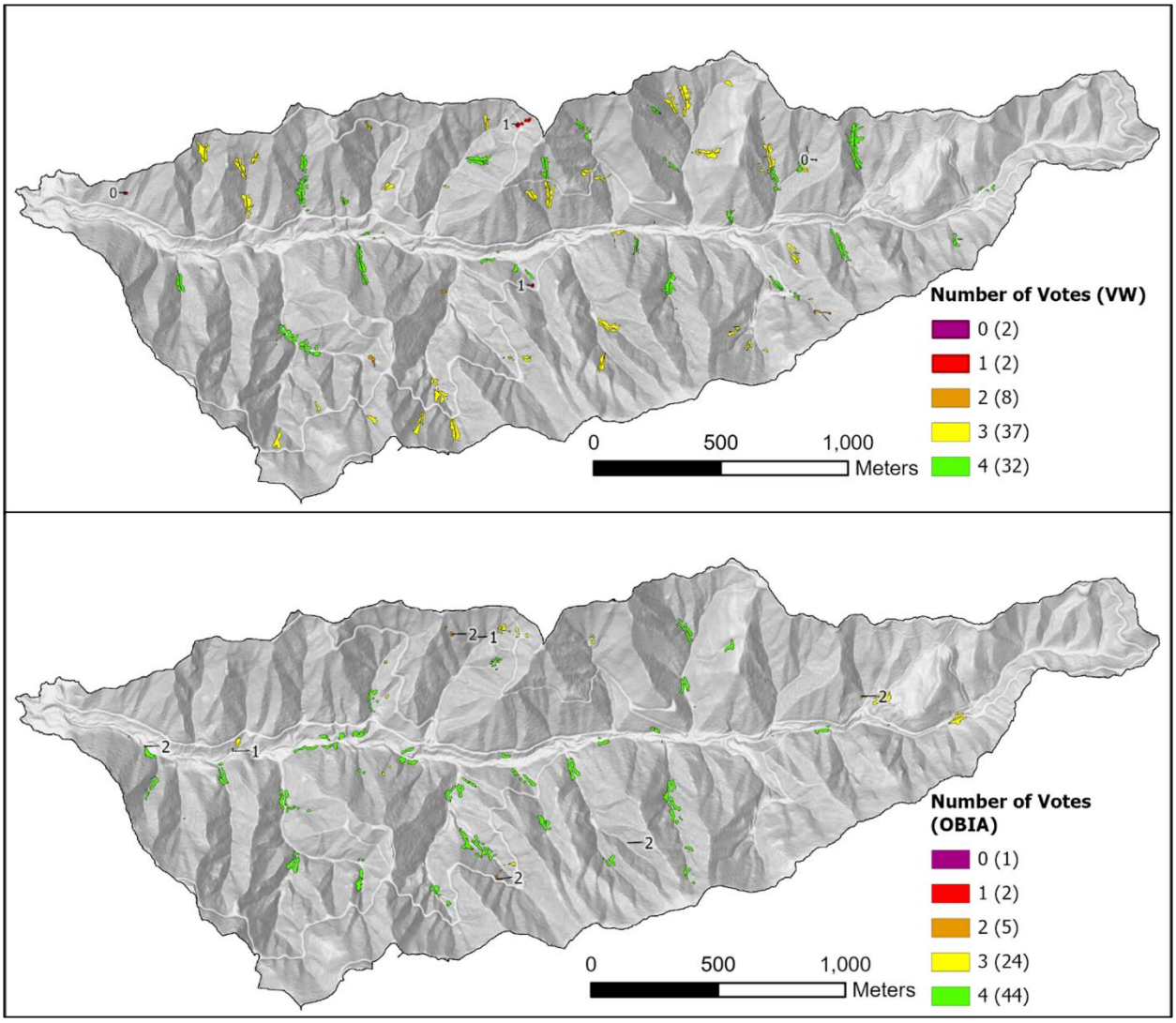


Figure 38: Number of reasonable-screen yes-votes received for each polygon in the potential inner gorge sets (VW top and OBIA bottom) for Site 1A or North Fork Calawah and how many polygons received that number of votes.

Site 2 or Willapa Hills

Bedrock Hollows

The voting results for Site 2 potential bedrock hollow polygons for each model are shown in Tables 16 & 17. 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 or Willapa Hills (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/51)	%Yes (CD/RS)
M1	51	8	100%	16%
M2	41	15	80%	37%
M3	49	34	96%	69%
M4	44	12	86%	27%
Average			91%	37%

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 or Willapa Hills (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/66)	%Yes (CD/RS)
M1	58	13	97%	22%
M2	41	28	62%	68%
M3	56	54	85%	96%
M4	41	11	62%	27%
Average			74%	54%

The mappers voted "yes" unanimously on 64 (55%) of the 117 total potential bedrock hollow polygons produced by both models. Figure 39 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 39 shows the distribution of mapper agreement on RS designation.

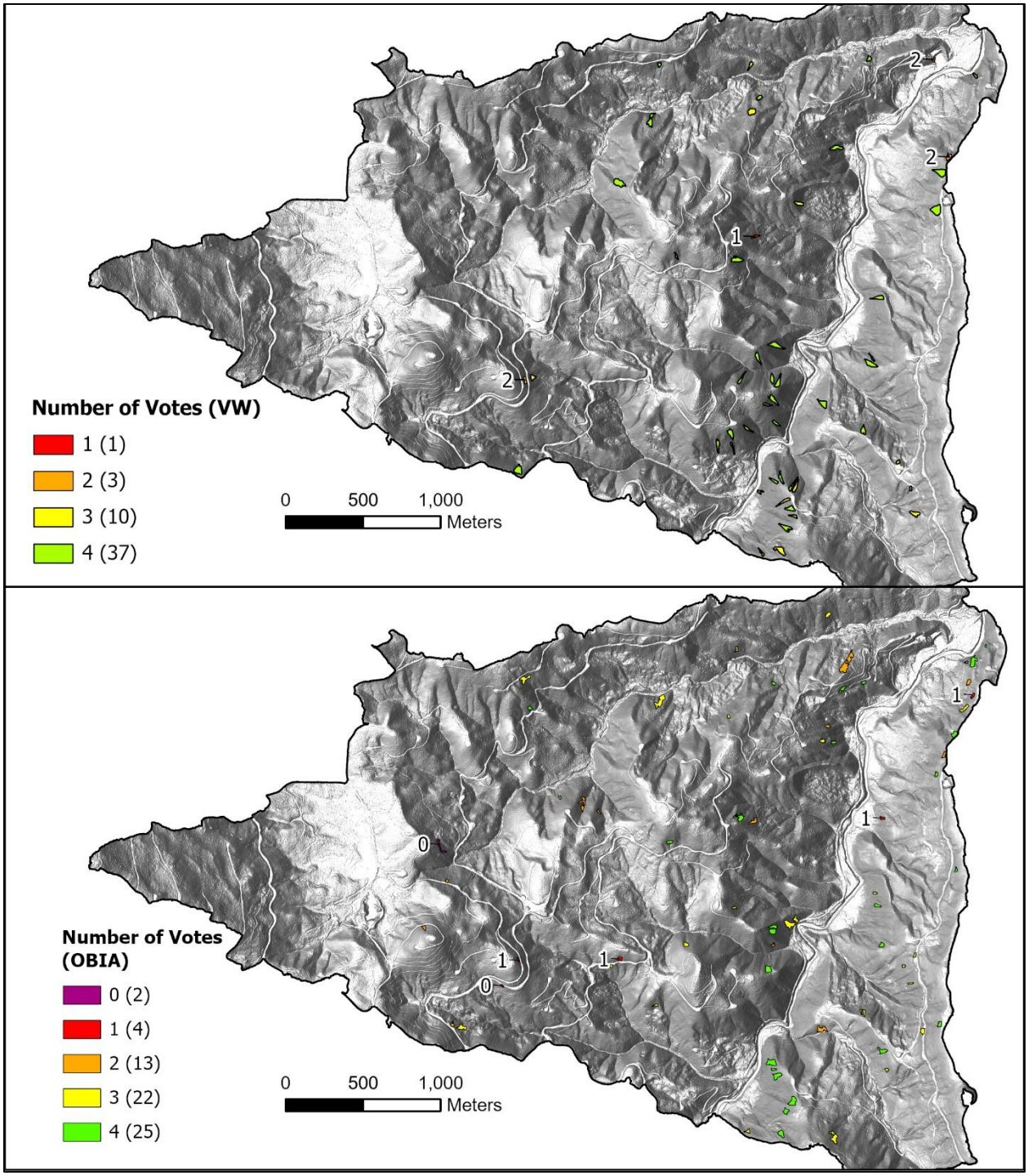


Figure 39: Number of yes-votes received for each polygon in the potential bedrock hollow sets (VW top and OBIA bottom) for Site 2 or Willapa Hills and how many polygons received that number of votes.

Inner gorges

The voting results for Site 2 potential inner gorge polygons for each model are shown in Tables 18 & 19. 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 or Willapa Hills (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/84)	%Yes (CD/RS)
M1	71	67	90%	94%
M2	76	63	90%	83%
M3	79	67	94%	85%
M4	47	17	56%	36%
Average			81%	78%

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 at Site 2 or Willapa Hills (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/81)	%Yes (CD/RS)
M1	65	61	92%	94%
M2	71	50	88%	70%
M3	74	65	91%	88%
M4	49	24	60%	49%
Average			80%	77%

The mappers voted "yes" unanimously on 85 (52%) of the 165 total potential inner gorge polygons produced by both models. Figure 40 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 40 shows the distribution of mapper agreement on RS designation.

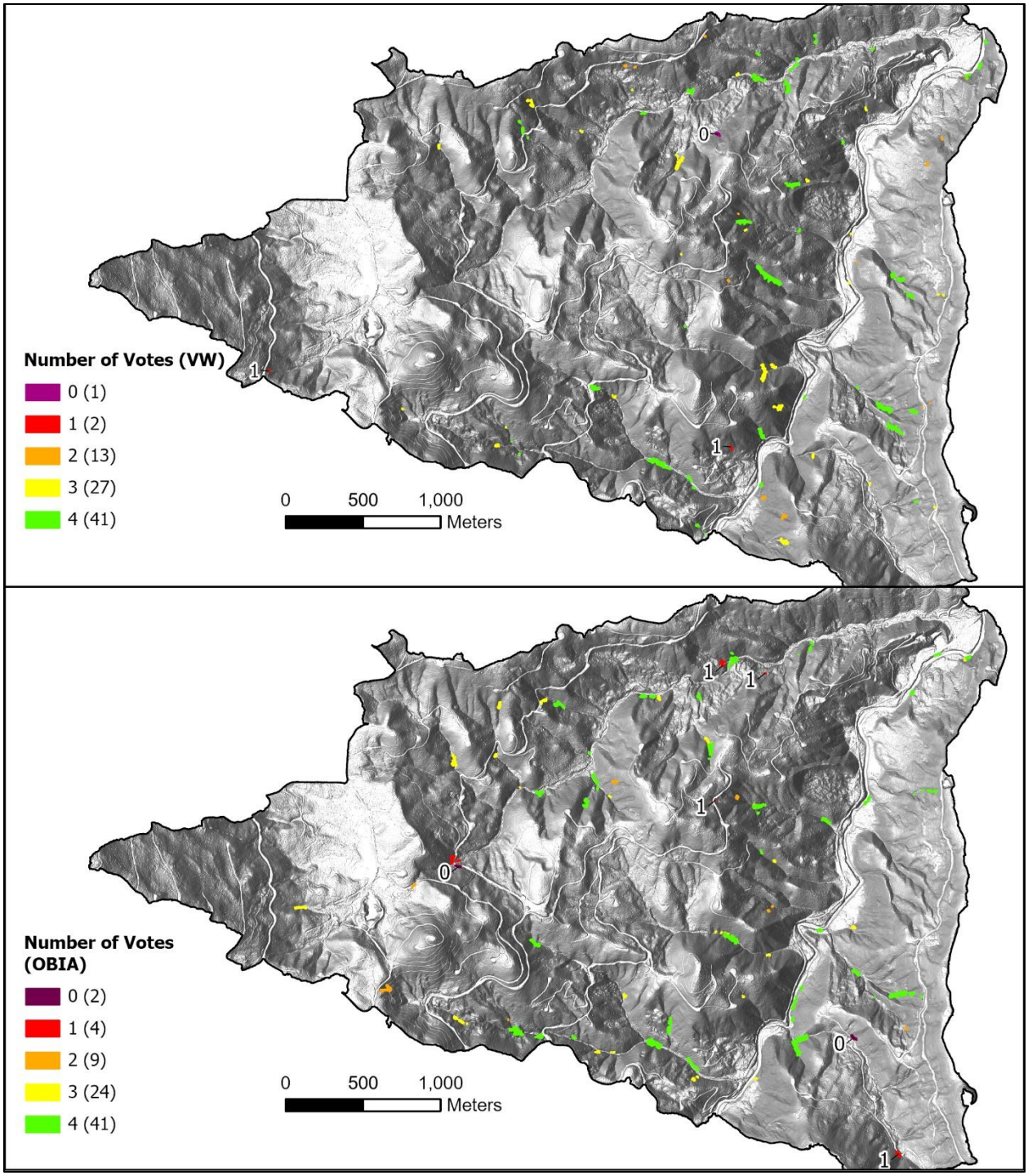


Figure 40: Number of yes-votes received for each polygon in the potential inner gorge sets (VW top and OBIA bottom) for Site 2 or Willapa Hills and how many polygons received that number of votes.

Site 3 or Howard Creek

Bedrock Hollows

The voting results for Site 3 potential bedrock hollow polygons for each model are shown in Tables 20 & 21. 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 or Howard Creek (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/78)	%Yes (CD/RS)
M1	77	17	99%	22%
M2	67	34	86%	51%
M3	63	33	81%	52%
M4	52	10	67%	19%
Average			83%	36%

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 or Howard Creek (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/90)	%Yes (CD/RS)
M1	77	12	86%	16%
M2	41	21	46%	51%
M3	36	11	40%	31%
M4	42	12	47%	29%
Average			54%	29%

The mappers voted "yes" unanimously on 74 (44%) of the 168 total potential bedrock hollow polygons produced by both models. Figure 41 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 41 shows the distribution of mapper agreement on RS designation.

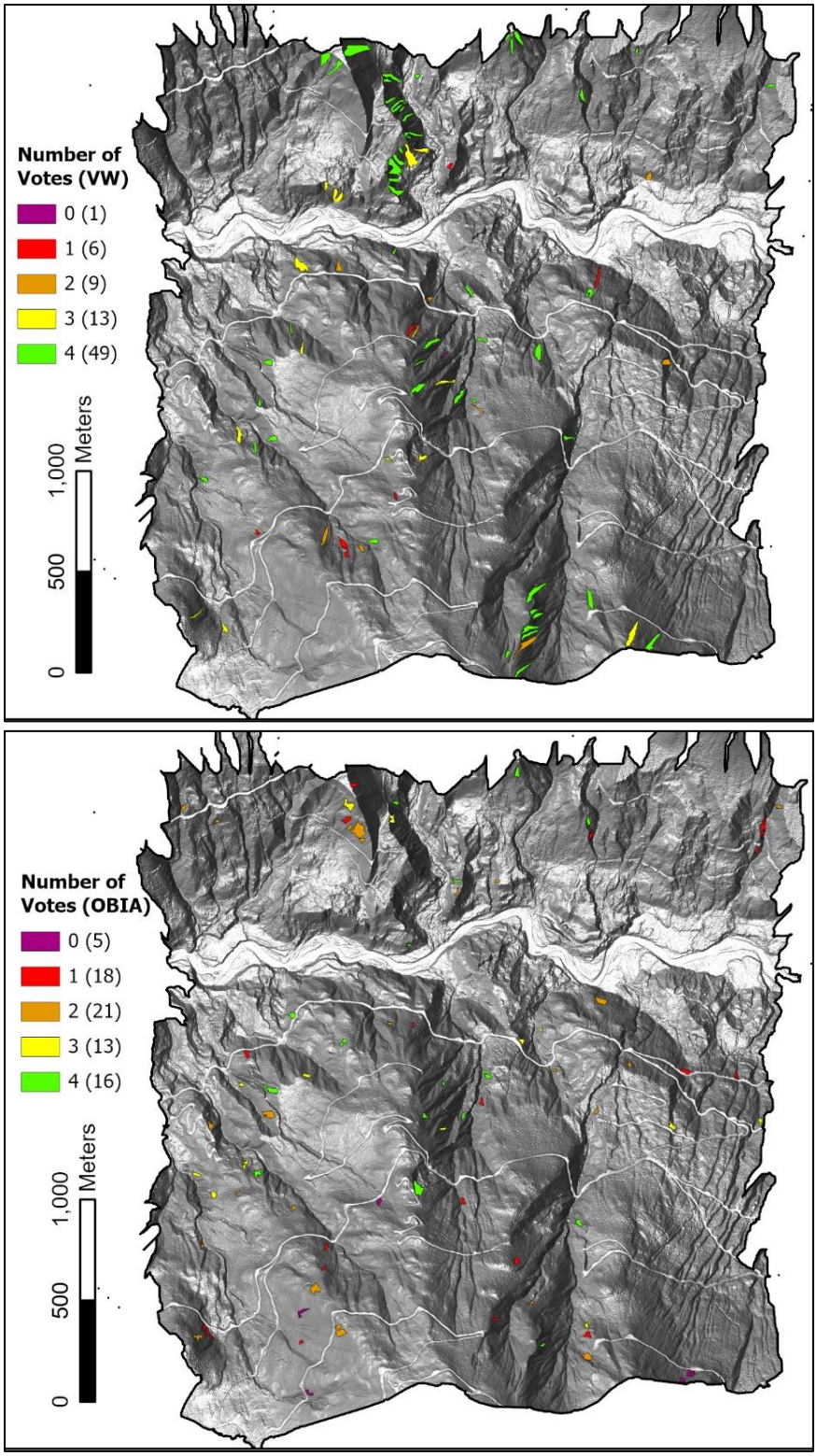


Figure 41: Number of yes-votes received for each polygon in the potential bedrock hollow sets (VW top and OBIA bottom) for Site 3 or Howard Creek and how many polygons received that number of votes.

The voting results for Site 3 potential inner gorge polygons for each model are shown in Tables 22 & 23. 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 or Howard Creek (RS = reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/88)	%Yes (CD/RS)
M1	88	79	100%	90%
M2	88	80	100%	91%
M3	85	52	97%	61%
M4	69	8	78%	12%
Average			94%	66%

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 or Howard Creek (RS = reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/80)	%Yes (CD/RS)
M1	80	55	100%	69%
M2	72	56	90%	78%
M3	68	49	85%	72%
M4	61	32	76%	52%
Average			88%	68%

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 42 shows the number of yes-votes each polygon in the set received. Figure 42 shows the distribution of mapper agreement on RS designation.

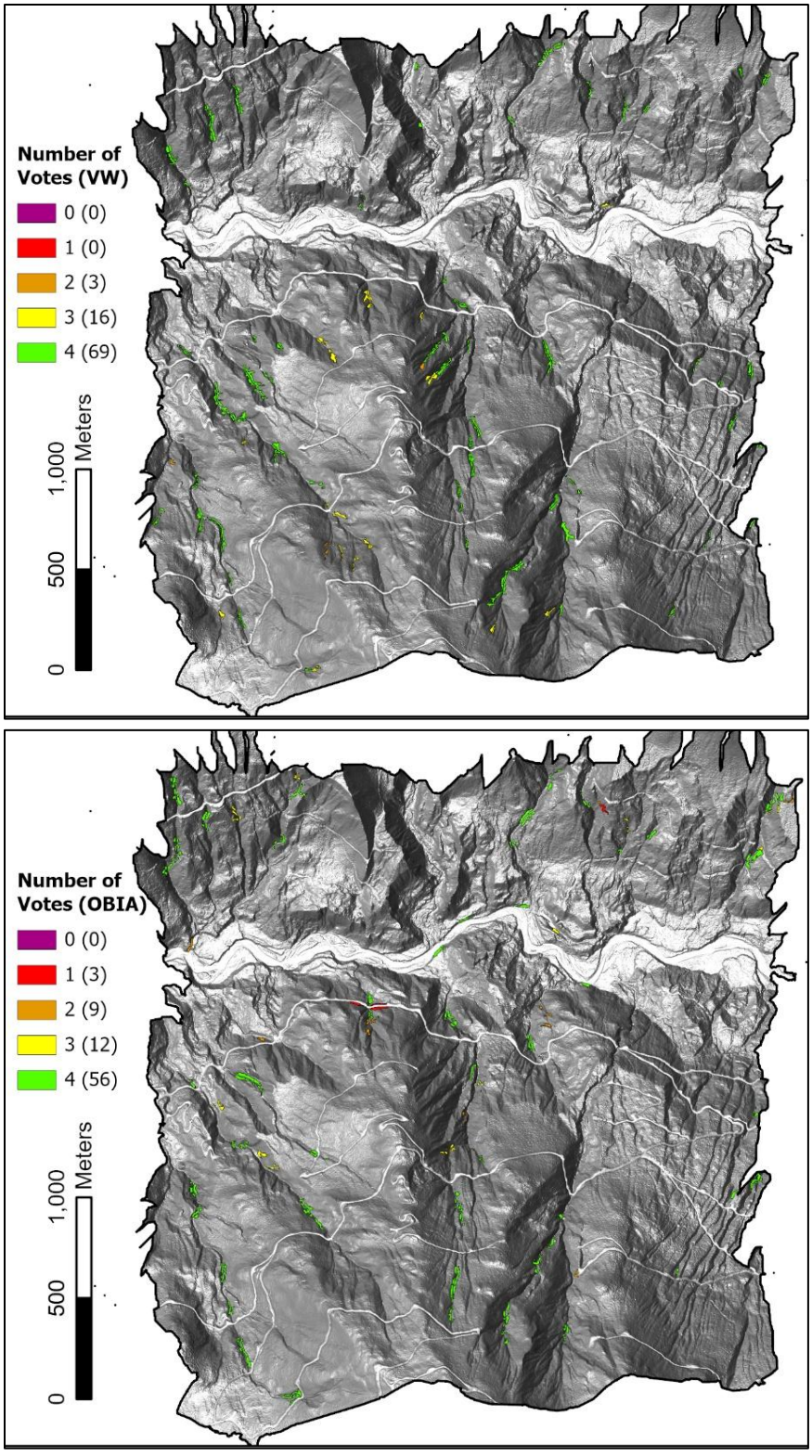


Figure 42: Number of reasonable screen yes-votes received for each polygon in the potential inner gorge sets (VW top and OBIA bottom) for Site 3 or Howard Creek and how many polygons received that number of votes.

Site 4 or Wishkah

Bedrock Hollows

The voting results for Site 4 potential bedrock hollow polygons for each model are shown in Tables 24 & 25. 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 or Wishkah (RS=reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/27)	%Yes (CD/RS)
M1	27	2	100%	7%
M2	22	15	81%	68%
M3	26	18	96%	69%
M4	20	3	74%	15%
Average			88%	40%

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 or Wishkah (RS = reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/59)	%Yes (CD/RS)
M1	54	18	92%	33%
M2	48	35	81%	73%
M3	46	42	78%	91%
M4	30	9	51%	30%
Average			75%	58%

The mappers voted "yes" unanimously on 43 (50%) of the 86 total potential bedrock hollow polygons produced by both models. Figure 43 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 43 shows the distribution of mapper agreement on RS designation.

