APPROVAL SHEET

Title of Thesis: Studying Water from the Air: Using new measures of aquatic habitat to assess stream restoration outcomes

Name of Candidate: Hayley Oakland
Master of Science, 2020

Thesis and Abstract Approved: ______________________________
Matthew Baker, Ph.D.
Professor
Geography and Environmental Systems

Date Approved: ________________

ABSTRACT

Title of Document: STUDYING WATER FROM THE AIR: USING NEW MEASURES OF AQUATIC HABITAT TO ASSESS STREAM RESTORATION OUTCOMES

Hayley Oakland, M.S., 2020

Directed By: Matthew Baker, Ph.D., Professor, Geography and Environmental Systems

Stream restoration is increasingly synonymous with local scale physical channel modification, with the assumption that a certain range of local physical conditions will improve physical and ecological function. These projects often target measures of channel habitat captured in conventional field surveys, which rely on either low-resolution data over broad extents, or high-resolution data over fine scales that must be extended to the sampling reach. Thus, the degree of habitat modification in the restoration process, and associated ecological relevance, may not be fully understood by conventional measures. Advances in drone-based aerial surveying methods allow for continuous, high-resolution measures of channel habitat over broader spatial extents.

Methods were developed to extract physical channel habitat features from aerial surveys of restored streams in the Piedmont physiographic province of Maryland and Pennsylvania (chapter 1). Resulting data were realistic in magnitude, pattern and extent (chapter 1). The physical habitat data generated from aerial surveys were extracted in an equivalent manner to data generated from field surveys in spatially paired reaches using two habitat assessment protocols (chapter 2). Spatially

paired data were significantly correlated, associated or equivalent for most sampling unit comparisons, and all reach-level comparisons (chapter 2). Restored and unrestored extents were compared at multiple scales to assess differences in variation of physical habitat data (chapter 3). A trend of reduced variation within restored extents was evident, and associated with restoration objectives and catchment characteristics (chapter 3). These data and analyses represent a new approach to assessing physical habitat conditions, which could significantly contribute to understanding ecological outcomes of stream restoration.

STUDYING WATER FROM THE AIR: USING NEW MEASURES OF AQUATIC HABITAT TO ASSESS STREAM RESTORATION OUTCOMES

By

Hayley C. Oakland

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County, in partial fulfillment
of the requirements for the degree of
Master's of Science
2020

Acknowledgements

First, I would like to thank the Maryland Sea Grant, the Pennsylvania Department of Environmental Protection, and the William Penn Foundation for funding this research. Much of this work was funded through my Coastal Resilience and Sustainability fellowship through the Maryland Sea Grant, with match funding from the PA DEP and William Penn Foundation. The impact of this work was bolstered by working across disciplinary boundaries, collaborating with scientists, practitioners and managers, to understand local freshwater management issues from multiple perspectives.

This work was nurtured by the unique perspectives provided by my committee. I have been incredibly lucky to have a committee comprised of distinguished and passionate freshwater scientists: Drs. Matthew Baker, Stefanie Kroll, Andrew Miller and Chris Swan. Each member served as a seemingly endless source of theoretical and technical knowledge, who exuded enthusiasm for their research, and inspired my desire to continue work in freshwater science. My advisor, Matthew Baker, deserves a special thank you for his patience, insight and guidance.

I would like to thank the numerous other colleagues, family, and friends who helped me along the way and acted as a great source of support. I would like to thank Gina Lee and Alex Rittle for their invaluable help learning the tips and tricks for Structure-from-Motion data processing. I would also like to thank Gina Lee for all of the time and assistance (and fun) in the field. Lastly, (but far from leastly), I owe much of my sanity throughout this process to my partner, Jon, whose presence in my life provides an endless source of joy, laughter, comfort and support.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction and method development of SfM-based habitat measurement