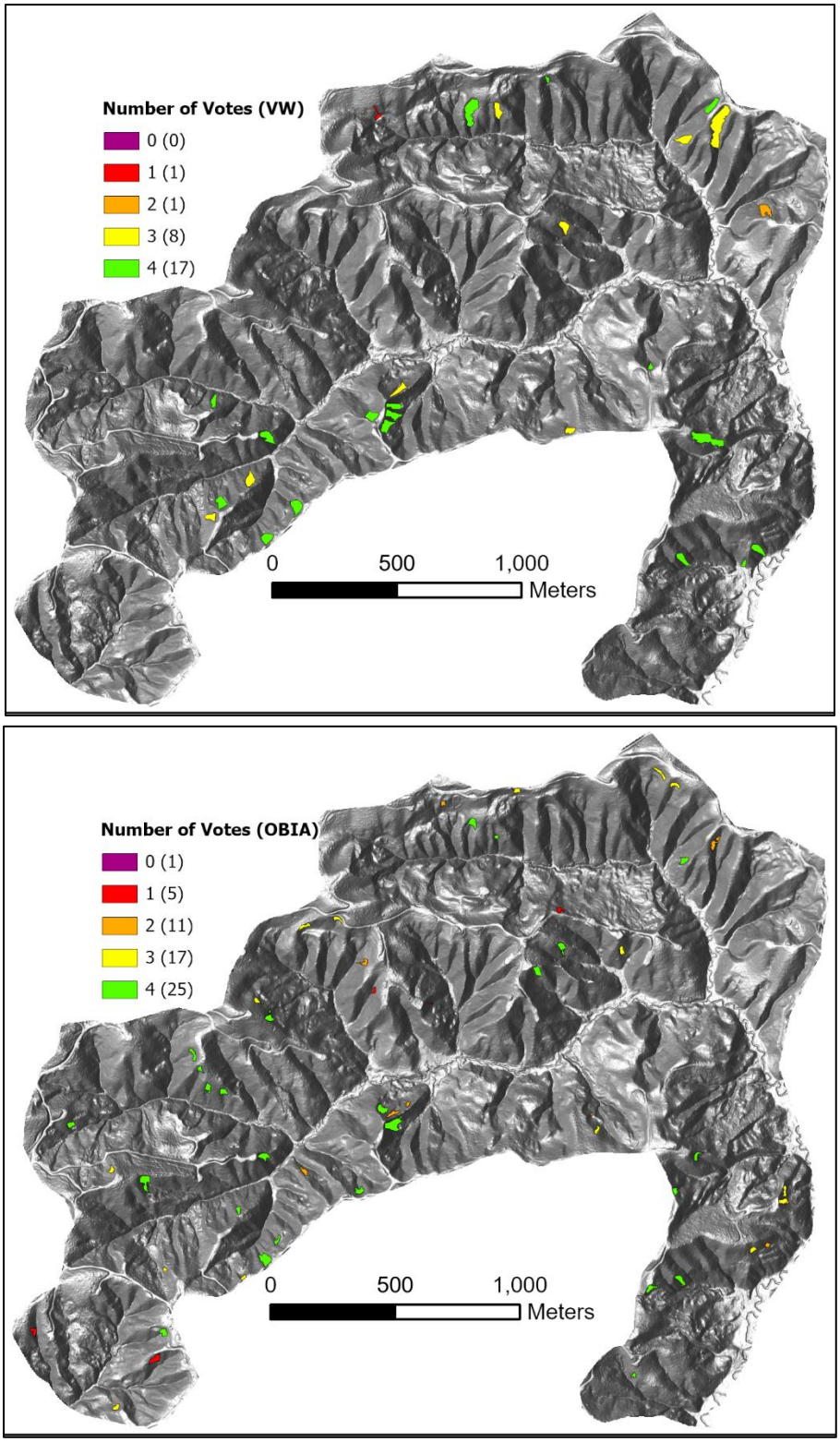


Figure 43: Number of yes-votes received for each polygon in the potential bedrock hollow sets (VW top and OBIA bottom) for Site 4 or Wishkah and how many polygons received that number of votes.

Inner Gorges

The voting results for Site 4 potential inner gorge polygons for each model are shown in Tables 26 & 27. 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 or Wishkah (RS = reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/68)	%Yes (CD/RS)
M1	68	61	100%	90%
M2	68	59	100%	87%
M3	68	40	100%	59%
M4	56	6	82%	11%
Average			96%	64%

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 or Wishkah (RS = reasonable screen; CD = correctly drawn).

	#Yes-votes for prompt 1 (RS)	#Yes-votes for prompt 2 (CD)	%Yes (RS/64)	%Yes (CD/RS)
M1	63	50	98%	79%
M2	63	33	98%	52%
M3	64	49	100%	77%
M4	55	12	86%	22%
Average			96%	59%

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 44 shows the number of yes-votes each polygon in the set received. Figure 44 shows the distribution of mapper agreement on RS designation.

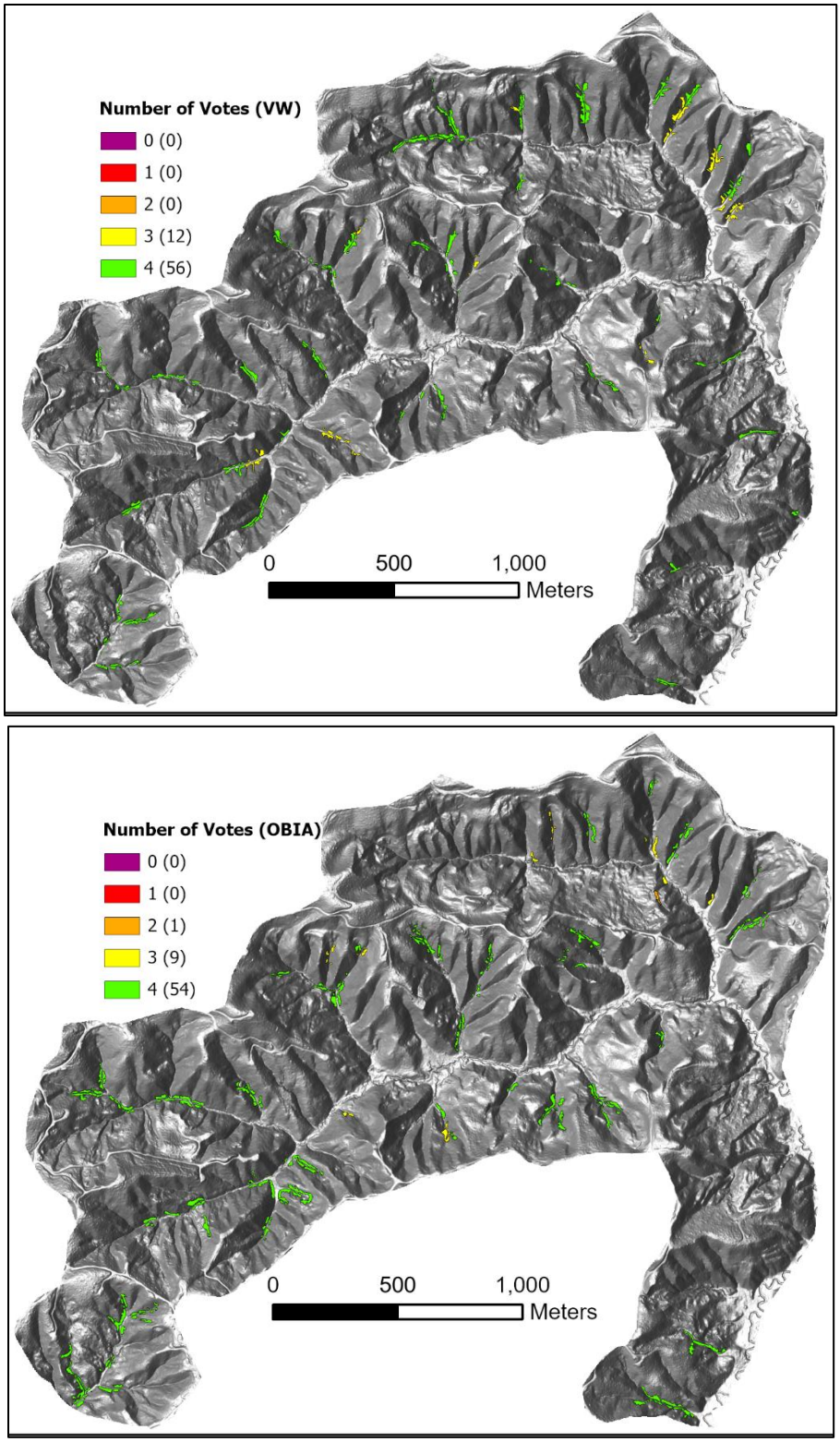


Figure 44: Number of yes-votes received for each polygon in the potential inner gorge sets (VW top and OBIA bottom) for Site 4 or Wishkah and how many polygons received that number of votes.

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 for the presence of target landforms and (2) correctly draw their boundaries.

Results indicated that both models generally performed well in identifying landform presence as perceived by the EPs, 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.

Delineation 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. Inner gorge polygons consistently received higher correctly drawn ratings than bedrock hollows, and VW model polygons tended to perform slightly better on average for both reasonable screen and correctly drawn criteria.

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 identify specific input parameters, particularly those controlling convergence and planform thresholds, to focus on as we compare modeled landform delineations to precise landslide locations inventoried with the next sub-project.

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 and Future Work

5.1 Why we compared OBIA results to a new set of manually mapped landforms

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 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, for example, 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 sub-project (Susceptibility and Runout), we seek to develop reference maps using lidar-differencing derived landslide inventories. Repeat lidar data is available through the Washington Lidar Portal. We will use these new reference maps to calibrate our OBIA models. As described in detail in Sections 5.3 and 5.4, we intend to identify the range of terrain elements 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. 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 5.3 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 sub-project, 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.

5.2 AUTOMATED MAPPING LIMITATIONS

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 headwall or topography without quantifiably defining convergence. As described in section 4.1.2, the OBIA-based models use tangential curvature to set specific convergence parameters. With the next sub-project, we will develop landslide inventories based on 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.

5.3 IMPLICATIONS FOR THE NEXT SUB-PROJECTS

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 manually developed reference maps. We parameterized our OBIA models to match those reference maps and evaluated performance based on model ability to match the empirical probability distributions of key terrain elements like slope and curvature. However, we observed high variability among experienced practitioners when manually delineating landforms from lidar-derived topographic data which limited reliability of the manual reference maps. Despite shared guidance and criteria, individual interpretation of terrain features led to unsatisfactory spatial agreement, suggesting that manual mapping alone introduces subjectivity. Voting exercises further revealed that mappers recognized valid landforms in model-generated polygons they had not originally mapped.

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 could be parameterized to perform consistently well in identifying potential RIL locations, especially for inner gorges. However, delineation accuracy, particularly for bedrock hollows, was more variable. As discussed in Section 4.2.3, EP review highlighted specific model limitations related to topographic convergence, slope thresholds, and stream adjacency.

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 (described in Section 5.4).

The next Susceptibility and Runout sub-project will incorporate expert review of the landslide-inventory-derived reference maps. 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.

Once we have a reliable reference map, we will calibrate our OBIA models to those datasets. 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.

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.

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. 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.

5.4 How we will achieve Objective 2 in the next sub-project

This sub-project did not fulfill Objective 2, "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." Below we provide an overview of the steps we will follow to achieve objective 2 in the next sub-project. Potential field verification and consultation of independent statisticians and Qualified Experts will be threaded throughout the implementation steps of the Susceptibility and Runout Project.

Step 1. Collect landslide inventories
Using lidar differencing to obtain these inventories, landslide locations will be precisely correlated with pre-landslide terrain features. We will then use the mapped landslide locations, sizes, and volumes to calculate spatial landslide densities in terms of the number, area, and volume of landslides per unit area.

Step 2. Correlate landslide density to terrain features
We will seek to measure landslide density as a function of measurable landscape characteristics. The terrain features we will examine include the types of elevation derivatives discussed in this report – gradient, curvature, contributing area – and other covariates, including substrate types (soil and geologic units), forest age, and time series of precipitation.

Step 3. See if inventoried landslides fall into distinct terrain-feature groupings
We will do this using both unsupervised machine-learning approaches (e.g., k-means clustering) and by visual interpretation of the landform type each inventoried landslide is located within.

The computer-generated groupings show how observed landslides are spatially distributed across distinct combinations of terrain-attribute types and values, which we can compare to the landform criteria used to identify and delineate RILs. Comparison of the visual interpretations with the computer-generated groupings will show how our current landform definitions and our interpretation of those definitions compare to where observed landslides are actually located in the landscape.

Step 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.

Step 5. Overlay the inventoried landslides on the delineated landforms.
We can then calculate landslide density for each landform type and identify landslides that fall outside of any modeled RILs. This step will be repeated as model parameters are adjusted based on results from step 3 above and step 7 below.

Step 6. Compare the frequency distribution of terrain-element values within the modeled landform polygons to the distributions within the mapped landslide polygons.
We can then adjust the computer-model inputs to ensure the modeled landforms encompass the same range of attributes.

Step 7. Iterate the previous steps.
Steps 3, 6 and 7 will provide alternative rule sets for the models. 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 modeled with these different rule sets. We will then look at how sensitive the densities and proportions are to variations in the rules used to delineate the landforms.

Step 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.

6.0 CONCLUSIONS

While the models produced results that are broadly consistent with expert-drawn maps, there is variation between model outputs and among expert interpretations. This variation could highlight 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 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 sub-project, 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 Sections 5.2-5.4) in the next project phase.

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.

7.0 LITERATURE CITED

Acker, S. A., Reeves, G. H., Hogervorst, J. B., Blundon, B., Ian-Huei, Y., Bell, D. M., 2023, Classification and Assessment of Riparian Ecosystems in Northwest Oregon for Restoration Planning, Northwest Science, v.96, no. 3-4. P. 184-205.

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

Atwood E. C., Englhart S., Lorenz E., Halle W., Wiedemann W., Siegert F., 2016, Detection and Characterization of Low Temperature Peat Fires during the 2015 Fire Catastrophe in Indonesia Using a New High-Sensitivity Fire Monitoring Satellite Sensor (FireBird). PLoS ONE 11(8): e0159410. doi:10.1371/journal.pone.0159410

Avcioglu, B., Anderson, C. J., 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., 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., 2011, Creating a catchment scale perspective for river restoration, Hydrology and Earth System Science, v. 15, p. 2995-3015.

Benda, L., Miller, D., Barquin, J., McCleary, R. J., Cai, T., 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., 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., 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., 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., Slater, L. J., 2014, Objective extraction of channel heads from high-resolution topographic data: Water Resources Research, v. 50, p. 4283-4304.

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

Dietrich, W.E., Wilson, C.J., 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

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

Dietrich, W. E., Bellugi, D. G., Sklar, L. S., Stock, J. D., Heimsath, A. M., Roering, J. J., 2003, Geomorphic Transport Laws for Predicting Landscape form and Dynamics. In P. R. Wilcock & R. M. Iverson (Eds.), Geophysical Monograph Series (pp. 103–132). American Geophysical Union. https://doi.org/10.1029/135GM09

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., 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.

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.

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., 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.

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., Meyers, K. D., 2002, Geologic Map of Washington - Northwest Quadrant.

Dragut, 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.

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., 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.

Gailleton, B., Mudd, S. M., Clubb, F. J., Peifer, D., Hurst, M. D., 2019, A segmentation approach for the reproducible extractyion and quantification of knickpoints from river long profiles, Earth Surface Dynamics, vol. 7, no. 1, p. 211-230.

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.

Gercek, 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.

Grieve, S. W. D., Hales, T. C., Parker, R. N., Mudd, S. M., Clubb, F. J., 2018, Controls on zero-oder basin morphology, Journal of Geophysical Research: Earth Surface, vol. 123, no. 12, p. 3269-3291.

Grieve, S. W. D., Mudd, S. M., Hurst, M. D., 2016, How long is a hillslope?, Earth Surface Processes and Landforms, vol. 41, p. 1039-1054.

Guilbert, E., and Moulin, B. 2017, Towards a Common Framework for the Identification of Landforms on Terrain Models. *ISPRS International Journal of Geo-Information*, 6(1), 12. https://doi.org/10.3390/ijgi6010012

Hattanji, T., Kodama, R., Takahashi, D., Tanaka, Y., Doshida, S., 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.

Hyde, K. D., Wilcox, A. C., Jencso, K., 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., 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., 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., Bai, R., 2020, A D8-compatible high-efficient channel head recognition method: Environmental Modelling & Software, v. 125.

Lin, S., Xie, J., Deng, J., Qi, M., 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.

Mazzeo, N., Ciganda, A. L., Nion, C. F., Peñas, F. J., González-Ferreras, A. M., Crisci, C., Zurbriggen, C. P, Pérez, D., Barquin, J., Diaz, I., 2024, Inter and transdisciplinarity strategies for evaluating and improving water quality monitoring systems: Uruguay as a study case, Environmental Science & Policy, v. 154.

McCleary, R. J., Hassan, M. A., Miller, D., 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., 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., 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. and Burnett, K. M., 2008, A probabilistic model of debris-flow delivery to stream channels, demonstrated for the Coast Range of Oregon, USA, 2008, Geomorphology vol. 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., 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.

Passalacqua, P., Trung, T. D., Foufoula-Georgiou, E., Sapiro, G., Dietrich, W. E., 2010, A geometric framework for channel network extraction from lidar. Nonlinear diffusion and geodesic paths, Journal of Geophysical Research, vol. 115,

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

Pazzaglia, F.J., and 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.

Pelletier, J. D., 2013, A robust, two-parameter method for the extraction of drainage networks from high-resolution digital elevation medels (DEMS): Evaluation using synthetic and real-world DEMS, Water Resources Research, vol. 49

Petras, V., Petrasova, A., McCarter, J. B., Mitasova, H., 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., 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., 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., Guðmundsson, S., 2015, A semi-automated method for mapping glacial geomorphology tested at Breiðamerkurjökull, Iceland: Remote Sensing of Environment, v. 163, p. 80-90.

Rodriguez-Castillo, T., Estéveez, E., González-Ferreras, A. M., Barquin, J., 2018. Estimating Ecosystem Metabolism to Entir River Networks, Ecosystems, v. 22, no. 4, p. 892-911.

Roering, J. J., Marshall, J., Booth, A. M., Mort, M., 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., 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., 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., 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., 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., 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., 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.

Struble, W. T., and Roering, J. J., 2021, Hilltop curvature as a proxy for erosion rate: wavelets enable rapid computation and reveal systematic underestimation, Earth Surface Dynamics, vol. 9, no. 5, p. 1279-1300.

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., and 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., 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., 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., Schasse, H. W., 1987, Geologic Map of Washington - Southwest Quadrant: Washington Department of Natural Resources.
Wang, S., Wu, Q., 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., 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.
Witharana, C. 2014. Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2 (ED2). ISPRS Journal of Photogrammetry and Remote Sensing. 10.1016/j.isprsjprs.2013.11.006.
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., 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., 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.

Zangana, I., Bell, R., Drăguţ, L., Sîrbu, F., Schrott, L. 2025, Inventory mapping of forest-covered landslides using Geographic Object-Based Image Analysis (GEOBIA), Jena region, Germany. EGU, Vienna, Austria, https://doi.org/10.5194/egusphere-egu24-10760.

Zhang, H., Loáiciga, H. A., Feng, L., He, J., 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.

Zheng, X., Tarboton, D. G., Maidment, D. R., Liu, Y., Passalacqua, P., 2018, River channel geometry and rating curve estimation using height above the nearest drainage, Journal of the American Water Resources Association, vol. 54, no. 4, p. 785-806.

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)

With the OBIA method we use, precisification, a logical process of making a vague concept, landforms, clearer, by defining distinct boundaries where none existed before, for example the outer extents of a bedrock hollow. To be implementable in a model the defining characteristic of a landform and the surrounding region need to be expressed by geometries that can be extracted from a terrain model (Guilbert, 2017). The defining RIL characteristics used in the OBIA model were drawn out of Board Manual descriptions and rephrased where appropriate to reflect model terminology (Table A1). Iteration through segmentation and classification routines allows us to add specificity to vague landform criteria by describing the mathematical process used to determine, for instance, where the uppermost extent of an inner gorge is located, i.e., break in slope. Building thresholds into the model allows us to describe the mathematical process quantitatively. It is anticipated that the thresholds could vary based on context and/or formative geomorphic process that exist in different landscapes.

Table A1. 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

The primary data source for the OBIA routine is a lidar DTM. 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. It is recognized that the use of more sophisticated smoothing and/or interpolation algorithms could potentially allow the use of smaller cell sizes. More effective representation of relevant micro-topography could produce a bare earth surface that more closely aligns with what practitioners encounter in the field and are able to use during the interpretation process (Roering et al., 2010).

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) and (b) maximum curvature. Horizontal curvature is a gravity field-specific measurement of flow convergence and divergence where flows are converged when values are less than zero and diverged when values are greater than zero. Horizontal curvature, unlike planform curvature (curvature of a smooth contour line curve), is a measurement of the intersection of a plane with the topographic surface. Both horizontal and planform curvature were calculated during the data exploration process. More noise was observed in the planform dataset and no detectable patterns of potential benefit for the segmentation/classification process were noted so horizontal curvature was used in the model. Vertical curvature, measure of acceleration or deceleration of gravity driven flows, were also explored with no detectable patterns of potential

benefits. Maximum curvature measures extreme bending at a point on the surface and is not dependent on gravity. Maximum curvature values larger than zero correspond to ridges while values less than zero relate to closed depressions. Other Gaussian classifications describing eliptic or hyperbolic portions of the landscape were not considered as hills, closed depression, saddles, and plains are not included in a targeted definition of bedrock hollow or inner gorge landforms. 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 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, is an iterative segmentation and classification process that allows 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.

Table A2. Image analysis routine to classify inner gorges.

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 begins with a multi-threshold segmentation algorithm, which segments and classifies an image into objects based on input data values - in this case the -difference-to-neighbor (DTN) image derived from the DEM. The mean difference-to-neighbor feature, which the DTN image was created from, quantifies the difference between an image object/pixel value and its surrounding neighbors. 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 channel incision given the potential energy of the streams in the test basins. This threshold value was held constant for all landscapes examined in this study. 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. 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. 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 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, measured as the sum of the pixel edge lengths, 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. *FAC: Flow Accumulation

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 bedrock hollow 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 upstream 4m^2 cells or 200 m^2), and a relatively low border, measured as the sum of the pixel edge lengths, 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, 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).

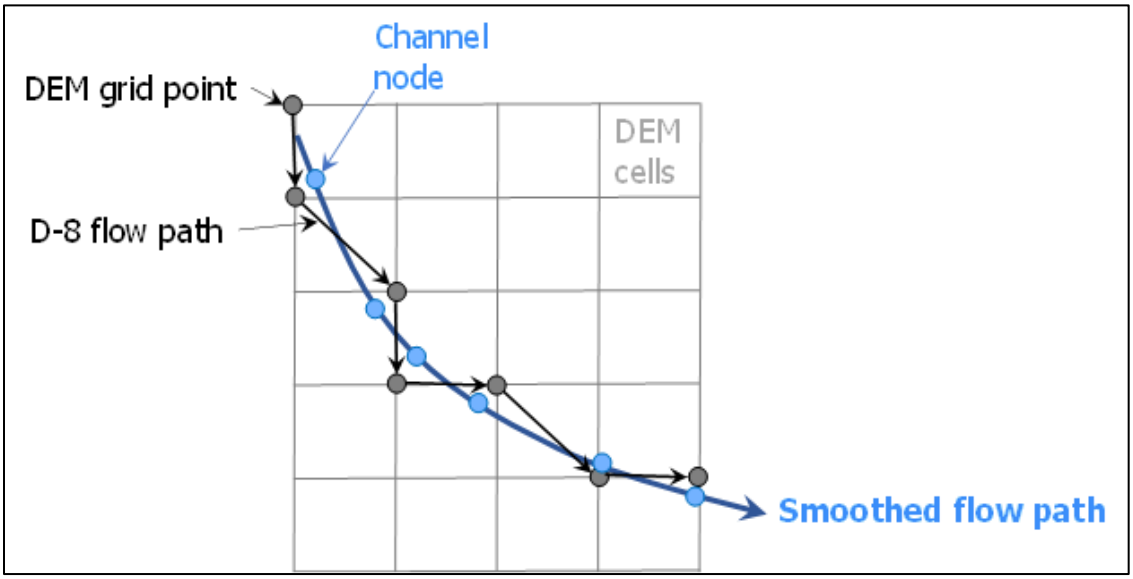


Figure B1. Channel centerlines are represented in a virtual watershed using a linked node data structure. Once the channel initiation criteria are met along any flow path, D8 flow directions are used to preclude dispersion and a node is created for each DTM grid point along the channelized flow path. 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 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"

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).

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 by tracking the amount of offset from the calculated aspect cell by cell along flow paths and shifting the downslope receiving cell as needed to keep the offset less than one cell width.

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 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).

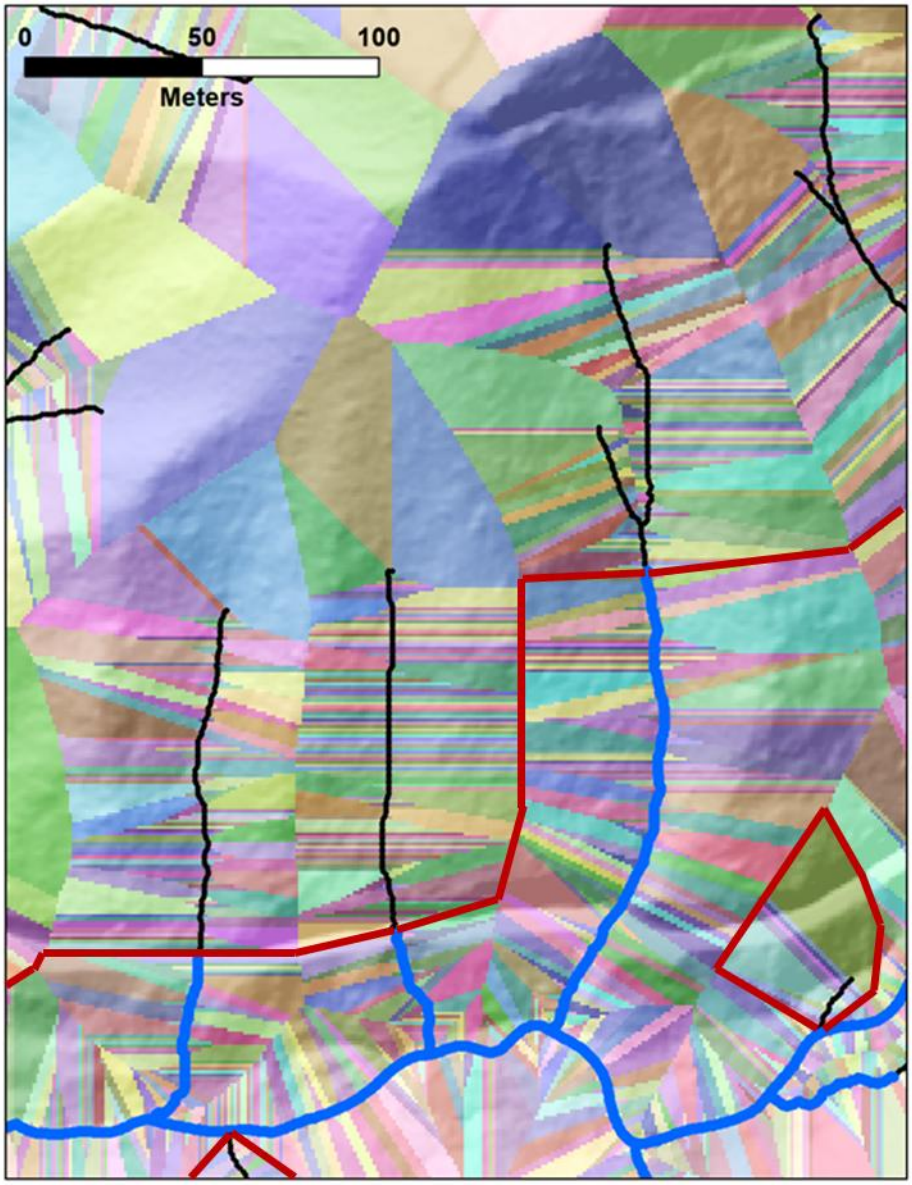


Figure B3. Closest node; each colored patch indicates the ID of the closest channel node, measured as horizontal Euclidean distance. Blue lines show traced channels; black lines show extension of the channel network into zero-order basins. Inner gorges are constrained to closest-node patches associated with blue-line channels, the thick dark red lines indicate the extent of blue-line-channel closest-node patches.

b. Examine each DTM cell. 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. Constraining the inner-gorge zones to closest-node patches associated with channels (the blue lines in Figure B3) prevents extension of the inner-gorge polygons into the zero-order basins.

c. The next three sub steps are for aesthetics 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.

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.

a. Examine each DTM cell . 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.

b. 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.

c. The following steps were 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.

B1. THRESHOLD VALUES

Table B1. Steps in generating 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, Montgomery and Foufoula-Georgiou, 1993) ~ $30m^2$, 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	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

Task	Description	Parameters
	only allowed within zero order basins and inner gorges only allowed downstream of the channel head.	bedrock hollow RIL polygons assuming, as per Board Manual, 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	This is calculated using a smooth polynomial surface fit	Length scale over which gradient is measured; 15 m to match

Task	Description	Parameters
gradient at each DTM cell	to 9 points, one at the DTM grid point and 8 on the 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.	typical sight distance on a vegetated slope.
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 thresholds	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		Minimum hole size = 1000m^2, minimum patch size = 50m^2

Task	Description	Parameters
patches; smooth polygon boundaries		
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

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%

APPENDIX C. PHASE 1 MANUAL MAP COMPARISONS

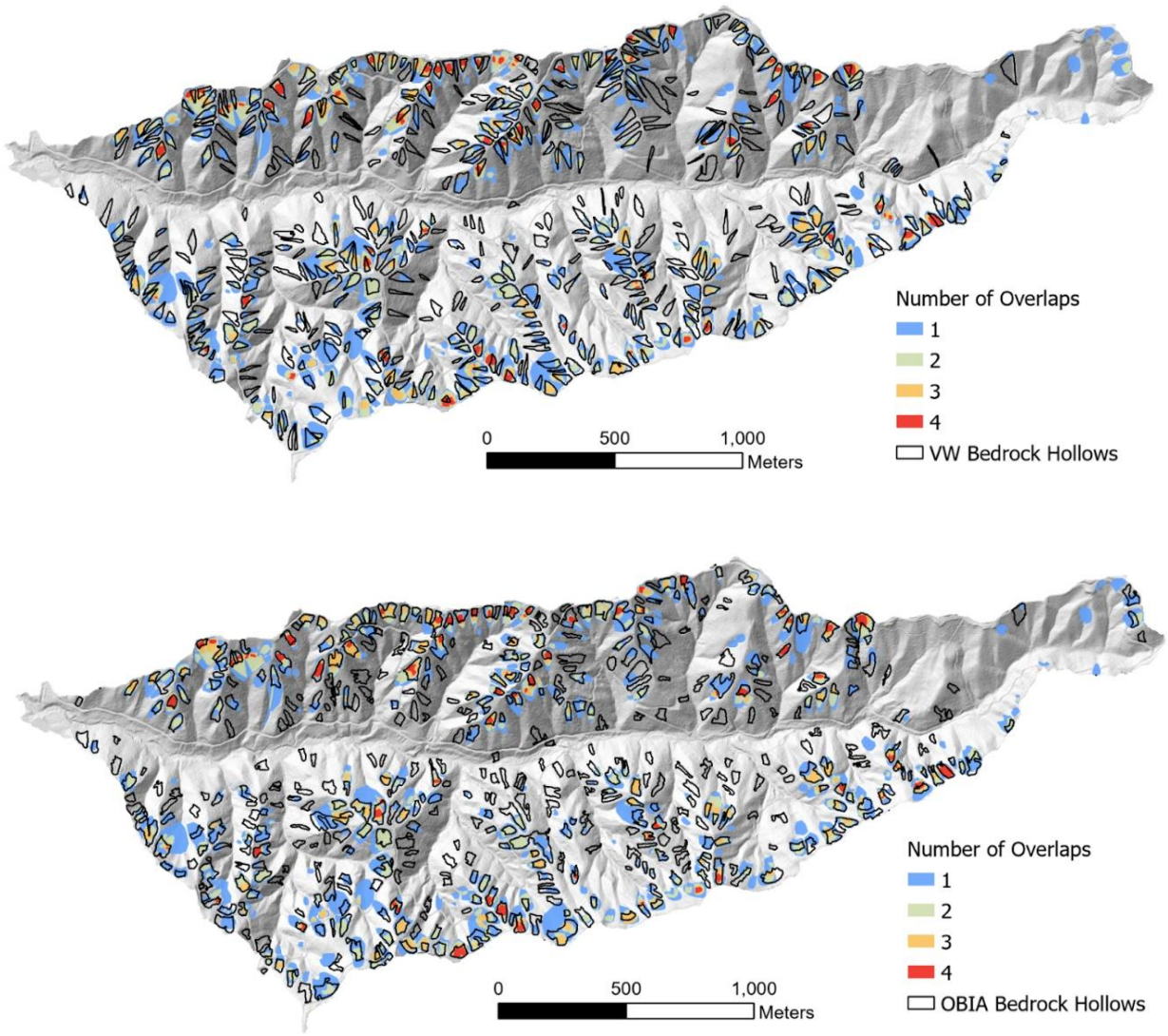


Figure C1: Overlap map of bedrock hollows in Site 1A or North Fork Calawah. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons. Each color represents the number of overlapping polygons in that area.

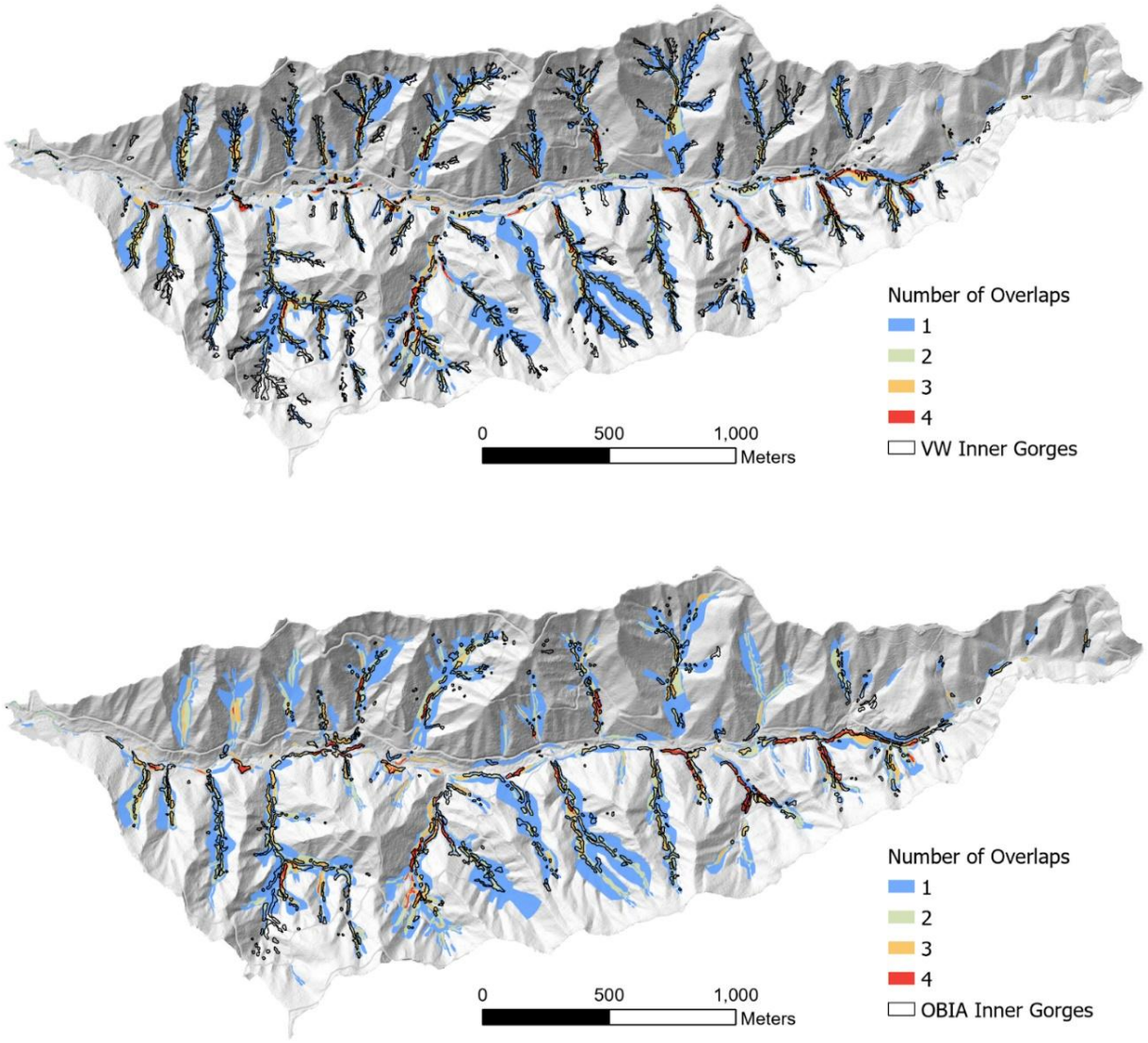


Figure C2: Overlap maps for inner gorges in Site 1A or North Fork Calawah. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

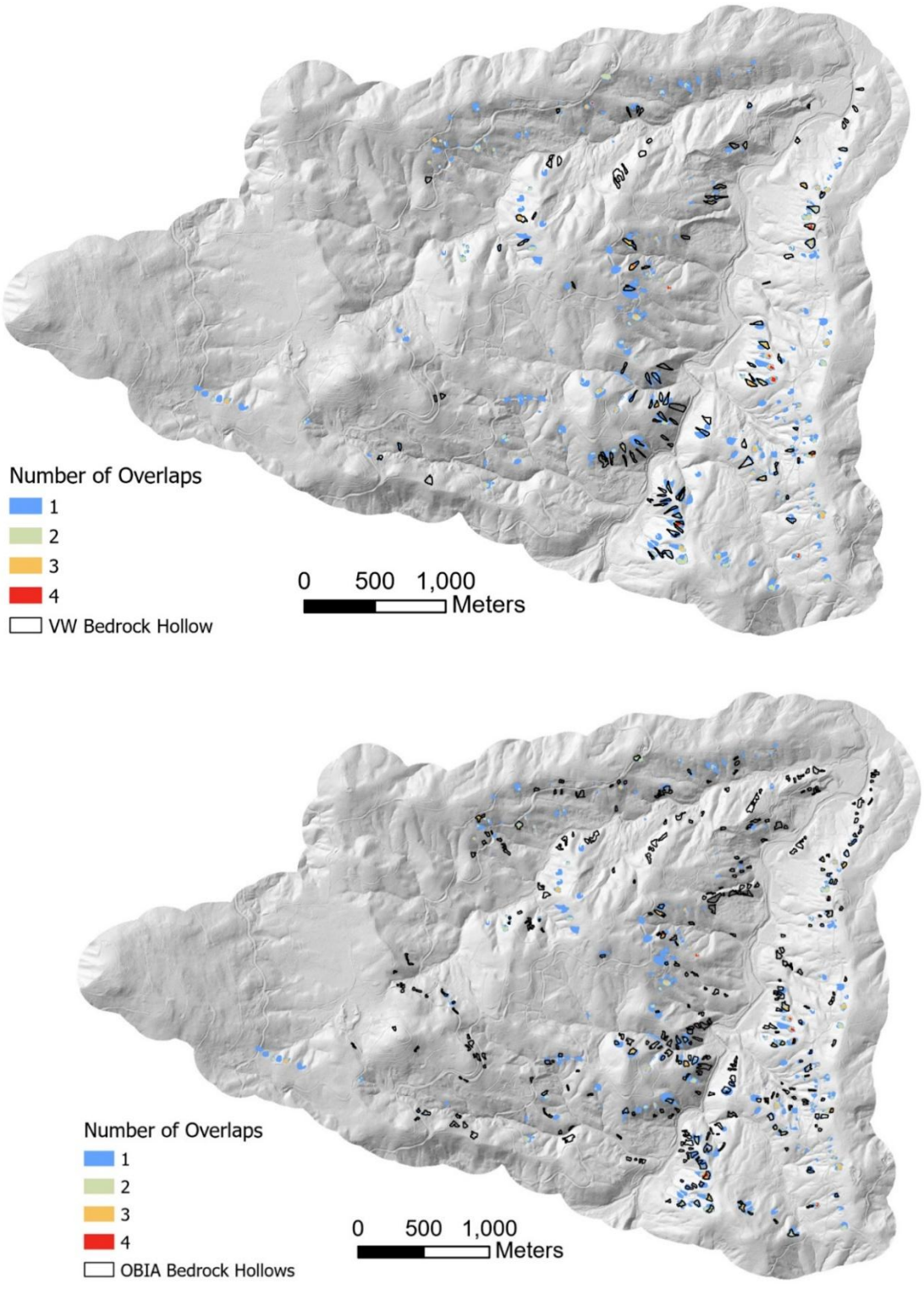


Figure C3: Overlap maps for bedrock hollows in Site 2 or Willapa Hills. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

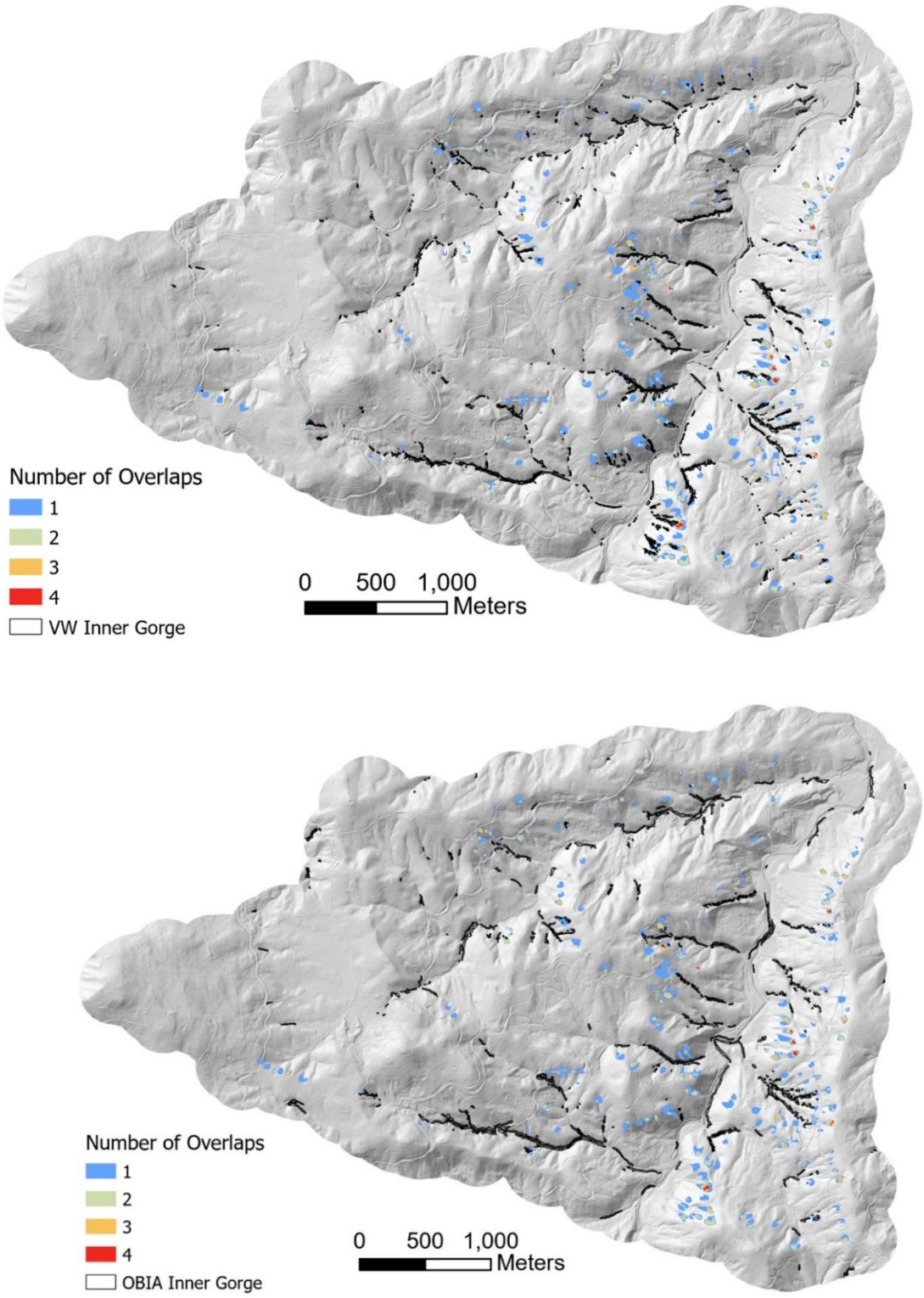


Figure C4: Overlap maps for inner gorges in Site 2 or Willapa Hills. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

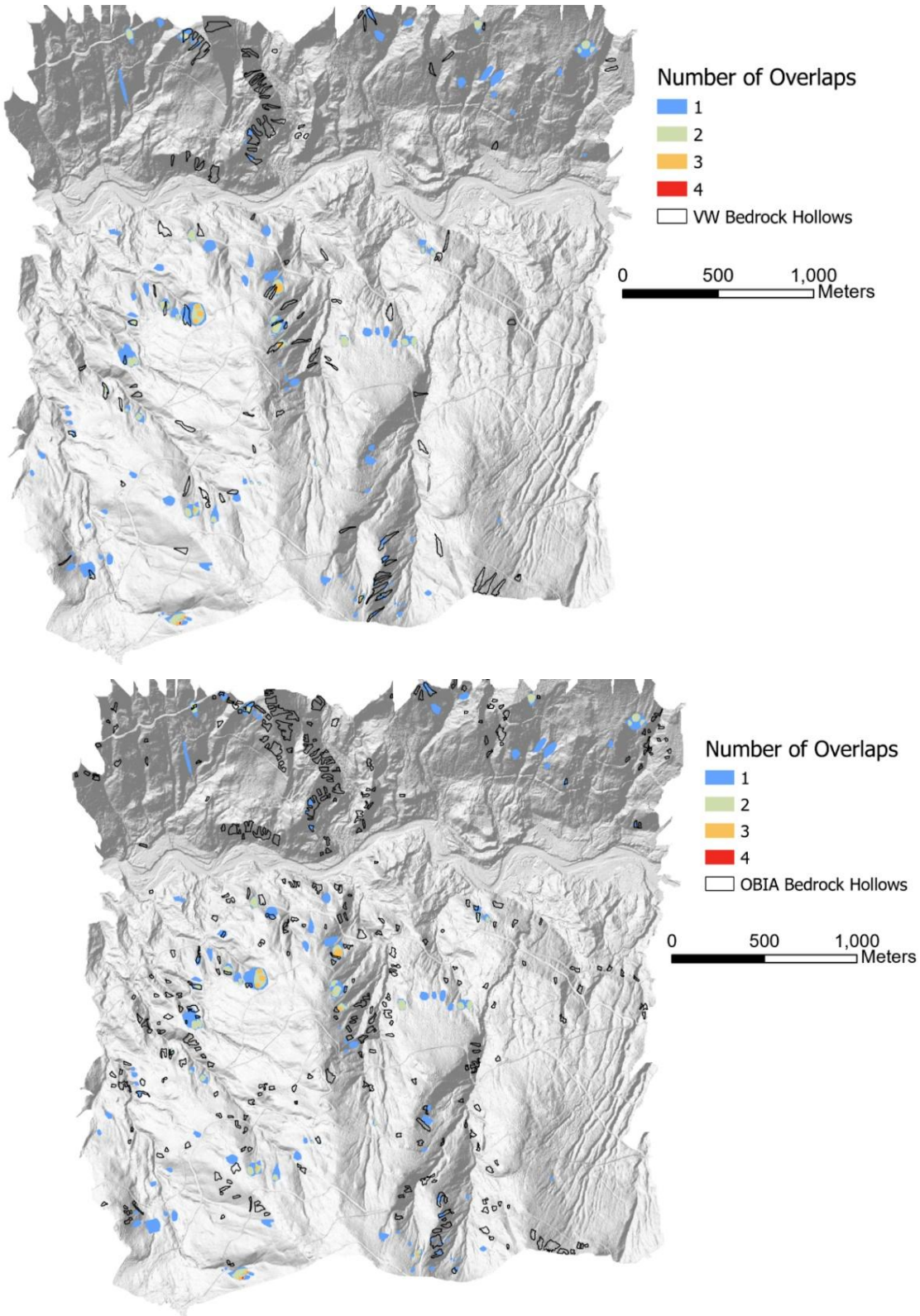


Figure C5: Overlap maps for bedrock hollows in Site 3 or Howard Creek. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