1.1 Introduction and objectives

Stream ecological assessment has a long and varied history, however, it has gained particular prominence in the U.S. since the amendments leading to the modern Clean Water Act (CWA) in 1972 (Karr and Yoder 2004). Initial iterations of the policy mainly emphasized human "uses" of fresh surface waters, such as "drinkability" and "fishability," but scientists and policy-makers alike have come to realize the necessity of thriving ecosystems in maintaining even the most human-centric uses (e.g., Patrick 1949; Karr and Yoder 2004). Therefore, ecological assessments of streams have gained increasing importance in water resource management. The three major components used to assess stream water quality for both human use and ecological function have included conditions of aquatic biological communities (e.g. fish, aquatic insects, and algae), water quality, and physical channel habitat (Plafkin et al. 1989; Barbour et al. 1999; Kauffman et al. 1999; US EPA 2004).

The integrity of aquatic ecosystems is approximated by assessing proxy conditions, which can be represented conceptually. Figure 1.1.1 displays a portion of such a conceptual model, where each component of the aquatic ecosystem – water quality, physical habitat quality and biotic communities – acts as an individually useful proxy, and the interactions of the three components are also collectively important in assessing the condition of the system as a whole (Plafkin et al. 1989; Kauffman et al. 1999). These three components, green in the formative model represented in Figure 1.1.1, together comprise components of aquatic ecosystem integrity. The interactions between these proxies are not always completely straightforward or as expected due to both theoretical assumptions of association, and potential measurement error. For example, both high quality physical habitat and water chemistry can be associated with diverse biotic communities (e.g., Meador and Goldstein 2003). Yet improving just chemistry or habitat may not be enough to improve biotic community conditions (Hilderbrand et al. 2005; Bernhardt and Palmer 2011).

Understanding interactions among components of aquatic ecosystem integrity can also be affected by the way each is measured. Theoretical concepts of "quality" and "integrity" (circled in Figure 1.1.1) are approximated by empirical measurements (examples boxed in Figure 1.1.1). Therefore, variation associated with each component concept is limited to how it is measured, and the degree of error associated with each measurement. Even if measurements capture all that affects the "quality" of habitat or biotic communities, any error associated with those measurements will lead to weaker relationships between the measurements and the "quality" concept, which can propagate to lower association throughout the conceptual model (e.g., attenuation; Whitlock and Schluter 2015). Any measurement errors can propagate to faulty approximations and associations. Each concept and its associated measurements can be scrutinized in this way – as to whether the applied theoretical assumptions and methodological techniques yield an adequate representation of the true conditions of the system.

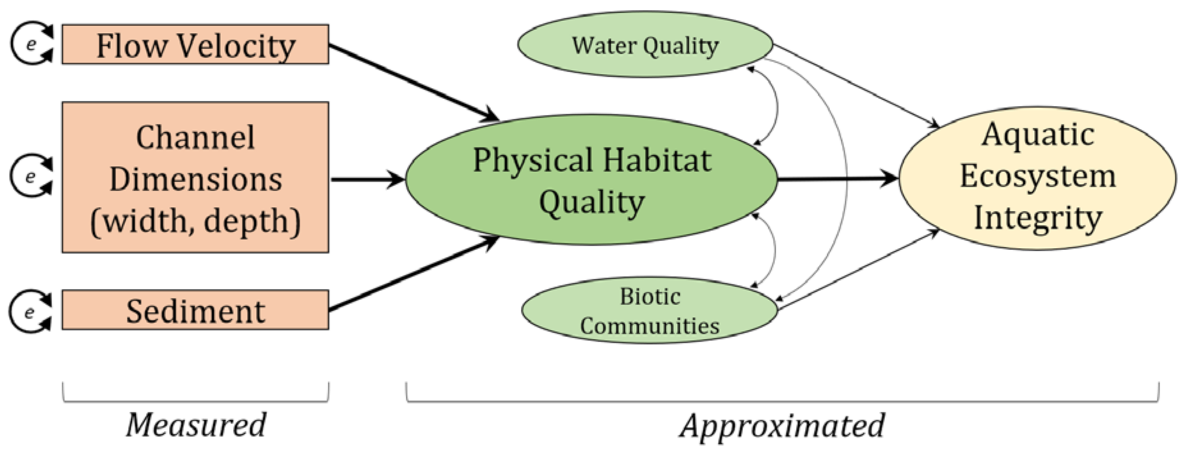


Figure 1.1.1. Conceptual model of aquatic ecosystem integrity approximation. Circled items are approximated concepts while rectangular items are measures contributing to approximated concepts, with associated measurement error (*e*). Bold arrows represent the focus of the current study. (There are many more interactions associated with aquatic ecosystem integrity not portrayed here.)