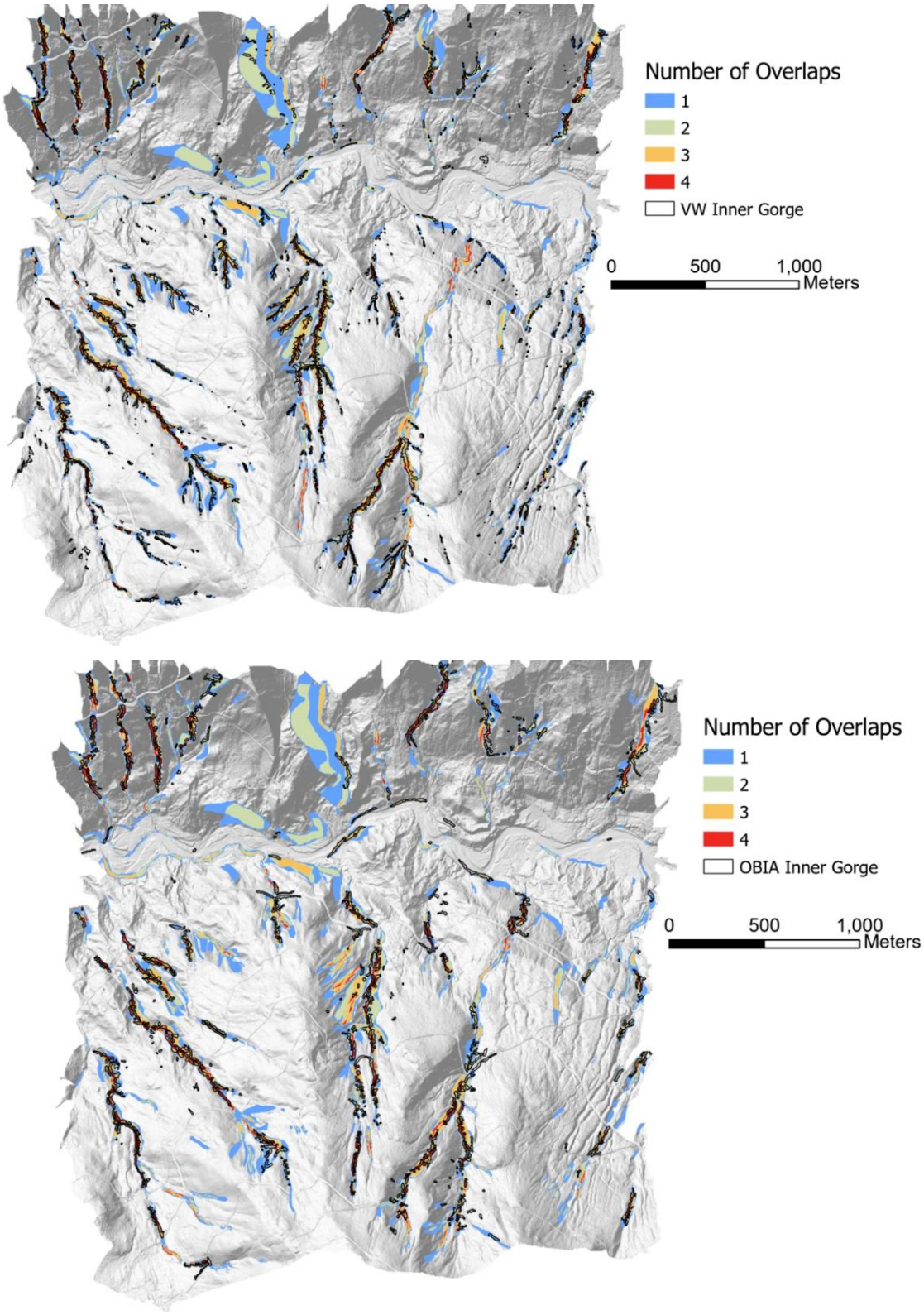


Figure C6: Overlap maps for inner gorges in Site 3 or Howard Creek. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

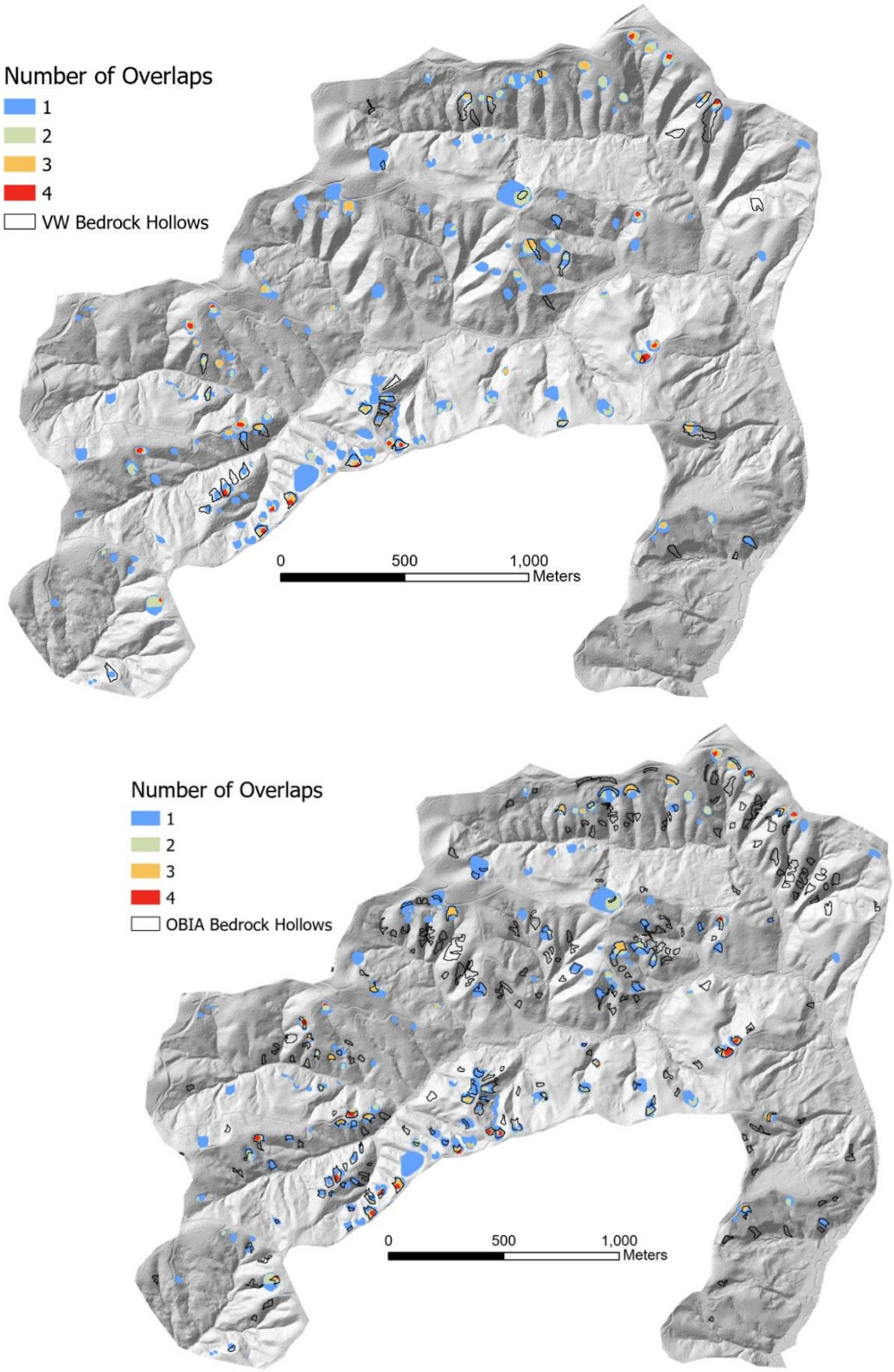


Figure C7: Overlap maps for bedrock hollows in Site 4 or Wishkah. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

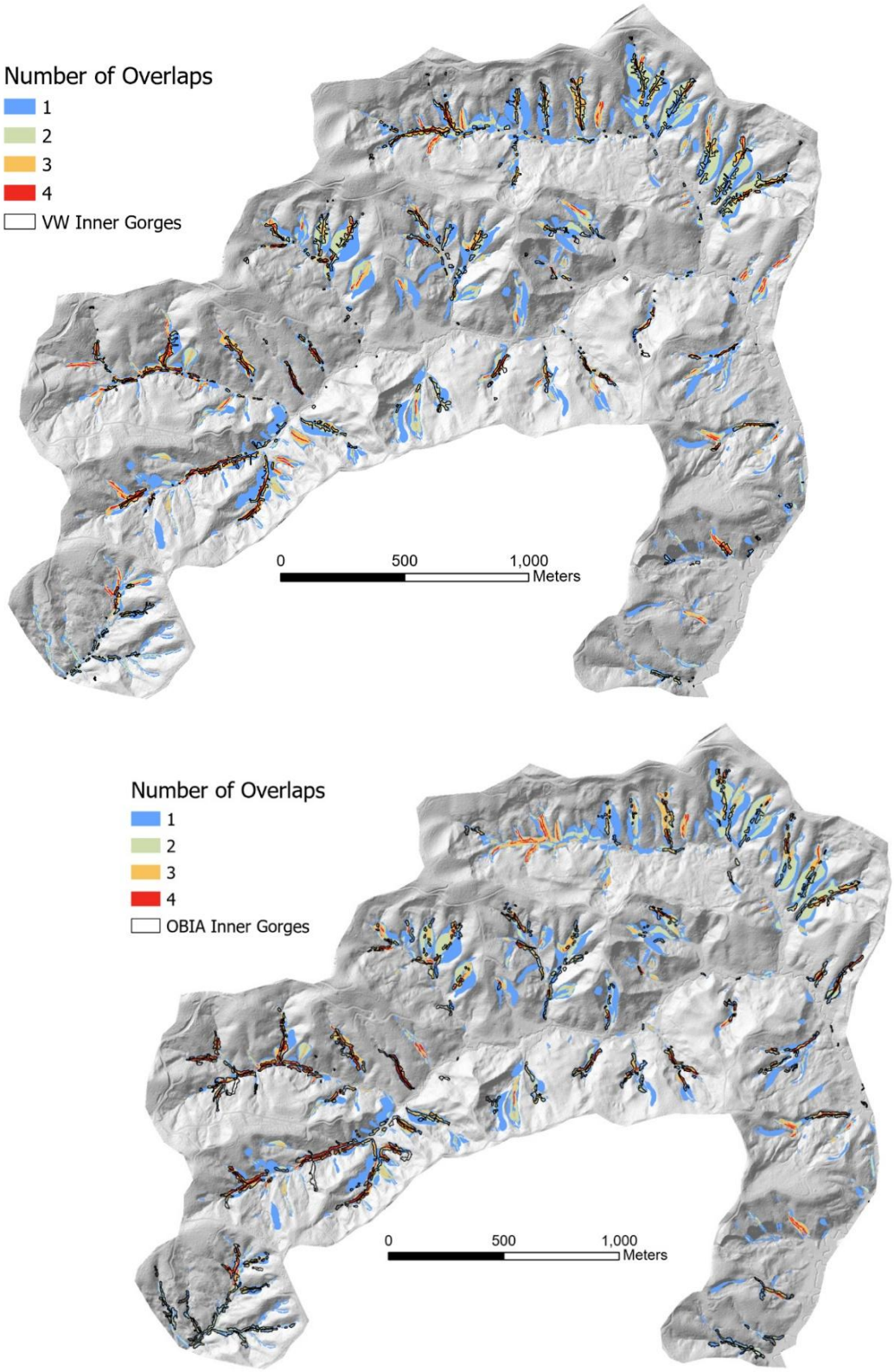


Figure C8: Overlap maps for inner gorges in Site 4 or Wishkah. Each color represents the number of overlapping polygons in that area. In the top panel manual mappers are compared to the VW model generated polygons and in the bottom panel compared to the OBIA generated polygons.

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.

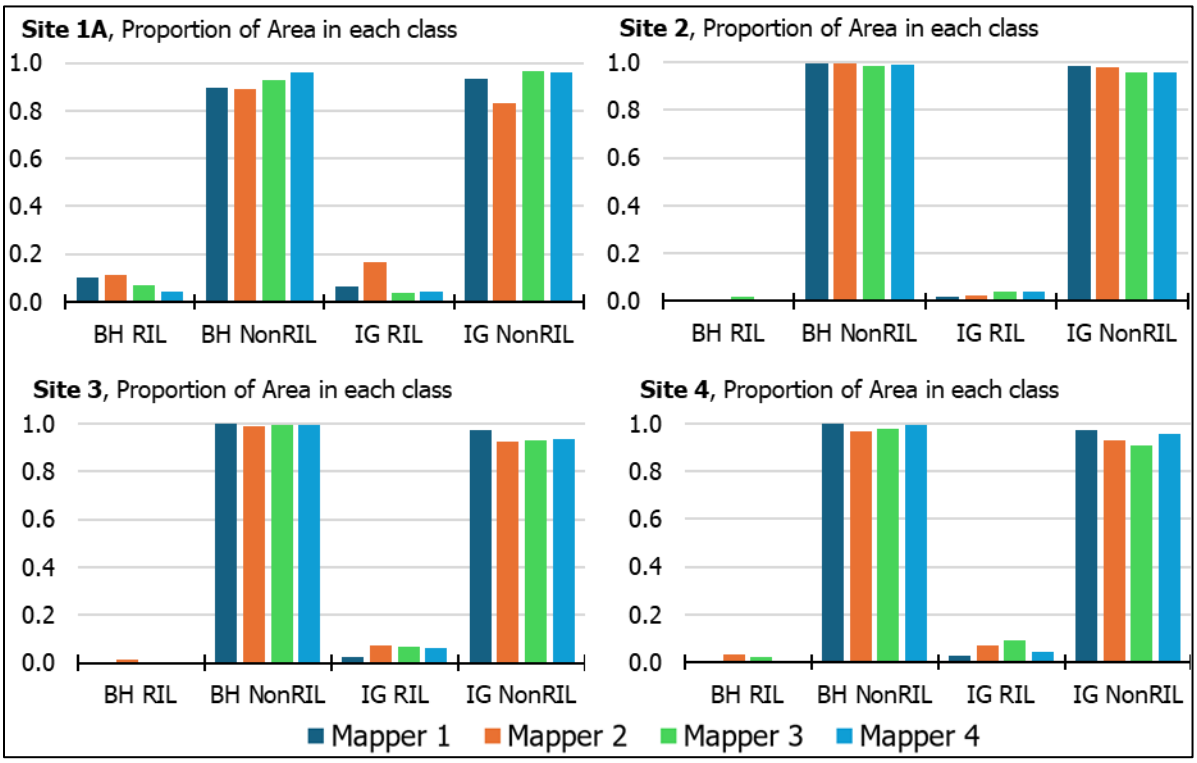


Figure C.9. The proportional area of mapped RIL polygons and NonRIL zones for Bedrock Hollows (BH) and Inner Gorges (IG) for each mapper at each study site. For both Bedrock Hollows and Inner Gorges, the proportion of area mapped as NonRIL is considerably higher than that mapped as RIL. This imbalance in area between the RIL and NonRIL classes complicates measures of accuracy (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

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, the previous measure of recall becomes the precision value, the precision value becomes the recall value, and the balanced accuracy will be slightly different.

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.

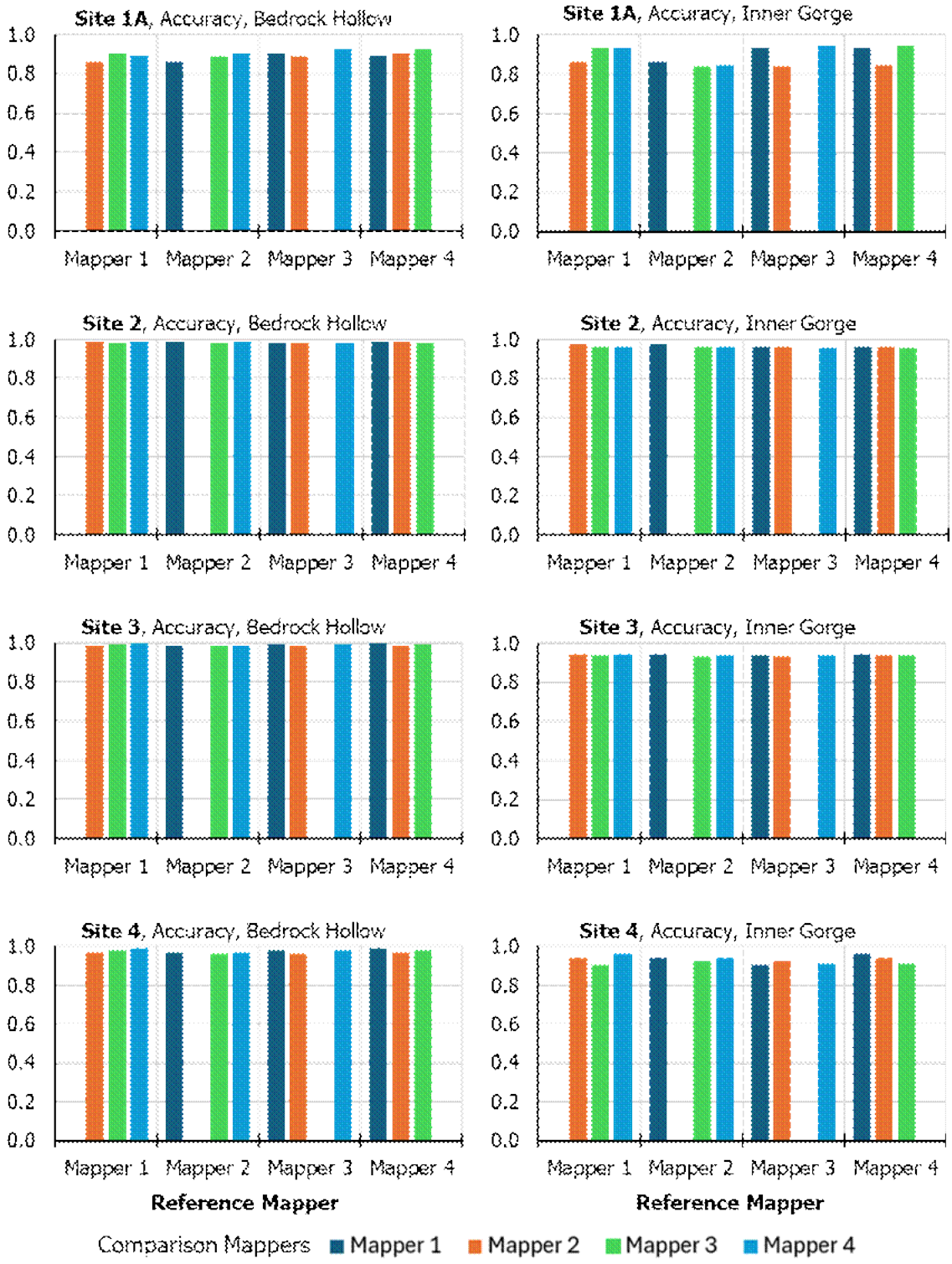


Figure C.10. Comparison across mappers using accuracy. The x-axis shows the reference mapper; the bars refer to the comparison mapper. Because the NonRIL area is much larger than the RIL area, this measure primarily reflects the agreement in NonRIL area between mappers (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

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 (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

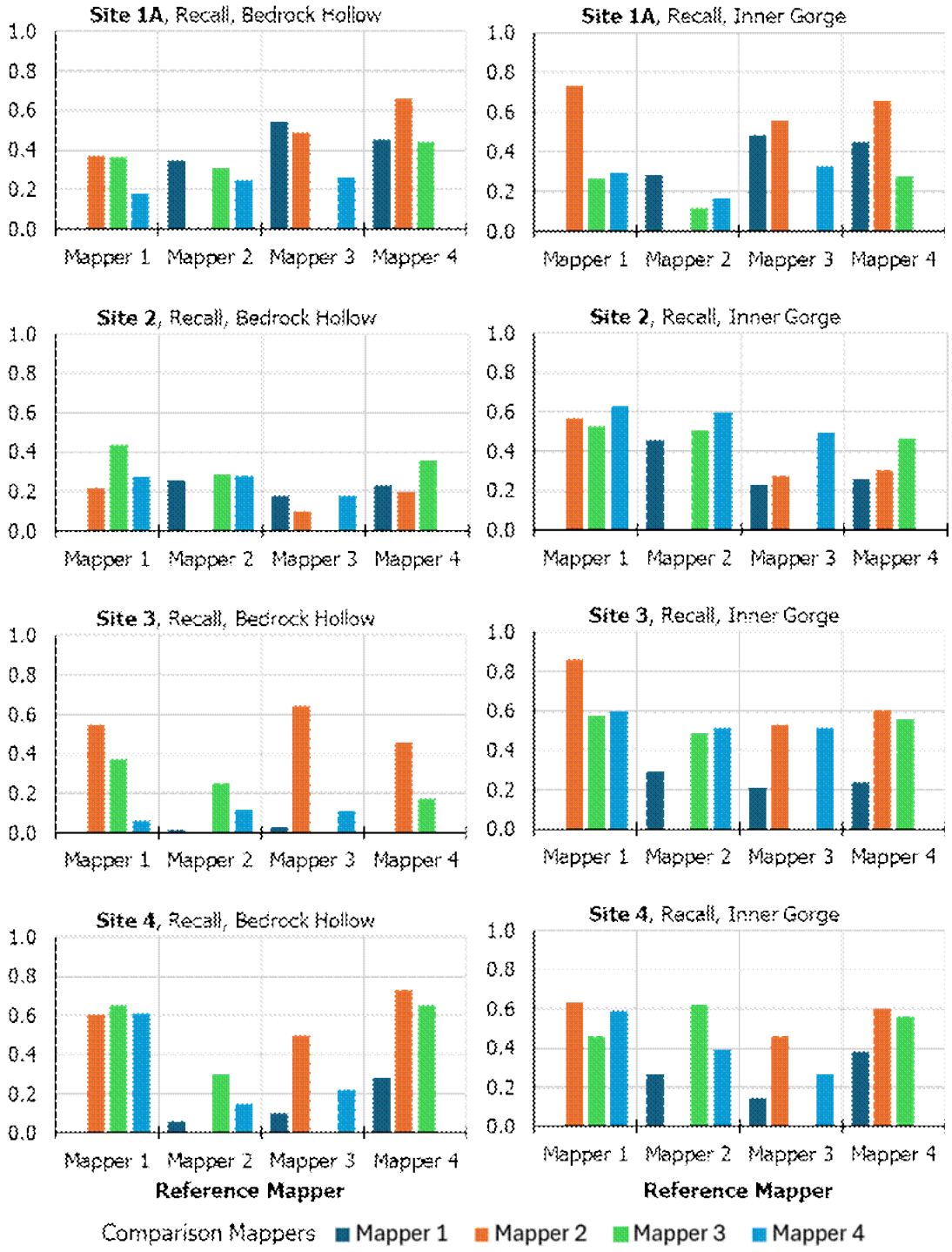


Figure C.12. Comparison across mappers using recall. The x-axis shows the reference mapper; the bars refer to the comparison mapper. Recall indicates the proportion of the reference RIL area matched by the comparison map. If the reference-map RIL polygons are large compared to those of the comparison 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 (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

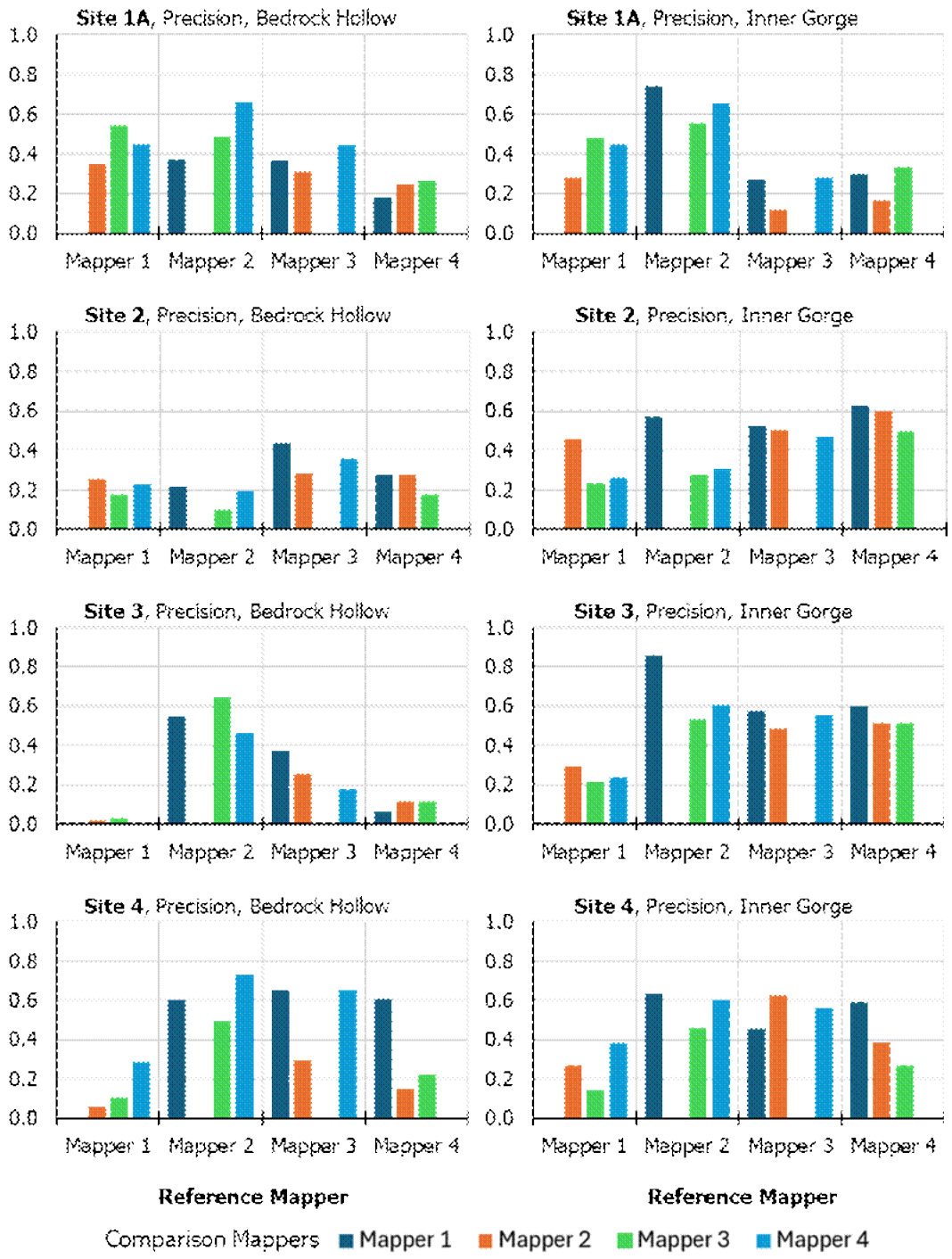


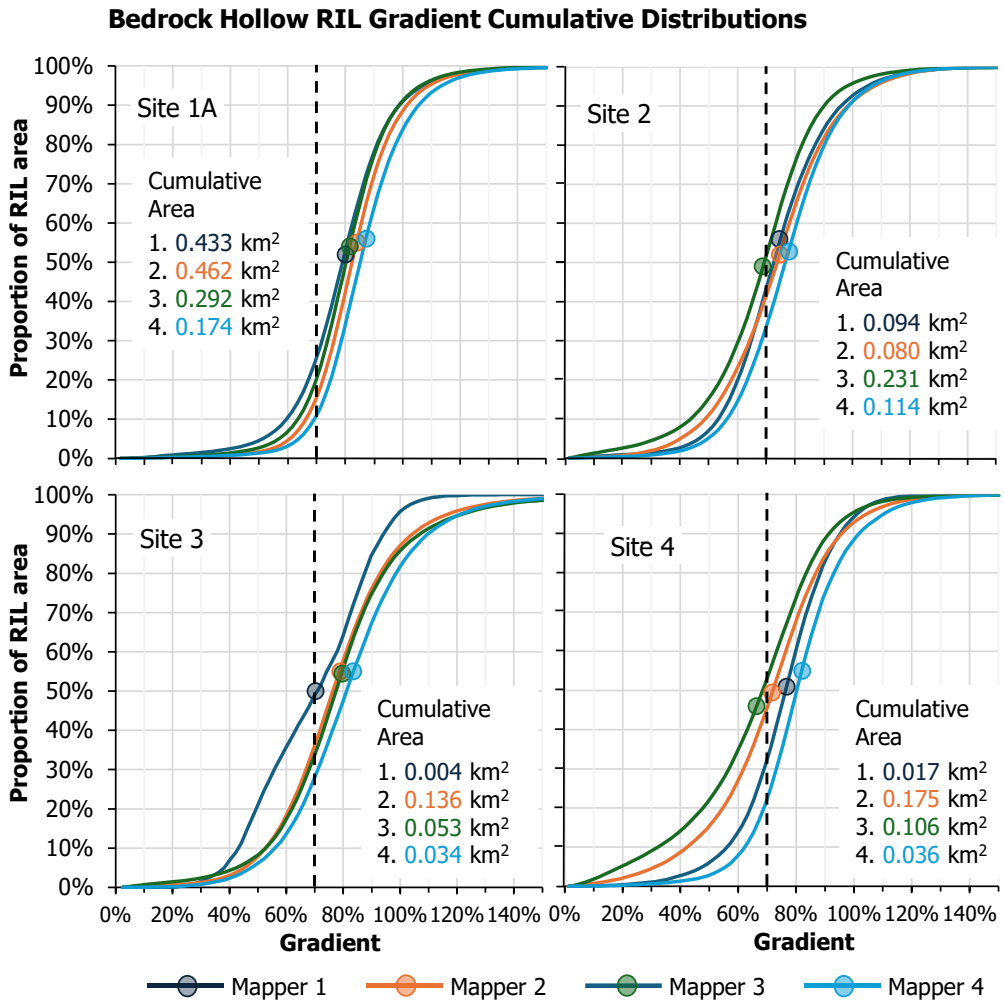
Figure C.13. Comparison across mappers using precision. The x-axis shows the reference mapper; the bars refer to the comparison mapper. Precision indicates the proportion of the comparison RIL area that matches the reference map. If the reference-map RIL polygons are large compared to those of the comparison 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 (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

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 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 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 RIL polygons for each of the mappers at each of the study sites. The cumulative distributions show 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 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 (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

The Board Manual describes hollows as convergent (spoon-shaped) landforms with gradients of 70% or greater. The Board Manual 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 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. 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.

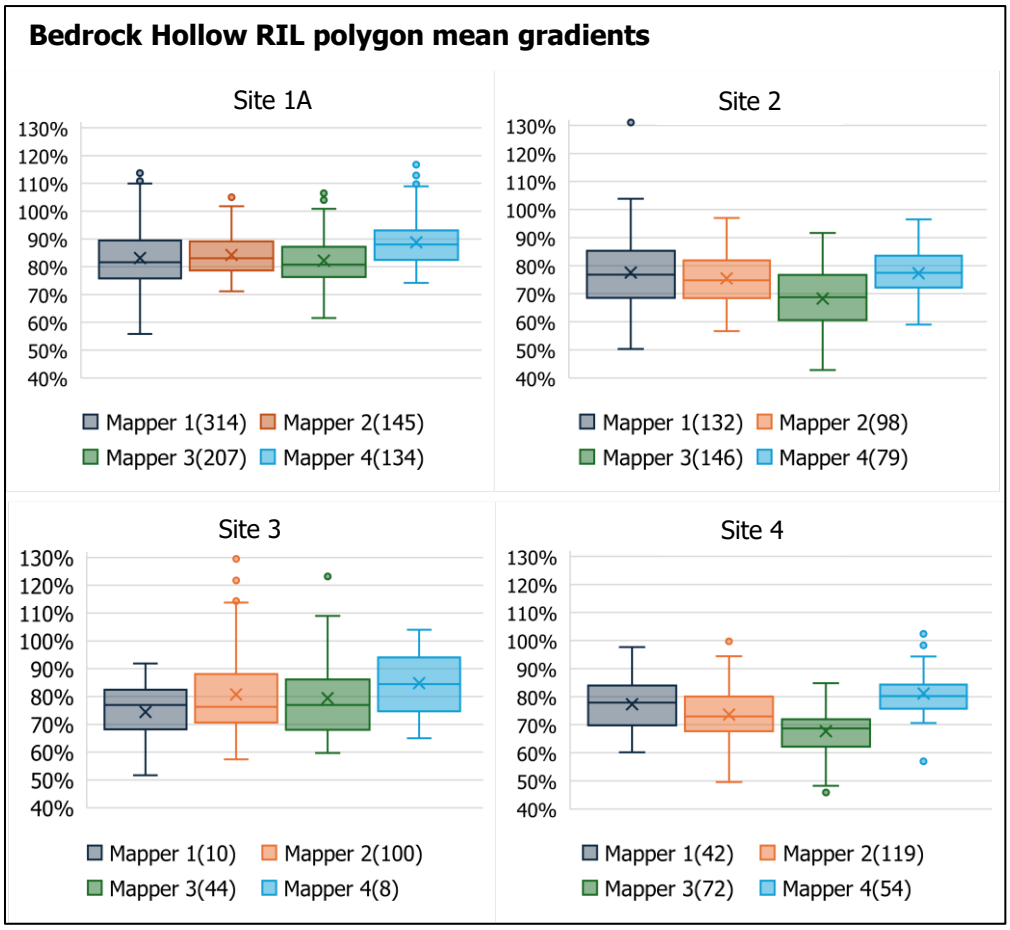


Figure C.15. Box and whisker plots showing distributions of mean gradient for each mapper's set of bedrock hollow RIL polygons for each study site. Boxes show extent of the 2nd to 3rd quartiles, whiskers show lower limit of the 1st quartile and upper limit of the 4th quartile, excluding points that fall outside 1.5 times the inter-quartile range (size of the box), which are plotted as circles (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

Figure C.16 shows cumulative distributions of tangential curvature for each of the mapper's sets of 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.

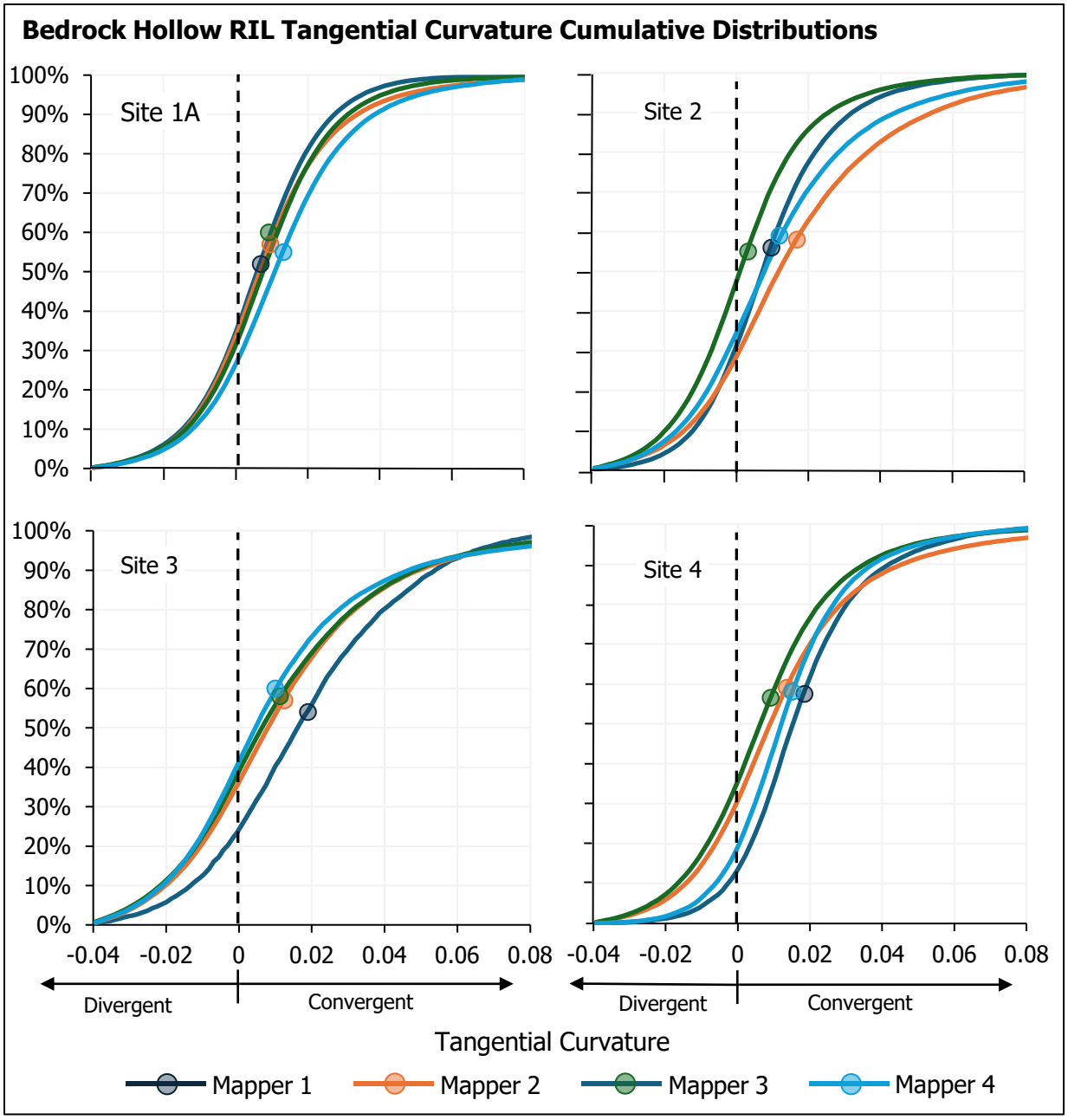


Figure C.16 Cumulative distributions of tangential curvature for the total area from each mapper's set of polygons for each study area. Circles show mean values. Dashed lines indicate zero curvature (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

Figure C.17 shows box and whisker plots for the mean tangential curvature of each polygon in each mapper's sets of bedrock hollow polygons. Similar to gradient, each mapper's polygons exhibit a range of mean values and these ranges differ between mappers.

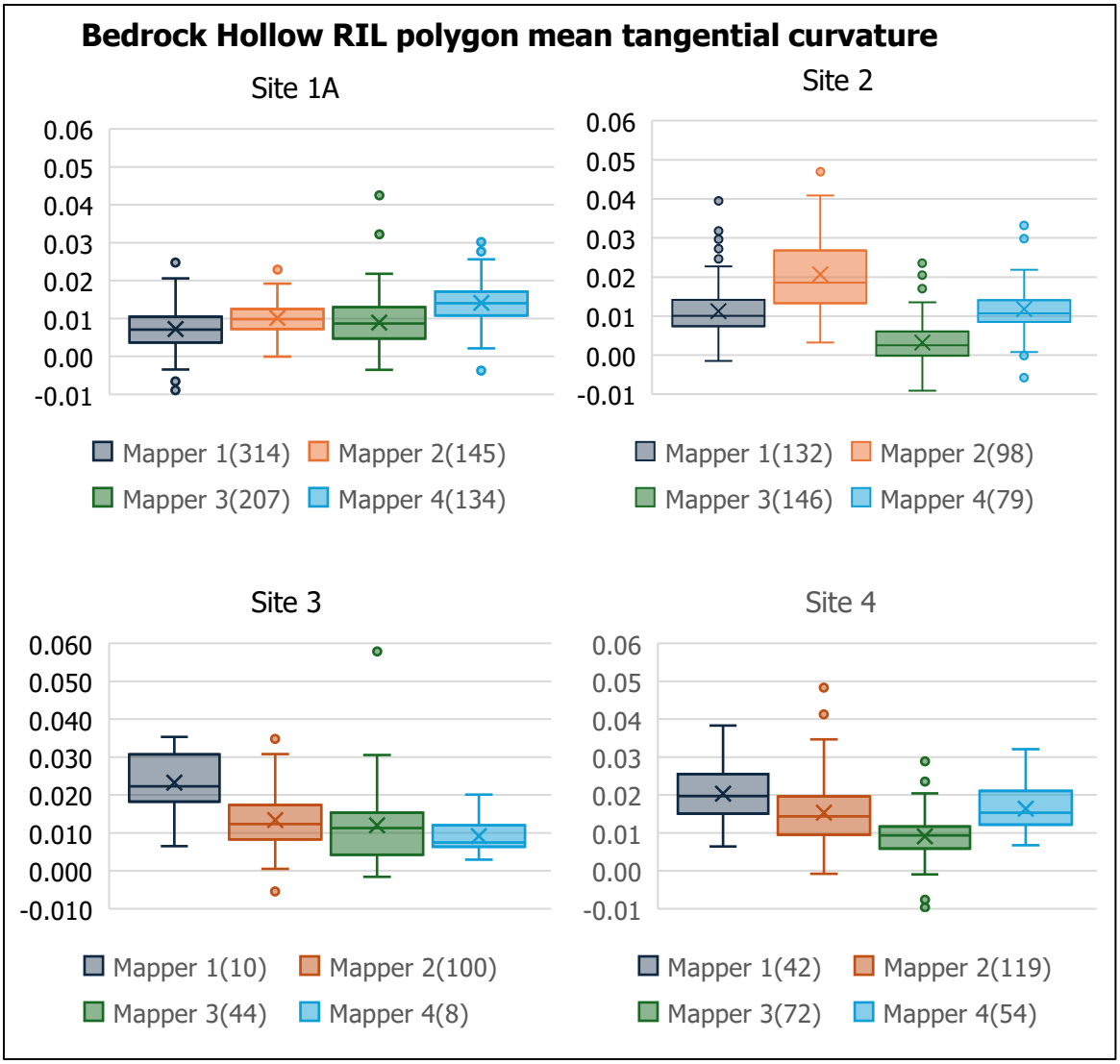


Figure C.17. Box and whisker plots of mean tangential curvature for each mapper's set of bedrock hollow RIL polygons at each study site (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

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[7]. 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.

[7] Given the potential for spatial intermittency of channel formation at the upslope extent of channel formation, such as, associated with large wood or boulders trapping sediment, there may be more than one channel head.

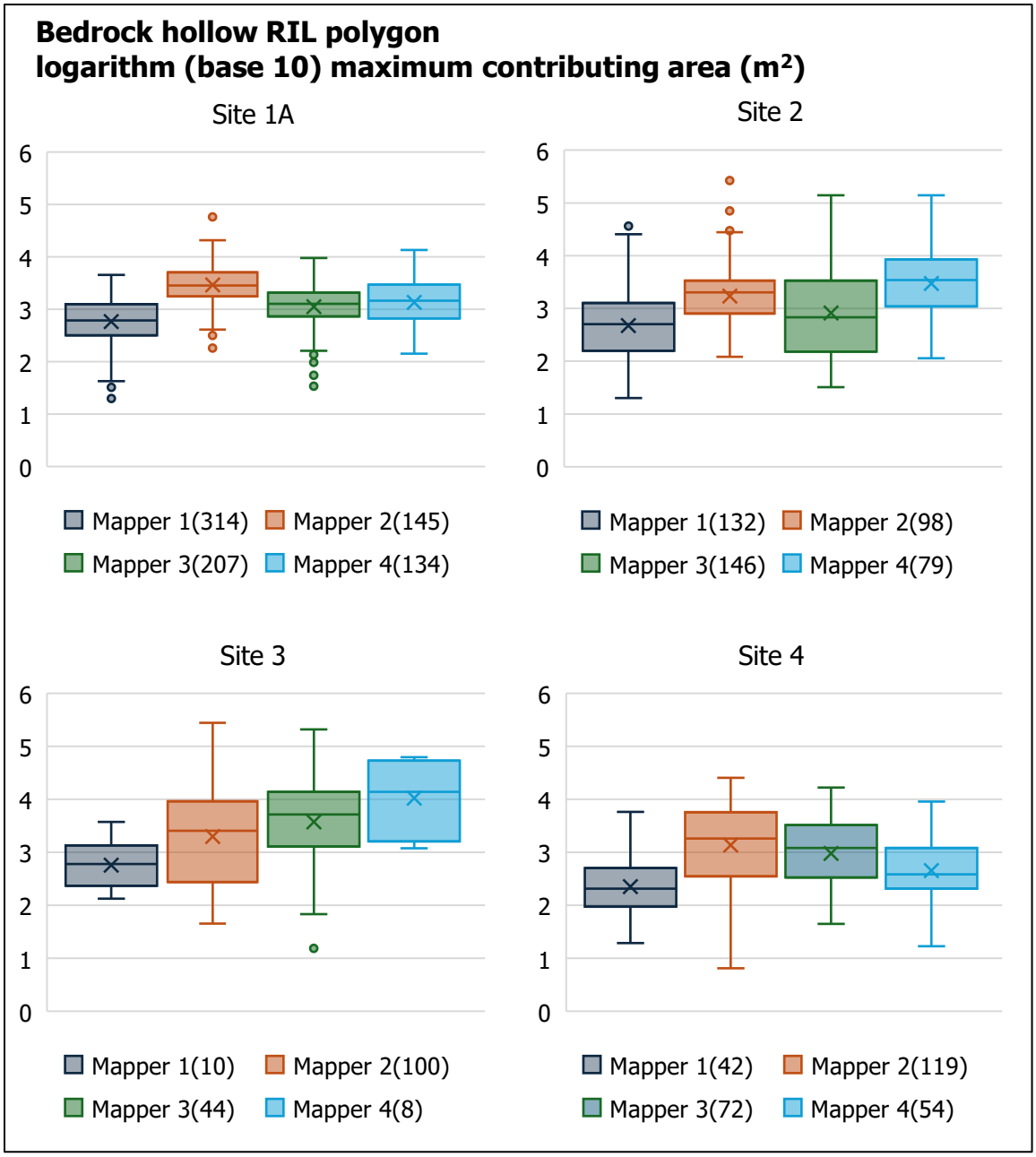


Figure C.18. Box and whisker plots of the base-10 logarithm of maximum contributing area (m²) for each mapper's set of bedrock hollow RIL polygons at each study site. Numbers in parentheses indicate the number of polygons in each mapper's set (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

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

	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

The manual mapping and computer delineation of 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 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 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 (Grieve et al., 2016; Struble & Roering, 2021).

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 (e.g., Zheng et al., 2018). 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).

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 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 and inner gorge landforms, a mapper must decide precisely where the landform criteria are met and draw a line delineating those areas inside and outside of each landform. 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.