Attempts to quantify and qualify physical stream channel habitat have become a standard component of stream ecological assessment. Measurements of the physical environment within a channel, such as channel dimensions, flow velocity, and substrate quality, contribute to our understanding of overall physical habitat quality, which itself contributes to our understanding of aquatic ecosystem integrity (Figure 1.1.1). The accuracy of measurements of aquatic habitat and its relationship with aquatic ecosystem integrity are thus of particular interest when implementing and assessing the manipulation of habitat via stream restoration. This study will focus on the habitat assessment component of stream ecological assessments, and specifically in-stream metrics (i.e., excluding floodplain and riparian zone assessments). Stream habitat assessments typically involve a combination of qualitative (e.g., Plafkin et al. 1989; Barbour et al. 1999) and quantitative (e.g., Wolman 1954; Kauffman et al. 1999; Turnipseed and Sauer 2010) measurements of physical attributes of the aquatic environment. Physical attributes represent aquatic habitat with the assumption that they help to explain the ecological patterns in the biotic communities present in the given stream system (Fausch et al. 2002; US EPA 2004; Seelbach et al. 2006; Fetscher et al. 2009; Woodget et al. 2017).

Issues of habitat measurement are intimately related to restoration because both rely on the conceptual models of habitat-biota relations discussed above (Figure 1.1.1). The measurement of physical habitat is itself a tradeoff between specificity and extent. Precise, spatially explicit, and objective measures must be interpolated over broader extents, or alternatively, imprecise and subjective estimations are collected over broader extents and assumed to be representative. The need for interpolation or estimation can yield error in measurement (Kauffman et al. 1999), leading to lower confidence in the relationships shown in Figure 1.1.1.When restored habitat is assessed using the aggregate scores to which projects were designed, the projects may appear successful even when they fail to reproduce the heterogeneity of natural systems (Rubin et al. 2017). Viewing the restored riverscape in a precise, spatially explicit and continuous manner may help to reveal previously unappreciated characteristics of both restored and unrestored habitats.

Stream habitat classification based on variable spatial scales of influence has improved how stream habitat is identified and quantified (e.g., Figure 1.1.2; Frissell et al. 1986; Hawkins et al. 1993; Poole et al. 1997). There have been extensive efforts since the mid twentieth century to better classify streams hierarchically, observing the ways that environmental factors – such as multi-scale geomorphic features, geology or other landscape attributes, and temperature and other climatic features – interact and affect each other in creating local stream conditions (e.g., Bisson et al. 1982; Frissell et al. 1986; Hawkins et al. 1993,2000; Poole et al. 1997; Seelbach et al. 1997,2006; Thomson et al. 2001). Classification efforts improved comparisons between stream ecosystems, as classification allows for habitat assessments that better capture the landscape and watershed-scale processes that impact local habitat features (Frissell et al. 1986; Hawkins et al. 1993; Poole et al. 1997; Seelbach et al. 2006). Approaches to classification that differentiated stream systems using hierarchical models, and hydrological and biological factors, in addition to the classically employed geomorphological factors, were particularly successful in finding the most comparable systems (e.g., Seelbach et al. 2006; Baker et al. 2006b).