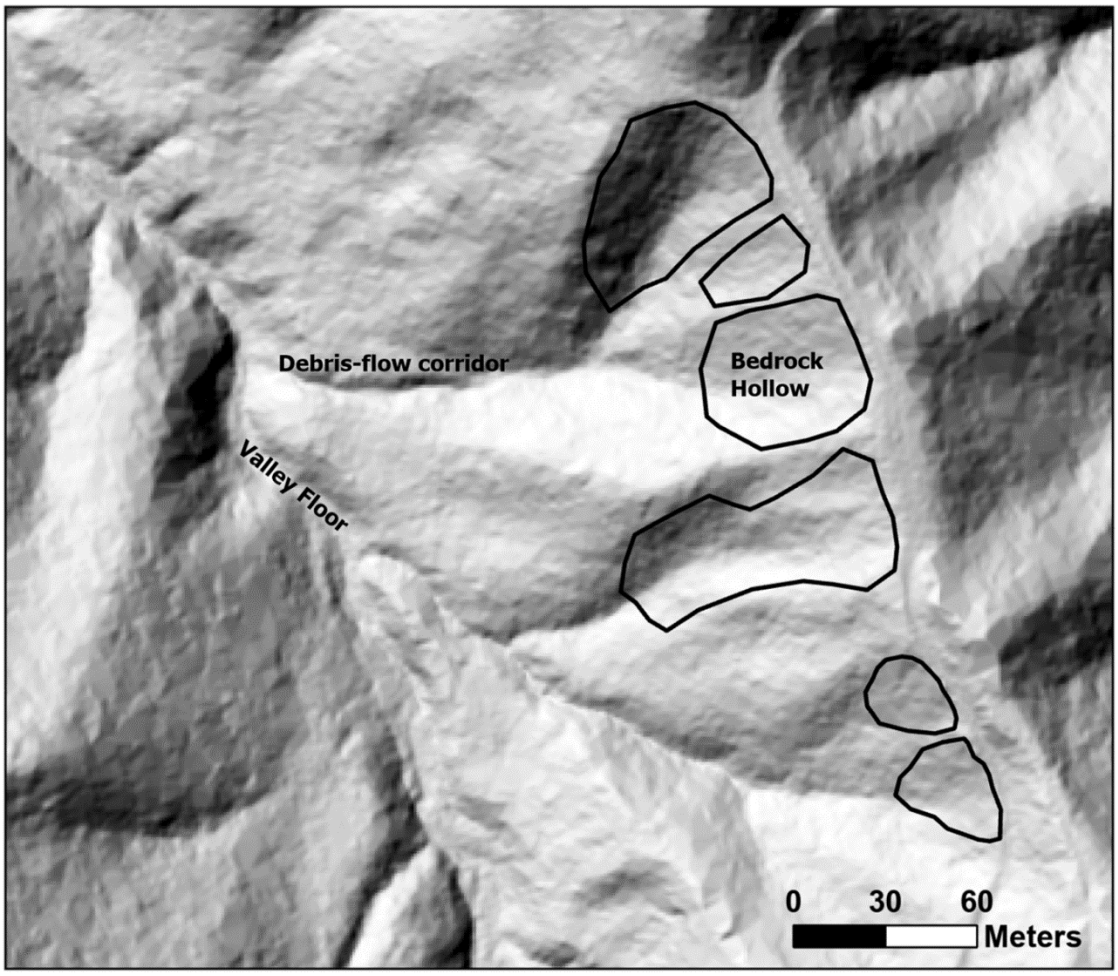


Figure E.1. A shaded relief image from Site 1A showing several landscape features. The valley floor extends diagonally from the southeast to northwest across the figure. A series of mapped bedrock hollow RILs lie on the upper portion of the hillslope.

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 landform, but one mapper might extend their landform 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 using DTM base maps. These distributions are not 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 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 (See Figures in Appendix F & G).

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[8]) 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.

[8] 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.

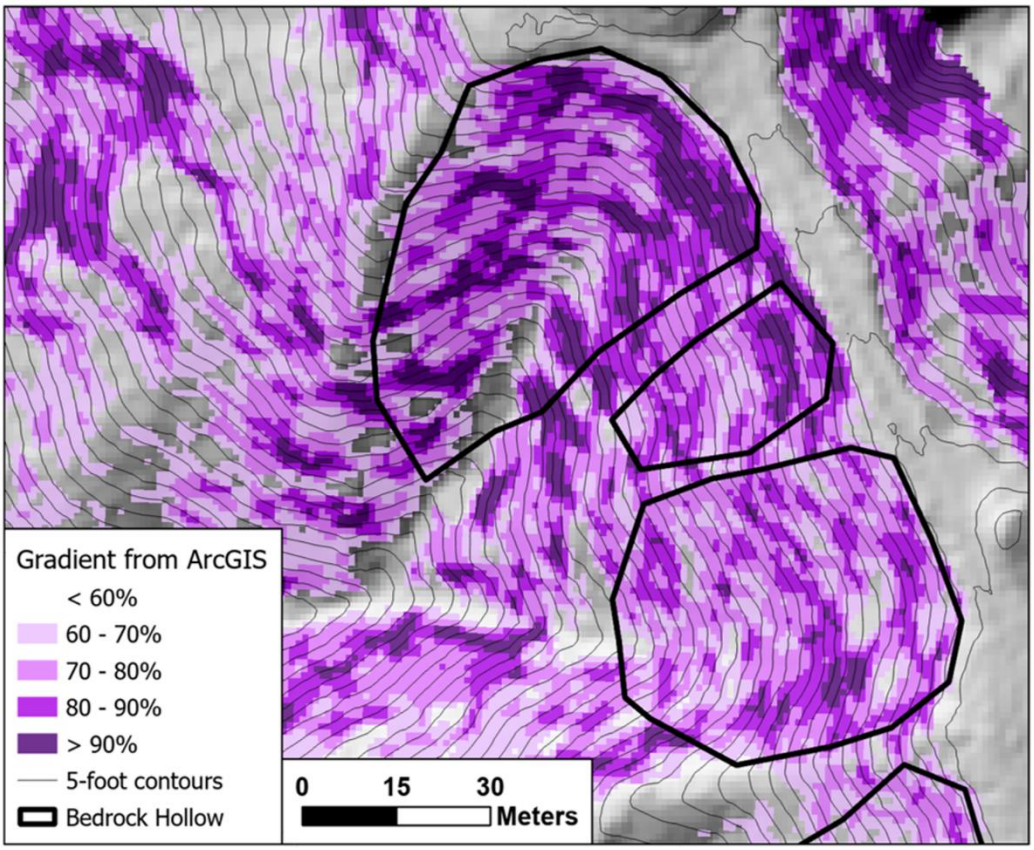


Figure E.2. Gradient classes, contour lines, and one mapper's bedrock hollow RIL polygons overlain on the shaded-relief image.

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 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.

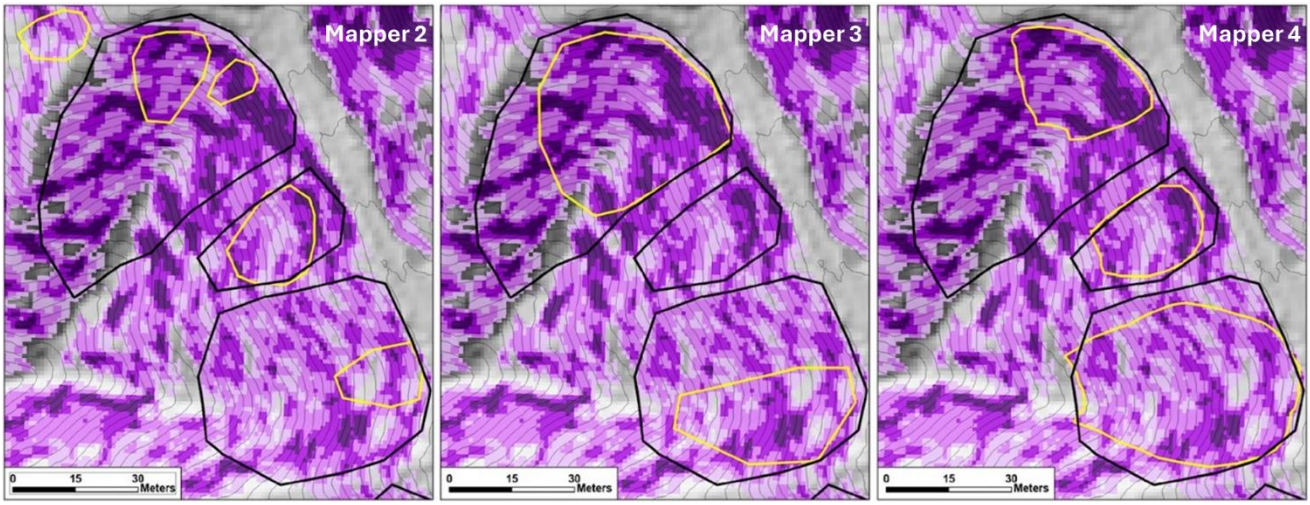


Figure E.3. Three other mappers' polygons (yellow lines) overlain on the original three polygons.

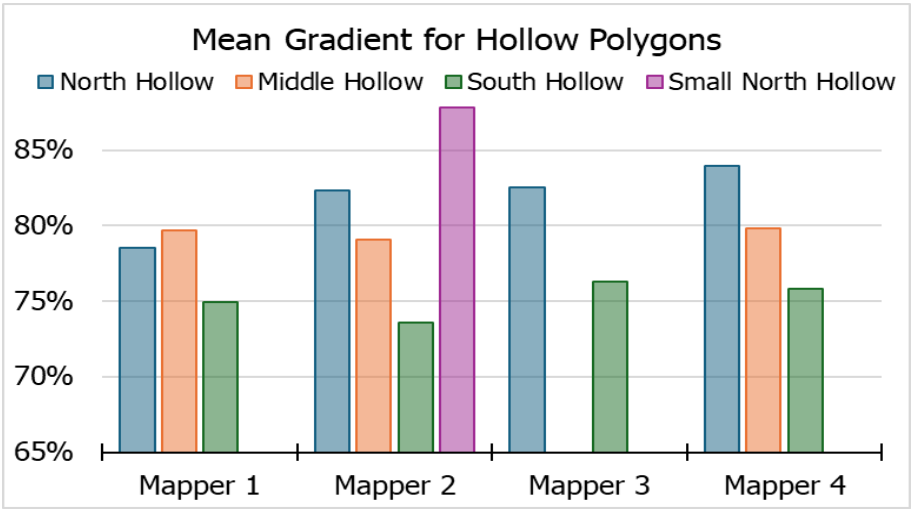


Figure E.4. Bar chart of mean gradient within each polygon.

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.

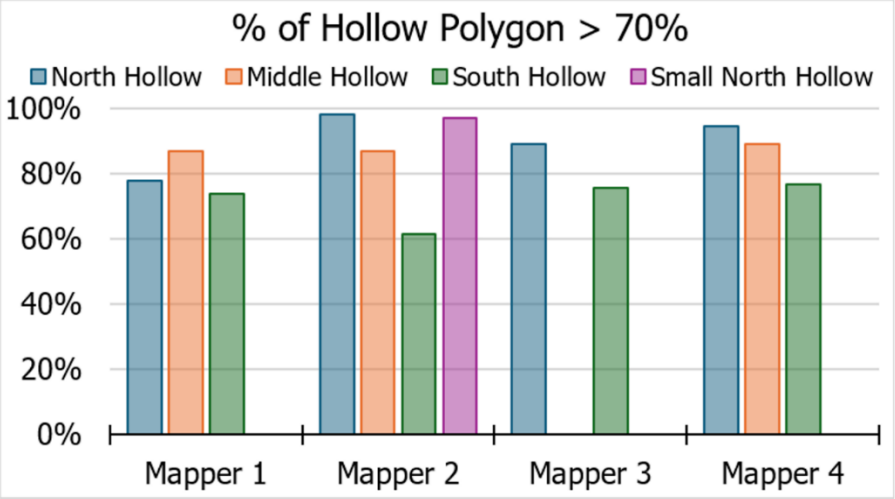


Figure E.5. Proportion of each mapped hollow with gradient exceeding 70%.

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[9]; 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.

[9] 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.

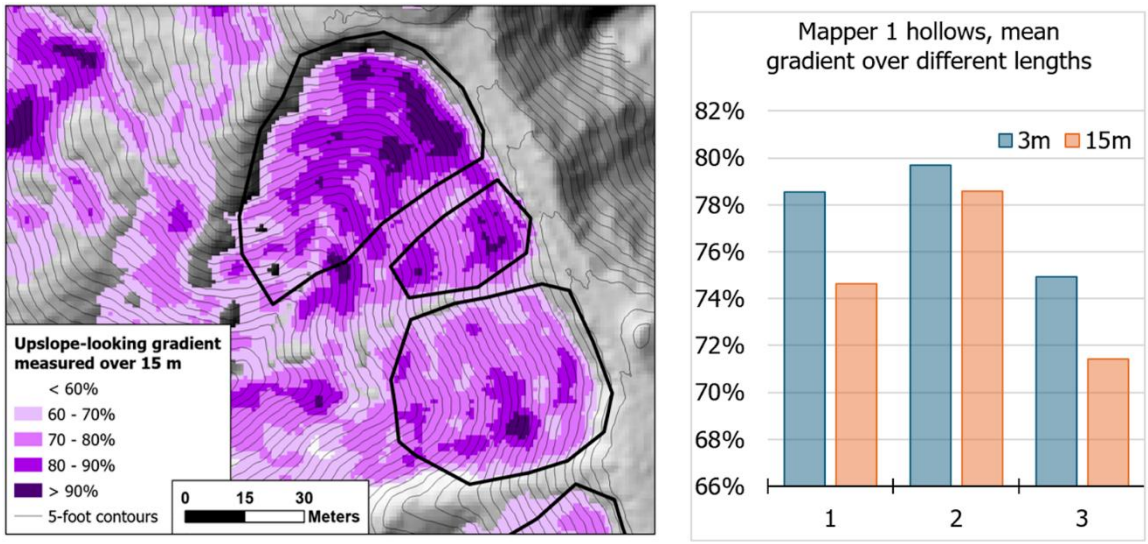


Figure E.6. Gradient looking upslope measured over a length of 15 meters. Bar chart compares the mean gradient for each hollow using gradient-measurement lengths of 3 meters and 5 meters.

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 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.

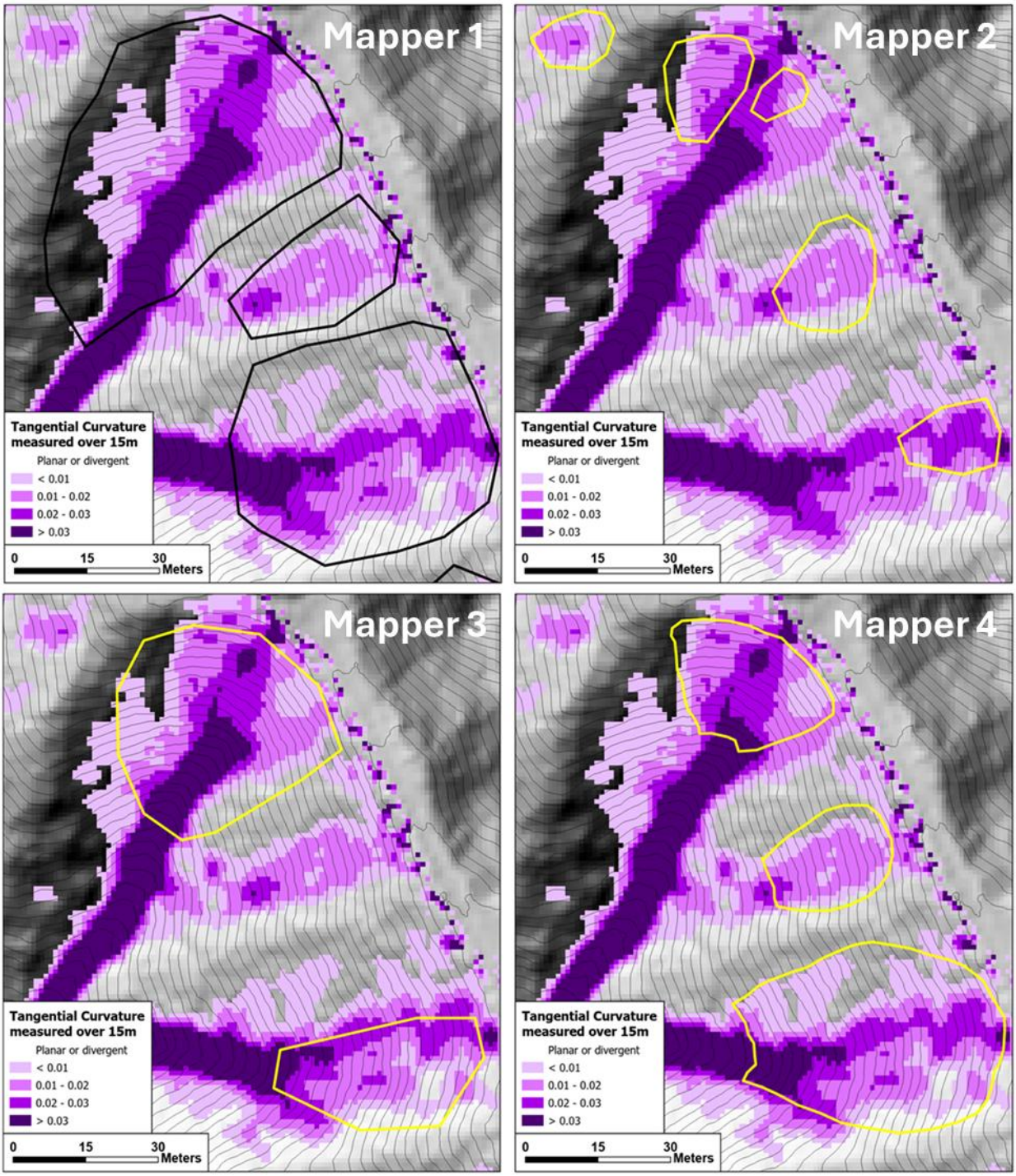


Figure E.7. Tangential curvature measured over a 15-meter length with mapped bedrock hollow polygons overlain. Planar (zero curvature) and concave-down (negative curvature) slopes are not colored.

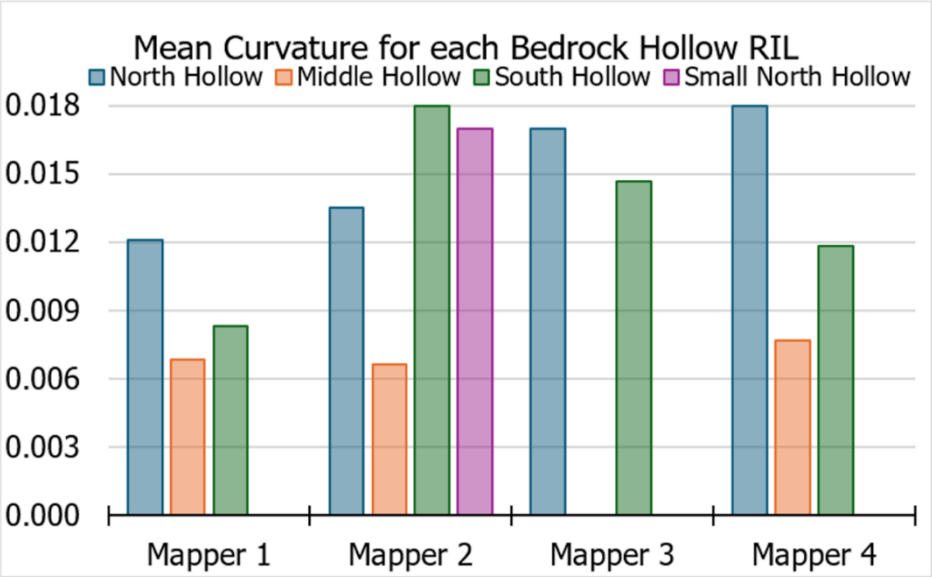


Figure E.8. Mean curvature for each mapped bedrock hollow RIL polygon.

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.

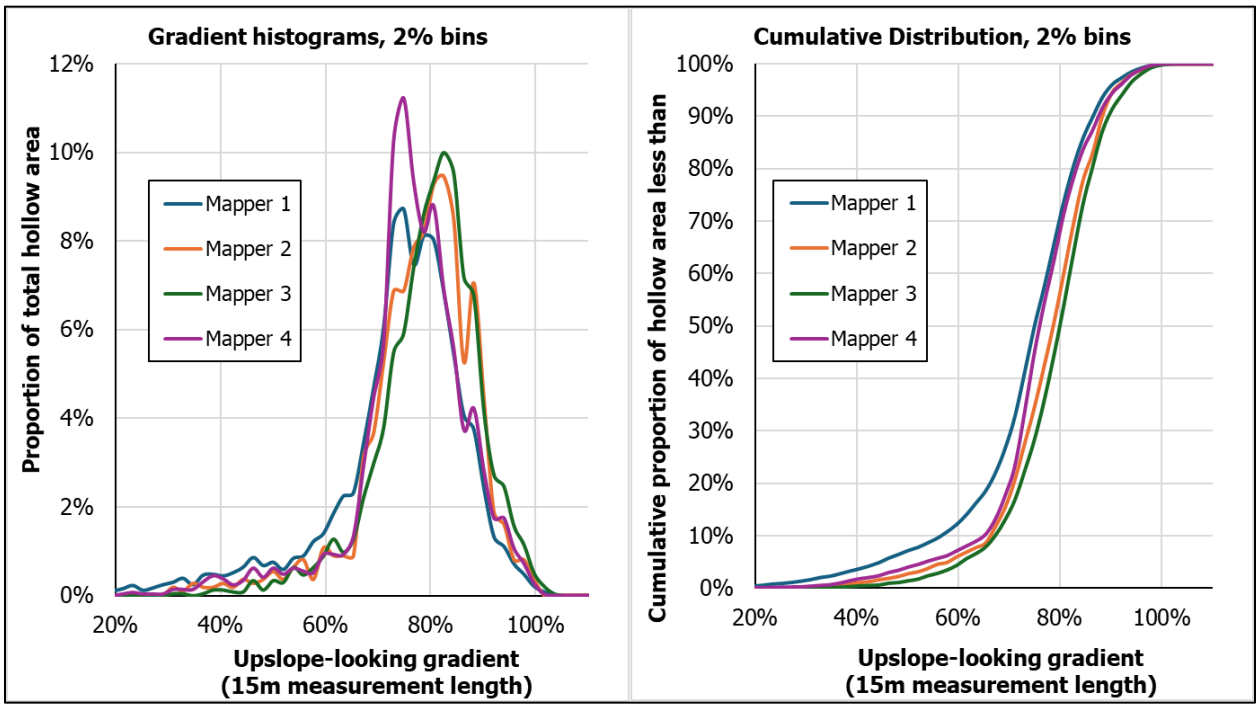


Figure E.9. Histograms and cumulative distributions of gradient over the area encompassed by the three hollows for each mapper.

Another important feature of the bedrock hollow 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 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.

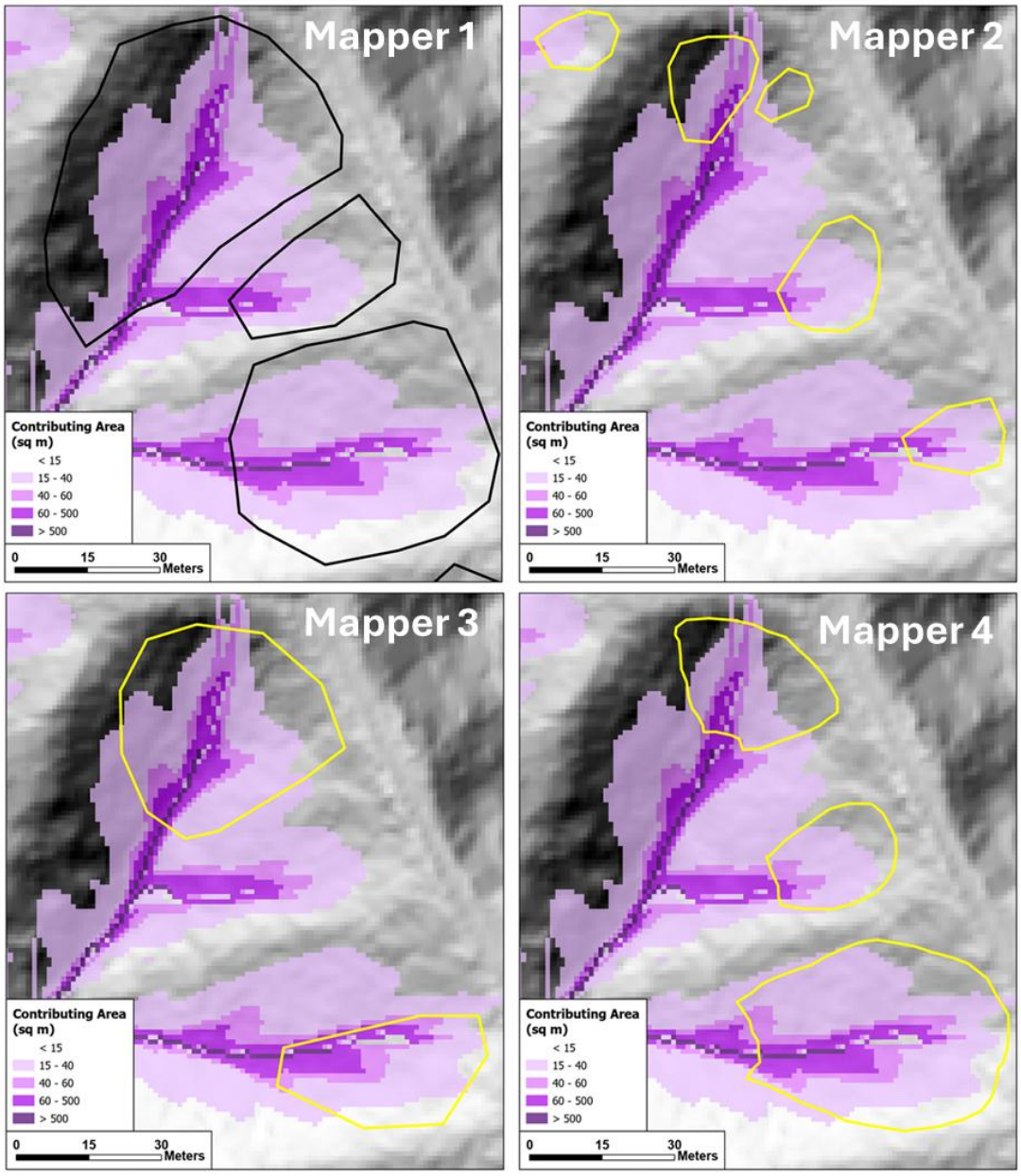


Figure E.10. Upslope contributing area with each mapper's polygons overlain.