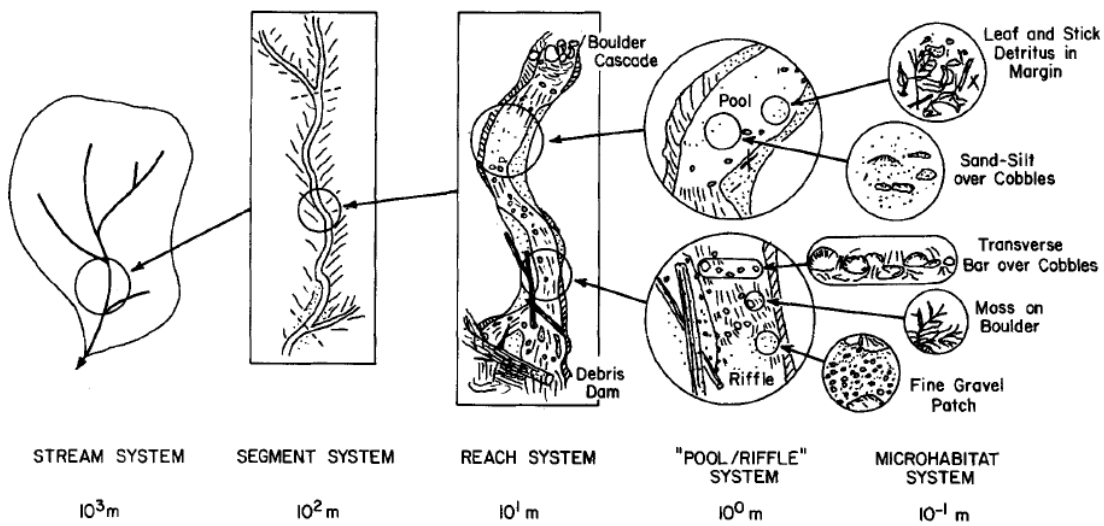


Figure 1.1.2. Hierarchical stream habitat classification at varying spatial scales, (figure reprinted from Frissell et al. 1986).

A further advancement in habitat measurement was made in scrutinizing the spatial scales of measurement in comparison to those of key ecological process (Fausch et al. 2002; Woodget et al. 2017). Stream habitats and aquatic communities are, for the most part, both affected by human activity and conserved by humans on an intermediate spatial scale (hundreds of meters to kilometers, Figure 1.1.2; Fausch et al. 2002). This is also the scale of key processes for aquatic communities (e.g., macroinvertebrate metapopulation dynamics; Fausch et al. 2002; Kominoski et al. 2010; Swan 2011). Fausch et al. (2002) discuss the “riverscape” to emphasize two key concepts: 1. habitat-biota relationships should be analyzed at the spatial scales that are most important to aquatic communities and the processes that create and maintain their habitats; and 2.

the fluvial landscape should be seen as continuous, rather than as the discrete reaches that we typically measure in stream ecosystem assessments (Figure 1.1.2).

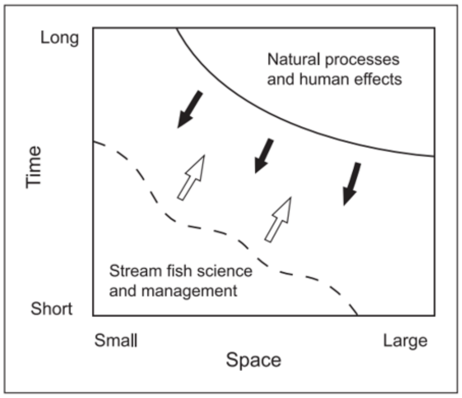


Figure 1. The gap in spatial and temporal scales between the scientific understanding and management of lotic fishes and the natural processes and human disturbances that affect the creation and maintenance of suitable habitat. The black arrows indicate that large-scale natural and anthropogenic effects interact in a hierarchical fashion with processes operating at smaller scales to influence habitat for declining stream fishes. The white arrows point toward the intermediate spatial and temporal scales at which stream fishes complete their life history, and at which fisheries ecologists and managers must begin working to effectively conserve them.

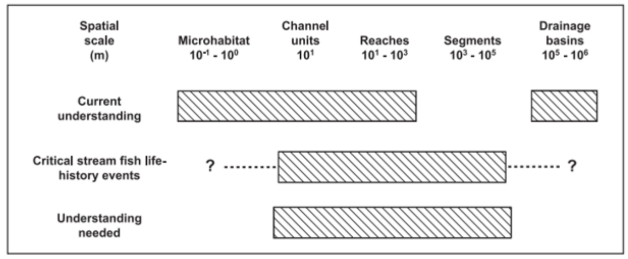


Figure 3. The spatial scales of our current understanding about lotic fishes, the probable range of scales spanned by critical life history events for many species, and the scales of understanding needed to adequately sample and predict attributes of fish populations and communities along riverscapes.

Figure 1.1.3. Spatial scales of biotic processes versus the scale of measurement and intervention, (figures reprinted from Fausch et al. 2002).

These concepts highlight the need for ecologists to understand the broader-scale processes that intimately affect the habitat features and communities they may observe at finer scales of measurement. Studies employing such understanding have been successful in discerning relative impacts and demonstrating relationships between communities and instream habitat (e.g., Seelbach et al. 1997; Hawkins et al. 2000; Lamouroux et al. 2004; Seelbach et al. 2006).

Recent technological advances in the field of remote sensing have allowed for high-resolution measurements of fluvial and riparian environments over broader extents, which may allow ecologists to measure aquatic habitat at scales closer to key processes (Carbonneau et al.

2012; Woodget et al. 2017). Most of these platforms – including Light Detection And Ranging (LiDAR), and satellite, spectral and structure-from-motion (SfM) -generated imagery – were designed and calibrated for measuring terrestrial features (Carbonneau and Piegay 2012). Recent research has highlighted potential for some remote sensing platforms to characterize fluvial landscapes, including both floodplain and in-channel features (Marcus et al. 2003; Carbonneau and Piegay 2012; Gleason et al. 2014; Tamminga et al. 2015; Woodget et al. 2015; Dietrich 2016; Marteau et al. 2017; Woodget et al. 2017; Dietrich 2017; Carrivick and Smith 2019).

For example, Marcus et al. (2003) is an early example of using remote sensing (hyperspectral imagery) to map components of stream habitat, including flow types, depths and woody debris. Fish habitat has been linked to geomorphic properties using unmanned aerial vehicle (UAV) -based SfM (Tamminga et al. 2015; Ventura et al. 2016) and land-based topographic surveys (Wheaton et al. 2010). Remote sensing data has been used to assess geomorphic change after restoration (Marteau et al. 2017) and to map aquatic habitat at equivalent intervals to field surveys (e.g., Marcus et al. 2003; Hall et al. 2009; Woodget et al. 2016). However, to our knowledge, remote sensing has not been applied to assessing aquatic habitat within restored streams, nor has it been applied to assessing variability in aquatic habitat. There have been limited applications of SfM data to ecological assessments and restoration assessments, and a lack of applications to assessing ecological outcomes of restoration, which is cited as a major gap in the literature (Carrivick and Smith 2019). A particular advantage in applying these platforms to stream ecological assessment is in the ability to generate a continuous view of the streambed, an important advancement in our ability to measure habitat on the scale of the "riverscape," (Fig. 1.1.3; Fausch et al. 2002; Carbonneau et al. 2012; Woodget et al. 2017).

Structure from Motion orthophotogrammetry (hereafter referred to as SfM) is an emerging method of fluvial remote sensing that approximates 3-D structures from continuous, overlapping, and high-resolution 2-D images. It has numerous advantages over the use of terrestrial and aerial LiDAR platforms, including affordability, flexibility, and the ability to analyze imagery in addition to elevation characteristics (Carrivick and Smith 2019). Published research has displayed the potential for SfM technology to collect submerged fluvial topographic data, which can be accomplished using an unmanned aerial vehicle (UAV; e.g., Woodget et al. 2015; Dietrich 2016). However, there have been relatively limited applications to instream habitat quantification (Marcus et al. 2003; Carbonneau and Piegay 2012; Tamminga et al. 2015; Woodget et al. 2016; Carrivick and Smith 2019).