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.

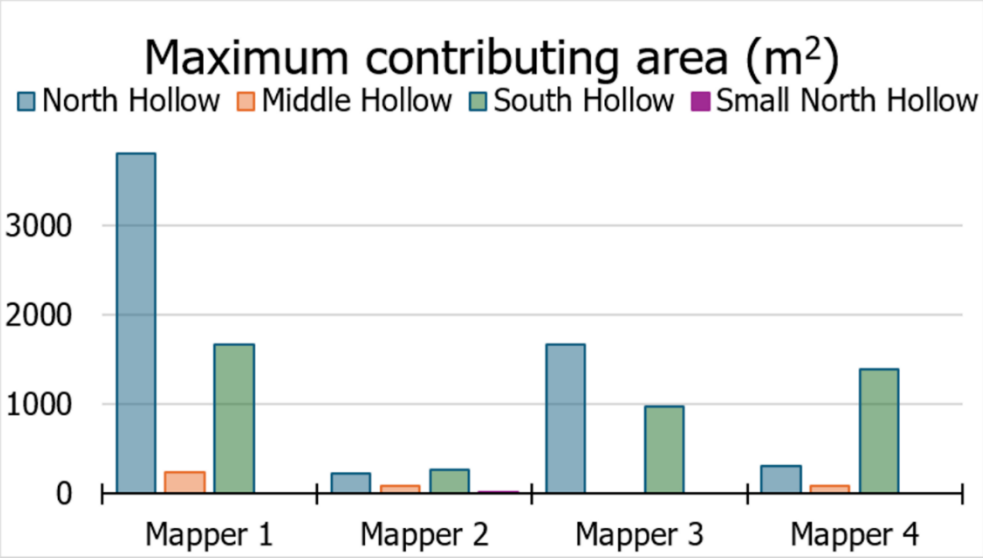


Figure E.11. Maximum contributing area at downslope exit point for each hollow.

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 in Appendix C 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.

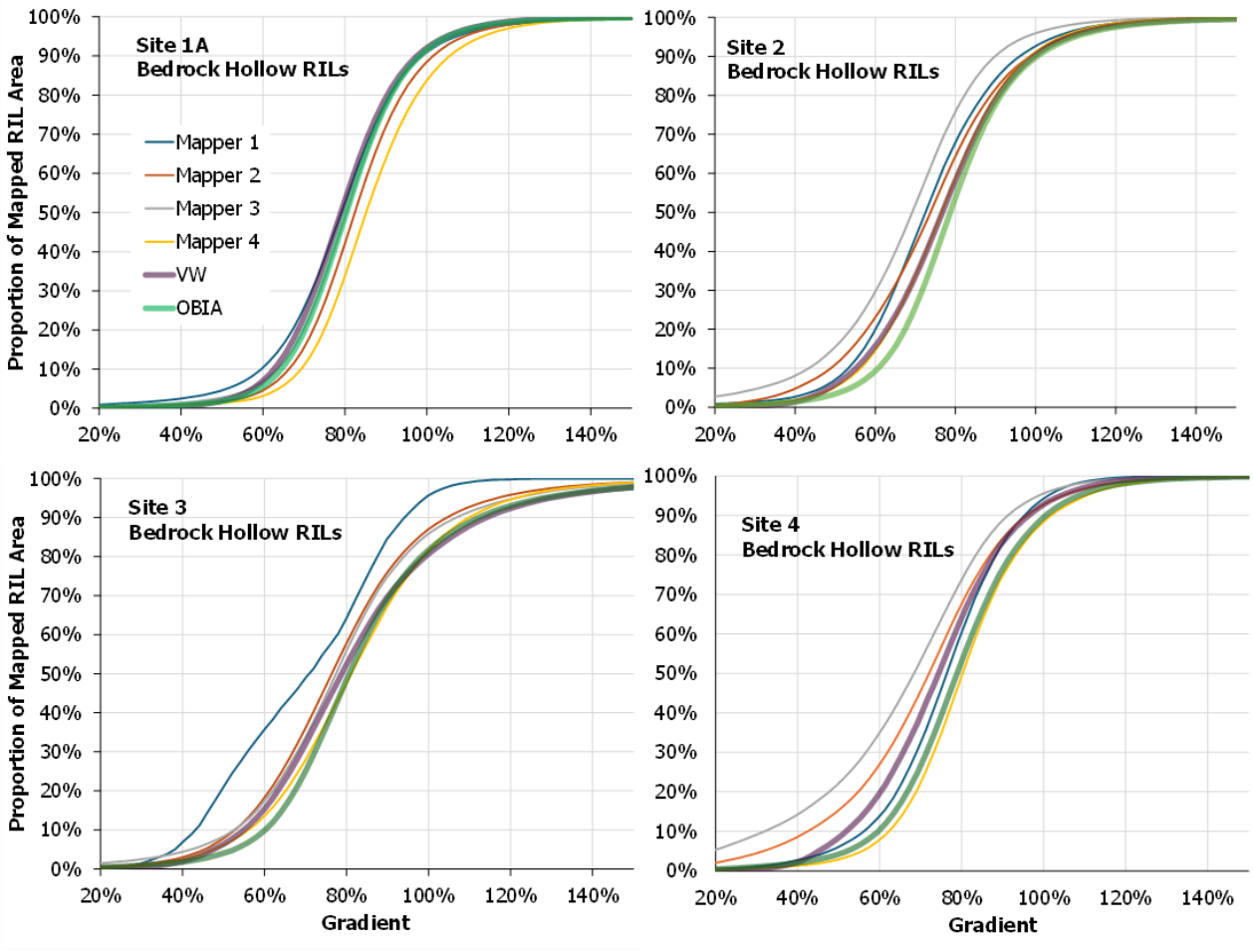


Figure F.1. Cumulative frequency distributions of upslope-looking gradient for bedrock hollow RILs for each of the four mappers and the two computer models at each of the four study sites (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

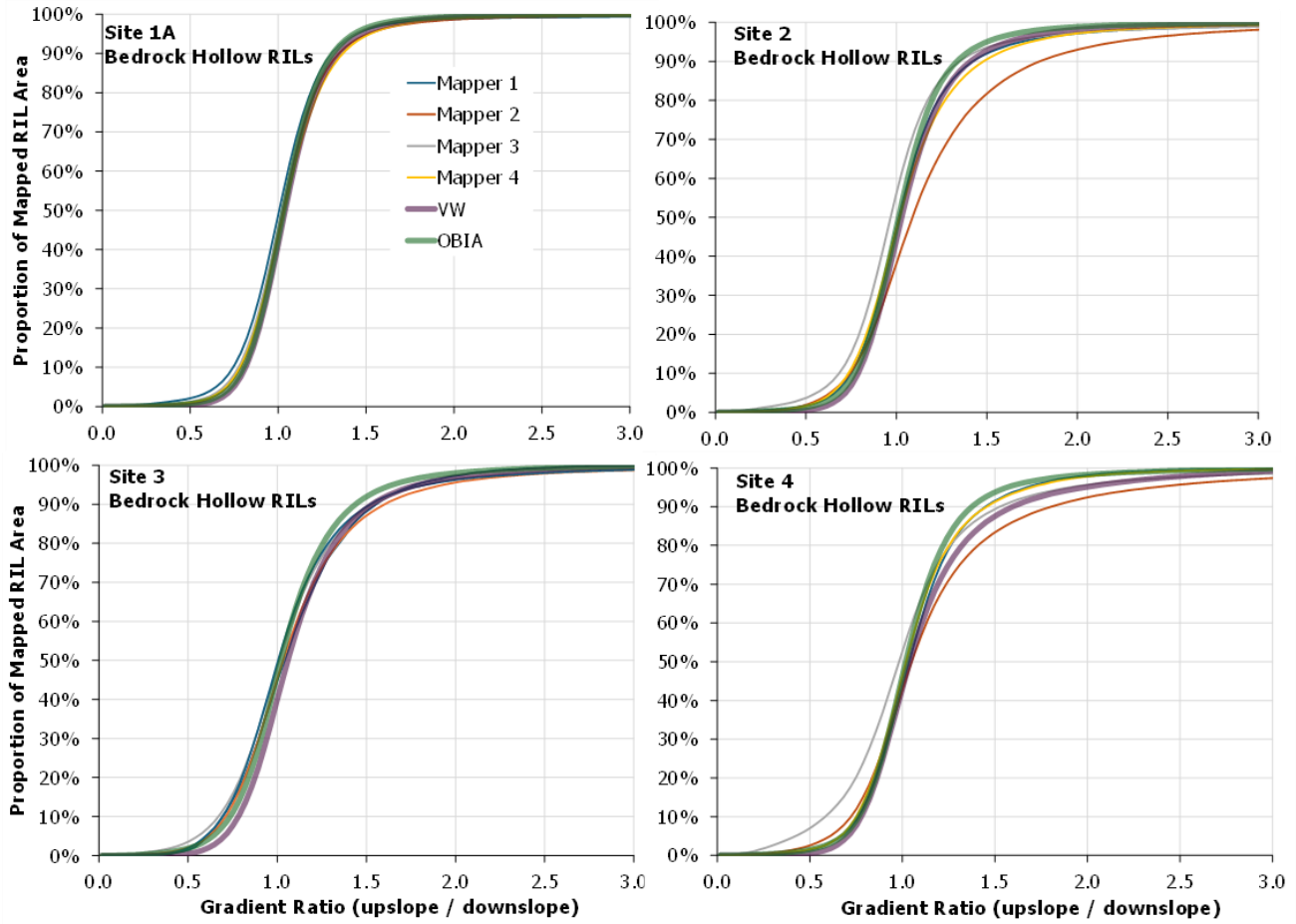


Figure F.2. Cumulative frequency distributions of the ratio of upslope-looking to downslope-looking gradient (profile curvature) for bedrock hollow RILs for all four mappers and the two computer models at the four study sites (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

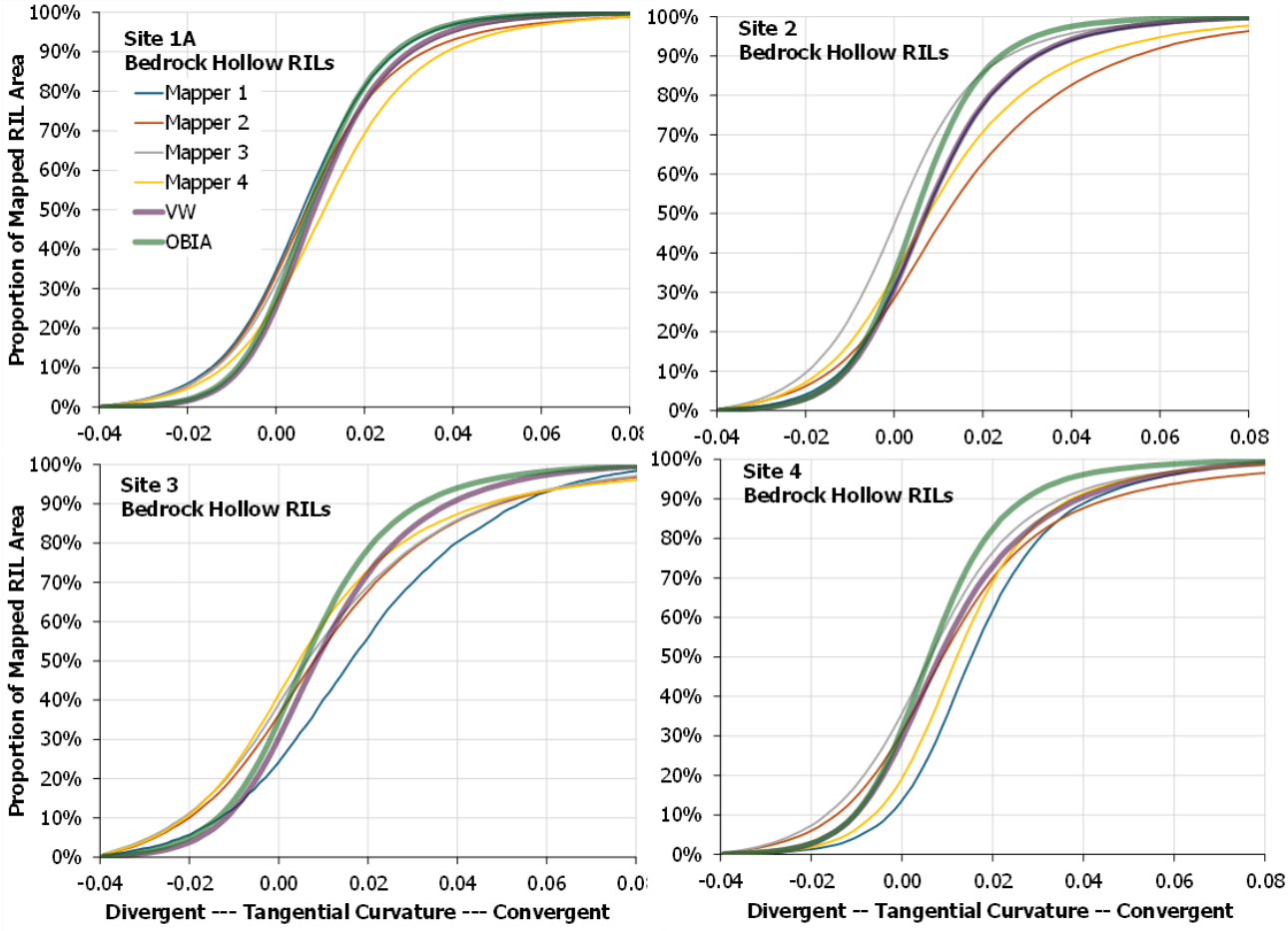


Figure F.3. Cumulative frequency distributions of tangential curvature for bedrock hollow RILs for each of the four mappers and the two computer models at each of the four study areas (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

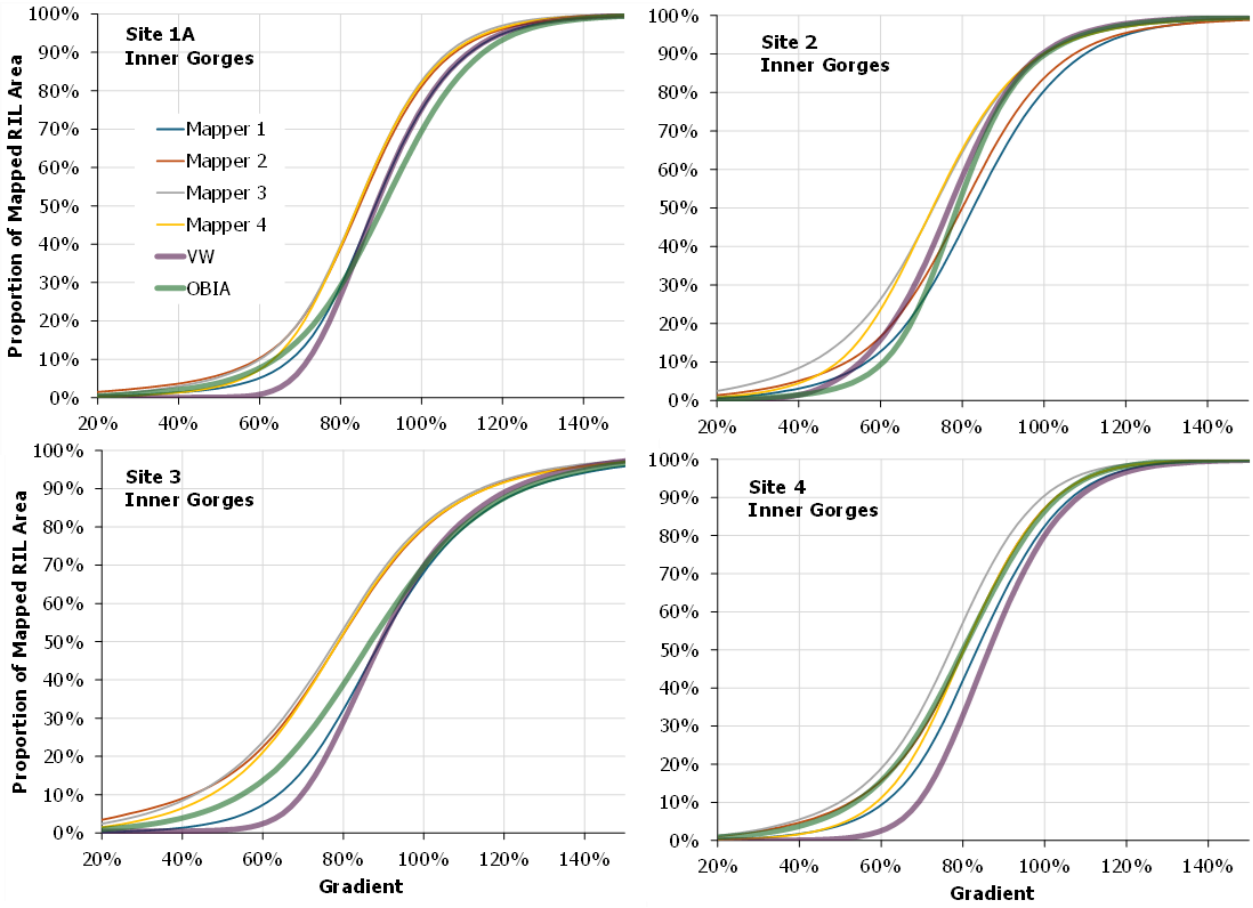


Figure F.4. Cumulative frequency distributions of upslope-looking gradient for inner gorge RILs for all four mappers and the two computer models at each of the four study sites (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

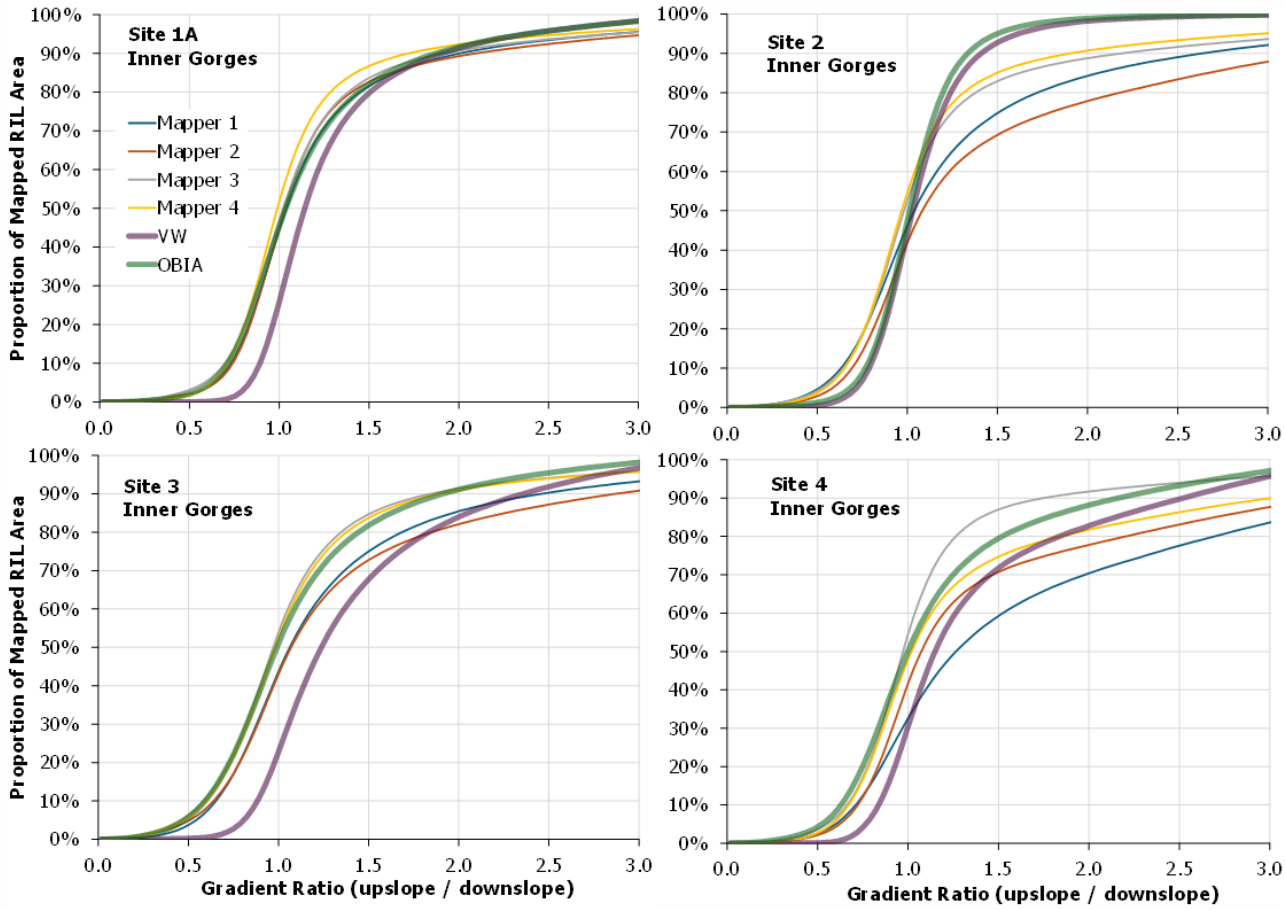


Figure F.5. Cumulative frequency distributions of the ratio of upslope-looking to downslope-looking gradient of inner gorge RILs by all four mappers and the two computer models for all four study sites. The horizontal axis extends to the maximum values found within the computer-generated polygons; mapper-drawn polygons included zones with larger gradient-ratio values (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