A major benefit of SfM data lies in its continuity over broad extents without compromising the precision of spatially explicit data. Carrivick and Smith (2019) discuss the limited application of SfM to the field of aquatic ecology, particularly in the context of habitat heterogeneity. They allude to the establishment of SfM as an increasingly common means of geomorphic assessment, but note the relatively limited, and still burgeoning application to ecological and restoration assessment. Woodget et al (2017) review efforts to improve physical habitat measurement through remote sensing, particularly with UAVs. Woodget and collaborators have performed seminal work using UAV-based SfM to more continuously map stream and river habitat, including mapping submerged sediment (e.g., Carbonneau et al. 2004, 2005; Woodget 2015; Woodget and Austrums 2017; Woodget et al. 2018), surface flow types (e.g., Woodget 2015; Woodget et al. 2016), and bathymetry (e.g., Woodget 2015; Woodget et al. 2015, 2017).

Data collection via UAV with mounted digital camera equipment is cheaper and more flexible than traditional remote sensing collection techniques (such as LiDAR), and has the capacity to capture both coarse and fine-scale stream channel measurements (Woodget et al. 2015). The technique is also beneficial to long-term studies in the ability to archive and re-visit the models and orthoimagery during future analyses. The ability to return to, scrutinize, and re-analyze the imagery from which measurements were derived will also reduce subjectivity currently inherent in certain aspects of stream habitat assessment. Therefore, it is evident that remote imagery has the capacity to augment habitat assessment qualitatively and quantitatively, as well as alter the economics of collection.

Despite ability to collect extensive imagery, there are, as yet, no standard methods for extracting measures of physical channel habitat from SfM data comparable to field surveys. To our knowledge, there has been no direct comparison of SfM-generated habitat measures and standard field habitat assessment protocols in the context of analyzing restored streams (Woodget et al. 2017; Carrivick and Smith 2019). The objective of this study is to assess metrics derived from low-altitude aerial imagery to inventory high-resolution physical stream channel habitat, and to examine variability in the aquatic habitat of restored and unrestored streams. This objective will be accomplished in three chapters: first by developing repeatable methods to extract habitat metrics from SfM data consistent with those taken during standard field habitat assessments (Chapter 1); then by comparing habitat measurements made by SfM data to those made in the field (Chapter 2); and finally by applying SfM data to restoration assessment (Chapter 3).

1.2 Methods

1.2.1 Study design

Seven restoration sites in the Piedmont physiographic province of Maryland (MD; 4 sites) and Pennsylvania (PA; 3 sites) were surveyed for this study (Figure 1.2.1). Each group were subsets of sites from two larger investigations in the respective states (Hilderbrand, *unpublished data*; Kroll, *unpublished data*), each of which had the objective to examine local restoration projects and associated outcomes in aquatic habitat and biotic communities. Sites for the current study were selected to utilize data from these ongoing efforts. The subset of sites from each state were chosen for comparability based on catchment characteristics (Table 1.2.1), regional physiography, and restoration practices employed. All restoration projects were completed at least two years prior to the start of this study. Restoration projects were at least 200 meters in extent along the stream, and each included instream habitat manipulation, such as channel regrading and/or instream structure placement. In addition to instream work, most sites included riparian buffer enhancement and/or bank stabilization. The sites were also selected in consultation with project practitioners, and state and county conservation representatives who were actively engaged in studying previously restored sites and in ongoing monitoring. A reference site was selected in consultation with the Maryland Department of Natural Resources, from the Maryland Biotic Stream Survey's Sentinel Sites program. This site was to serve as a reference, in that it was unrestored and represents high quality channel habitat and aquatic biota in the Piedmont. However, this site was ultimately excluded from the study due to logistical difficulties involved surveying, which are articulated below.

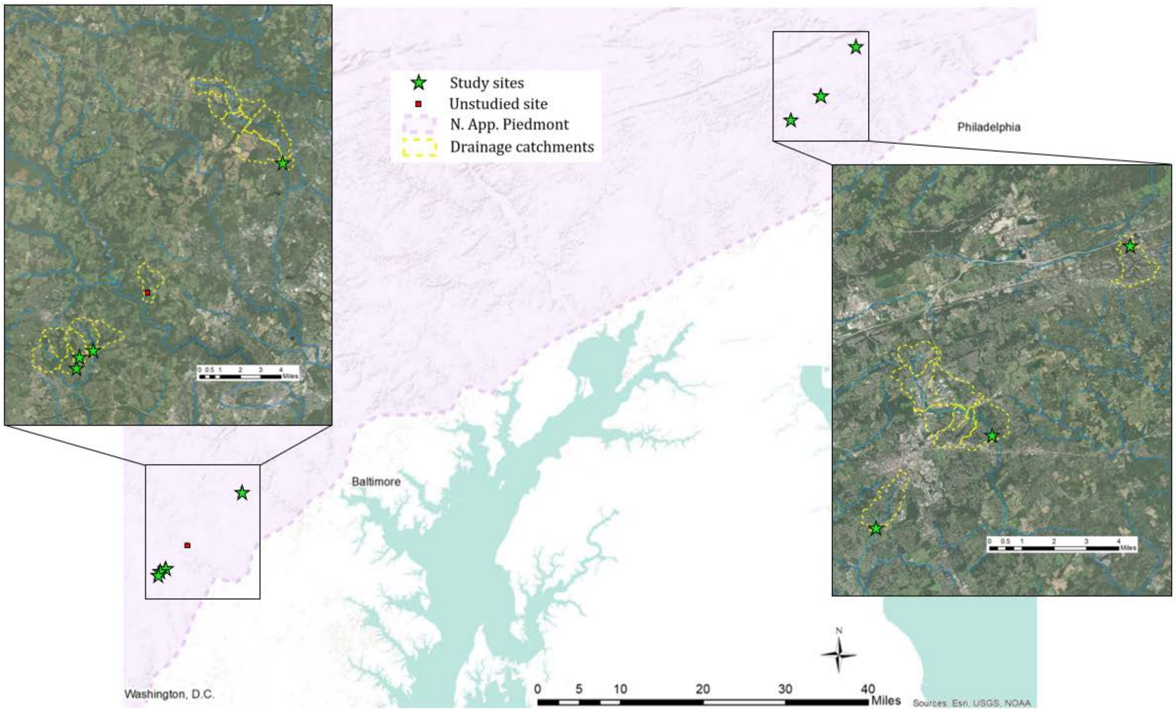


Figure 1.2.1. Distribution of sites, with closer view of Maryland and Pennsylvania sites with upstream catchments.

Table 1.2.1. Site upstream drainage area characteristics.

	144	253	CRAB	EBCH	PLUM	RKGR*	ULP2	WL
Sampling protocol	MBSS	MBSS	EMAP / SWAMP	EMAP / SWAMP	EMAP / SWAMP	MBSS	MBSS	MBSS
Drainage Area (km^2)[a]	3.02	8.82	3.12	16.26	3.25	3.48	26.03	6.69
Strahler stream order[a]	2	3	1	2	1	1	4	4
Drainage density[a]	1.01E-03	7.75E-04	8.86E-04	7.17E-04	8.85E-04	9.16E-04	8.12E-04	8.30E-04
Sinuosity[a]	1.06	1.14	1.14	1.13	1.14	1.1	1.23	1.09
Basin shape index (BSI)[a]	2.64E-02	2.82E-02	2.23E-02	2.38E-02	2.37E-02	2.39E-02	2.47E-02	2.57E-02
Basin Slope (°)[a]	3.1	3.2	6.5	2.9	4.3	4.3	3.2	2.6
Prim. rock type and % of US catchment[b]	Schist 96.7	Tonalite 79	Schist 74.1	Gneiss 62.9	Gneiss 98.5	Schist 45.1	Schist 46.4	Schist 98.7
Sec. rock type and % of US catchment[b]	Tonalite 3.3	Schist 21	Limestone 24	Schist 35.5	Serpentinite 1.4	Gneiss 38.4	Gneiss 40.7	Tonalite 1.2
US Area (km^2)[c ^]	2.9	8.83	2.8	15	3.24	3.44	26.16	6.58
% US Forested[c]	20.6	37.4	11	19	20	31.2	19.7	30.7
% US Developed[c]	27.4	34	52.1	73.4	74	11.7	48.5	36.1
% US Impervious[c]	3.78	3.86	11.2	17.7	13.1	1.8	7.61	5.91
Mean Annl. Precip. (")[c]	44.1	43.8	45	45	45	44.9	45.4	44.3