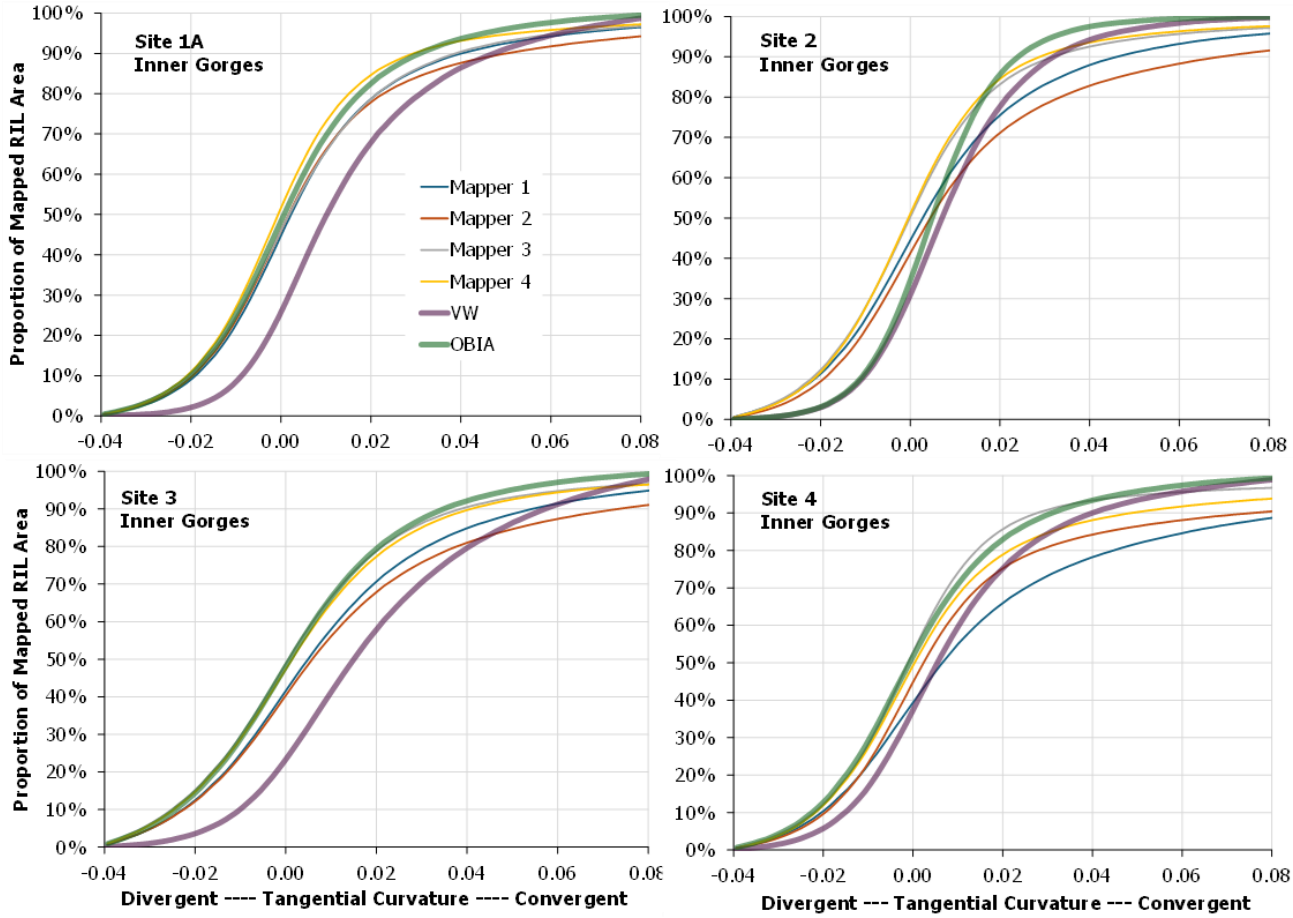


Figure F.6. Cumulative frequency distributions of tangential curvature in inner gorge RILs for the four mappers and the two computer models at all four study sites (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

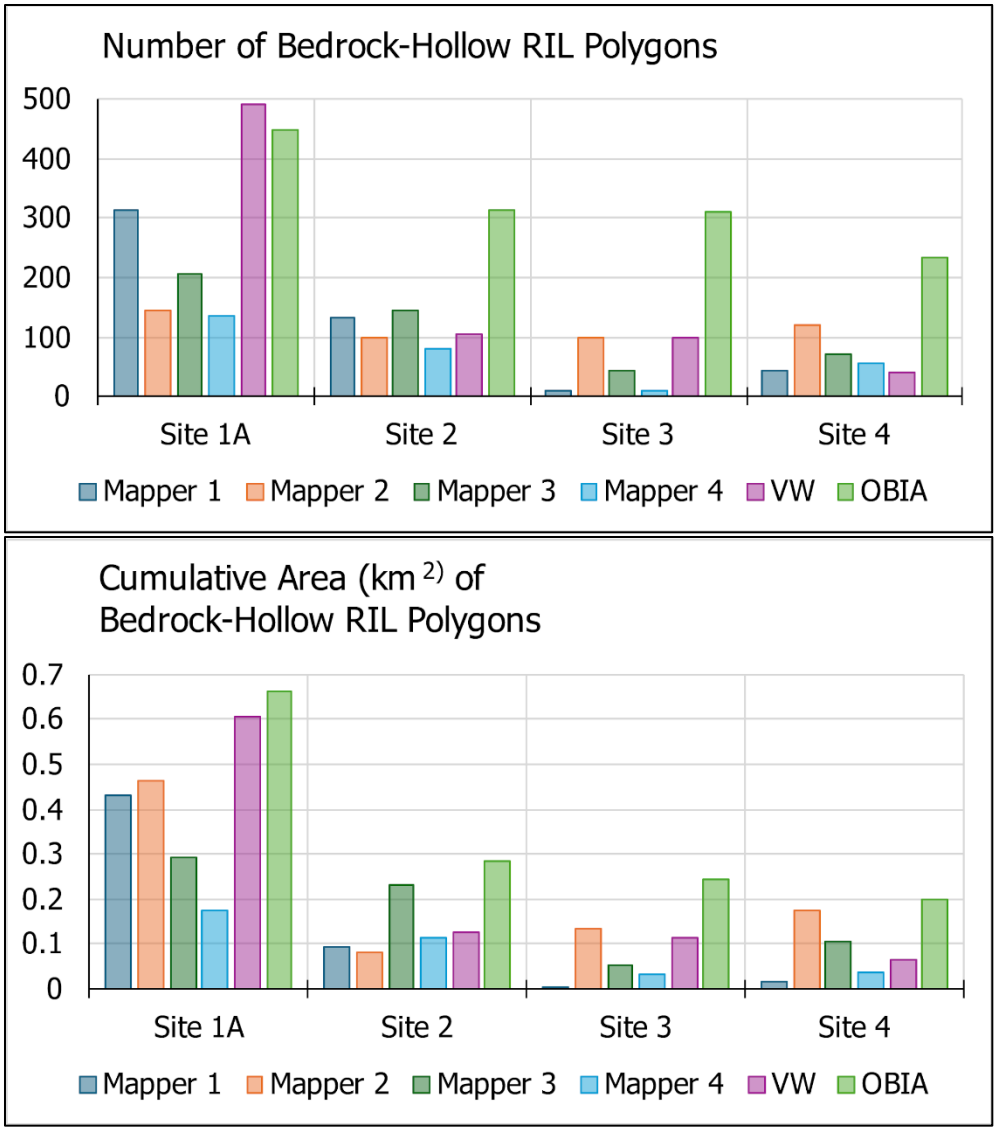


Figure F.7. Bar charts comparing the number and cumulative area of bedrock hollow RIL polygons for each mapper and the two computer models at each study site. Study-area size and zero-order-basin density varied from site to site, so the primary point of these charts is not the change in magnitude from 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 (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

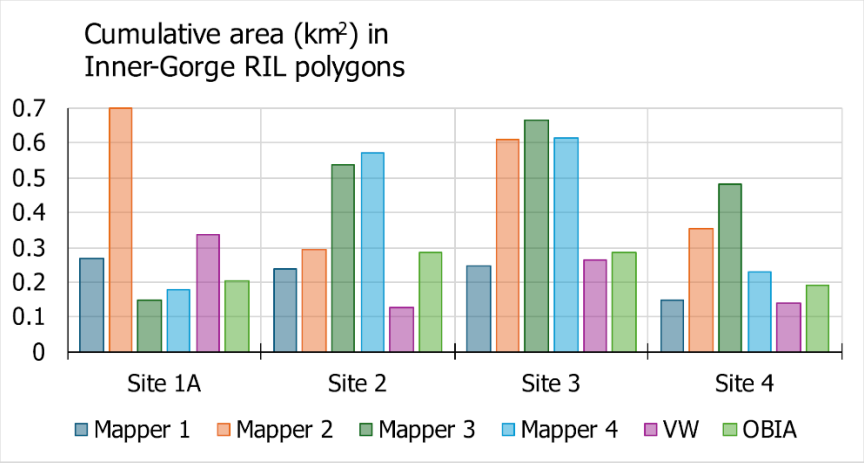


Figure F.8. Bar chart of cumulative area of inner gorges mapped by each mapper and the two computer models at each study site (Site 1a = North Fork Calawah, Site 2 = Fahnestock Creek, Site 3 = Howard Creek, Site 4 = Wishkah).

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.[10]

[10] 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.

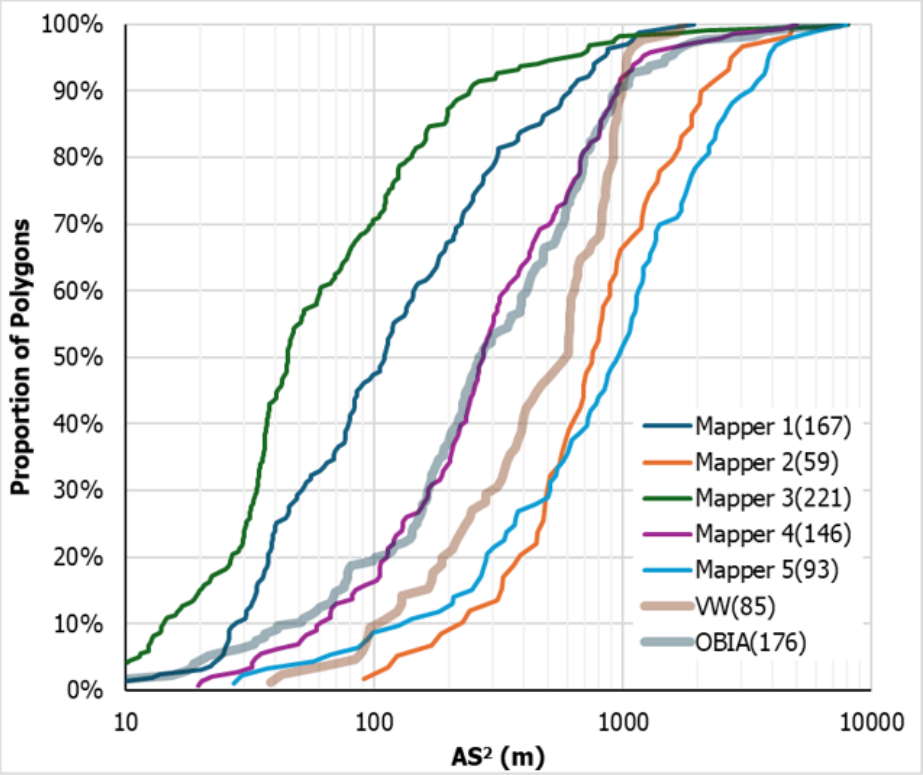


Figure G.1. Cumulative frequency distributions for the product of specific contributing area and channel gradient squared (AS^2) at the point of maximum contributing area for each mapped bedrock hollow RIL polygon for Site 1B. The number of polygons drawn by each mapper are indicated in parentheses in 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.

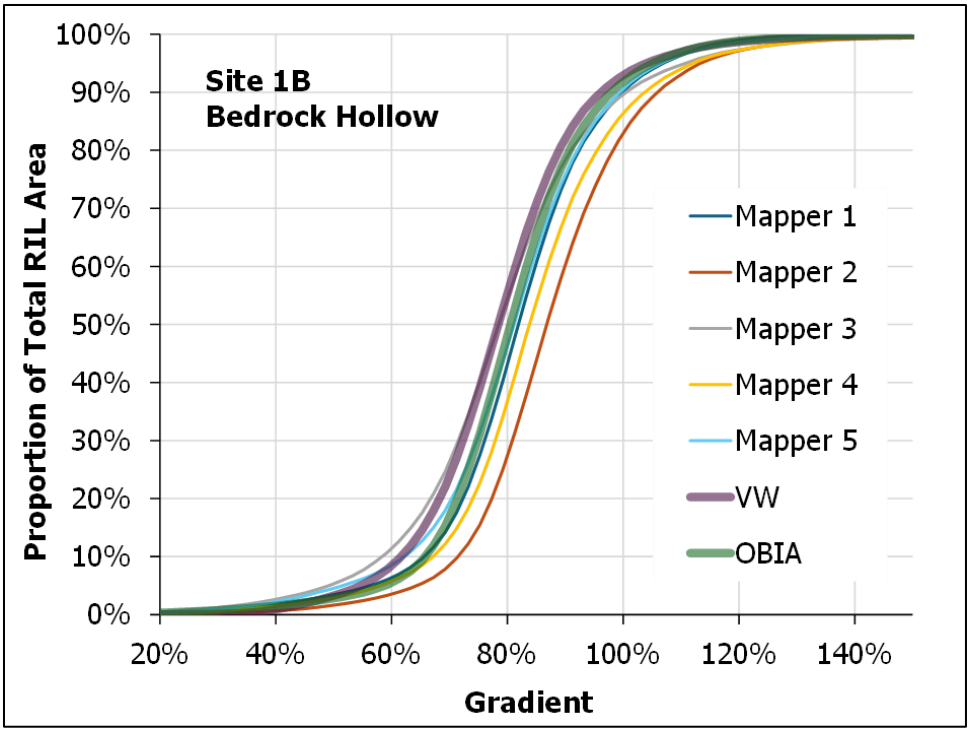


Figure G.2. Cumulative frequency distributions of gradient for mapping-team and computer-generated bedrock hollow RILs for Site 1B or Fahnestock Creek.

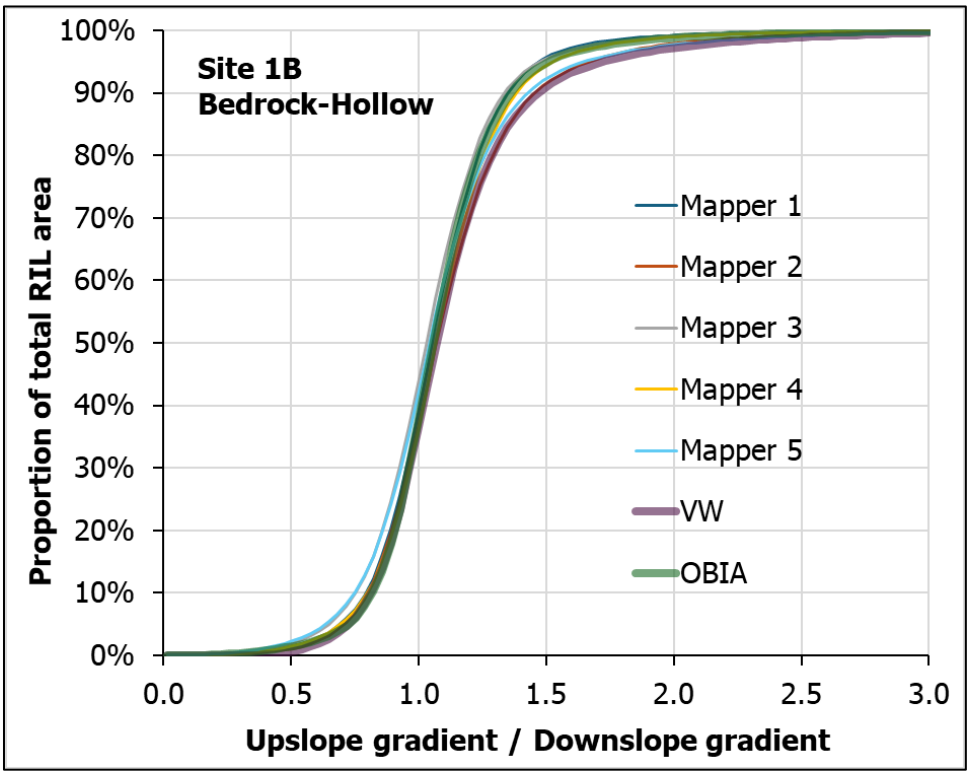


Figure G.3. Cumulative distributions for the ratio of upslope-looking to downslope-looking gradient for bedrock hollow RILs in the Site 1B or Fahnestock Creek.

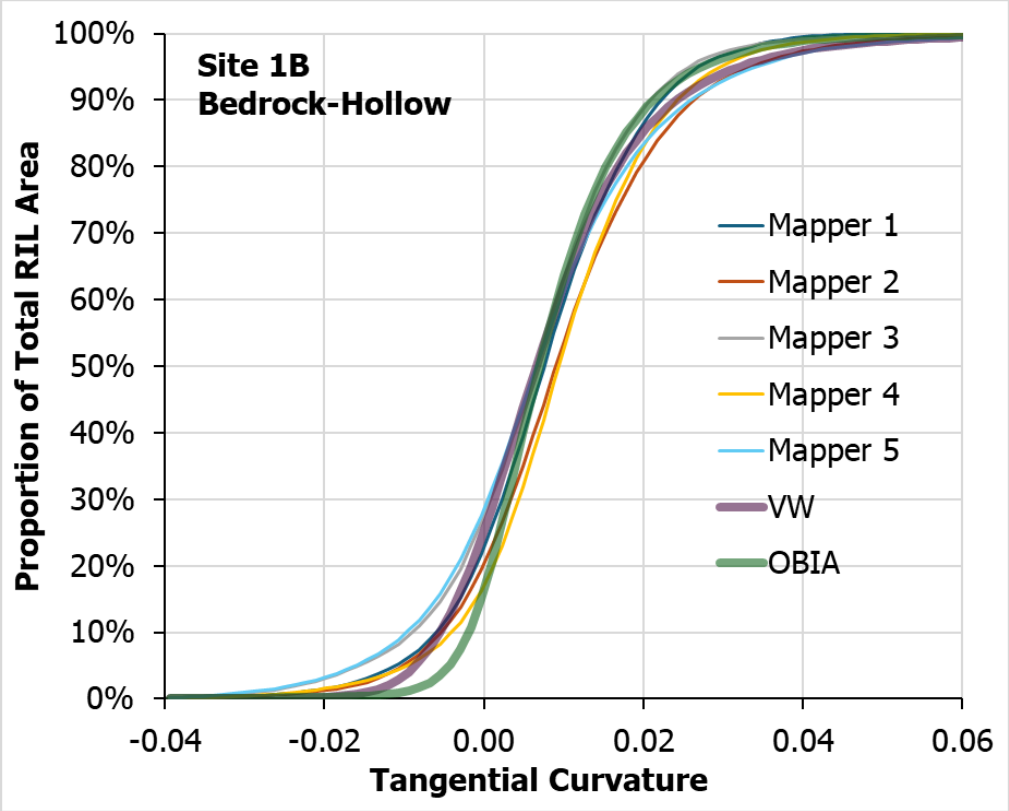


Figure G.4. Cumulative distributions of tangential curvature for bedrock hollow RILs, Site 1B or Fahnestock Creek.

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.

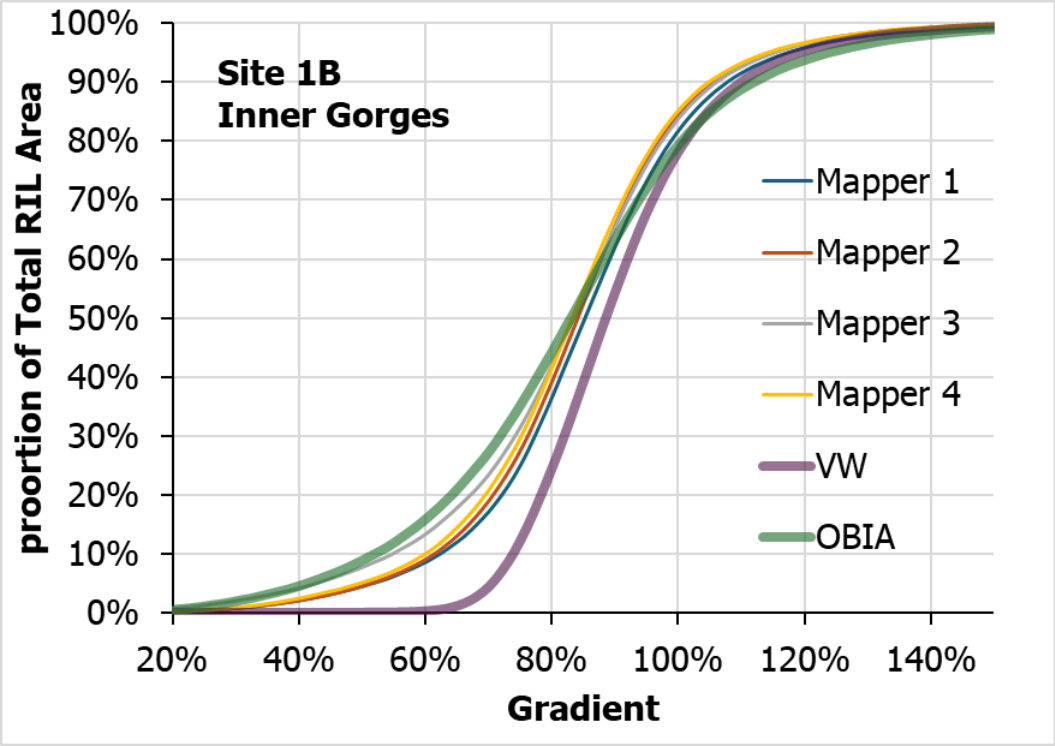


Figure G.5. Cumulative distributions of gradient for mapped and modeled inner gorge RILs in the Site 1B or Fahnestock Creek.

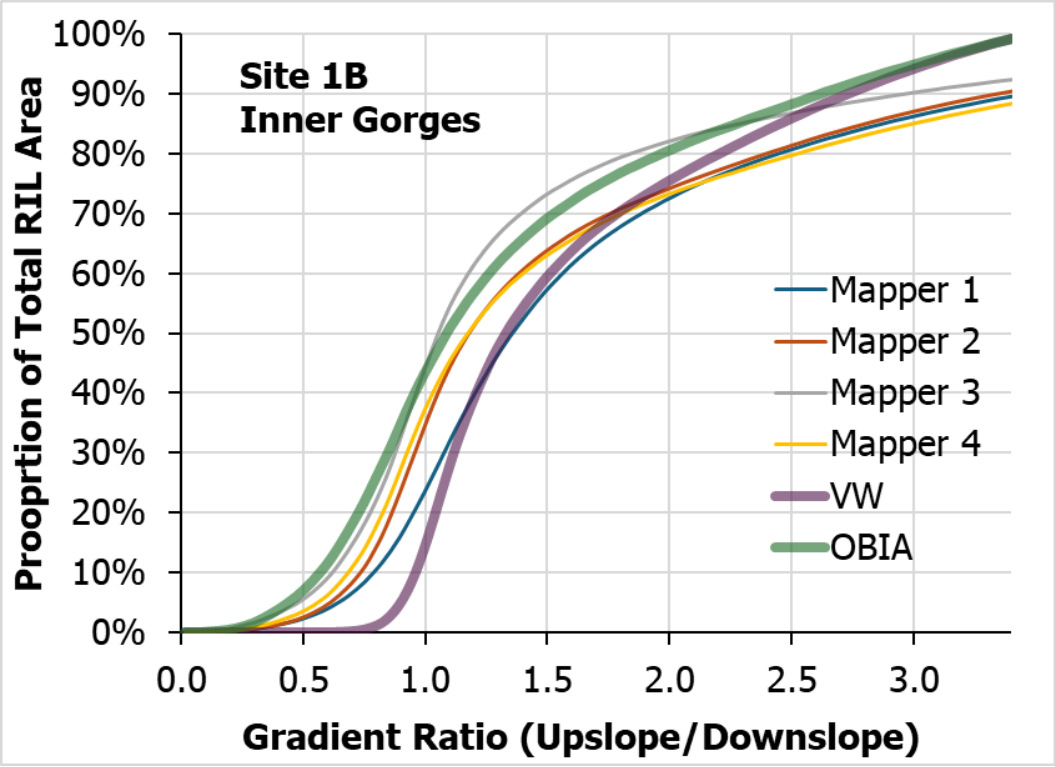


Figure G.6. Cumulative distributions of the gradient ratio for the Site 1B or Fahnestock Creek.

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.