(a) USEPA and USGS 2005; Wieczorek and LaMotte 2010; MD iMAP 2011, 2013; (b) Cleaves et al. 1968; PA DCNR 2015; (c) USGS 2016b,2016a. **RKGR is the unstudied reference site. ^US Area is the portion of the drainage area upstream of the point on the stream, (e.g., not including the area of the entire catchment that point lies within).*

Sites included field study reach pairs or triplets within the channel – upstream of (as reference) and within; or upstream of (as reference), within, and below – stream restoration projects. Choosing study reaches in this way isolated changes due to habitat alteration by restoration projects, as other parameters (water quality, hydrology, etc.) were largely controlled when comparing locations within the same stream. Reach pairs were assessed at PA sites, and triplets at MD sites to conform to the protocols used by Hilderbrand (*unpublished data*) and Kroll (*unpublished data*), respectively. The following workflow was established to extract measures of stream channel physical habitat consistent with field assessments of instream habitat conditions. The workflow was repeated at each study site from upstream to downstream to include each restored reach and unrestored neighbors (direct comparison in Chapter 2).

1.2.2 Field Surveys

1.2.2.a Aerial surveys

Prior to aerial assessment, a Leica GNSS 1200 Real-Time Kinematic (RTK) unit was used to survey precise (0.003 – 0.022m RMSE) global positioning system (GPS) latitude, longitude and elevation coordinates for points made visible during aerial surveys by placing orange bucket lids over the surveyed points. These ground control points (GCPs) serve to georeference the orthophotos and improve the accuracy of the point cloud generated from the orthophotos. Ten to fifteen GCPs were placed at each site, ensuring that each flight area included at least five GCPs, some of which overlapped with surrounding flight areas to aid in stitching flights together when processing the imagery.

Aerial surveys were completed between January and March 2018 to maximize the land surface visible during leaf-off season. The surveys were completed using a DJI Phantom 4 Pro, a small UAV mounted with an RGB camera. Flights were pre-planned using the GroundStationPro application for iPad. Pre-planning ensured automated flight paths that would capture the GCPs and would ensure line-of-sight access to the UAV (maximum distance 600m, site dependent), as stipulated by the Federal Aviation Administration Part 107 regulation. Within the GroundStationPro application, flight height was adjusted to control image pixel resolution, and image percent overlap was adjusted to control for the ability to align and stitch the resulting imagery. Flights were set to approximately 300ft above ground level, depending on site conditions, to obtain approximately 3cm/pixel resolution, with 70% overlap between images. Each flight area was flown once at nadir (camera pointing straight down), and twice at oblique angles (60° down angle). Oblique flights were flown approximately parallel to each stream channel to capture parts of the stream banks not visible in nadir imagery. Similarly, Dietrich (2017) found that multiple view angles of shallow streams could improve channel bathymetry over nadir view angles alone.

1.2.2.b Discharge cross sections

On the day of each flight, a cross-section discharge measurement was collected at the downstream end of the surveyed area with a Sontek Flow Tracker (accuracy ±1% of measured velocity; SonTek/YSI 2007). The cross-section location was chosen based on the downstream end of the study area, as well as where the channel bottom was relatively uniform. The cross-section width was measured, which was divided by 10-15 to determine the location of each incremental measurement along the cross-section transect. The depth is measured at each location, and the probe is adjusted to 60% of the depth. The probe is placed approximately perpendicular to flow for the 40 second measurement interval, which calculates a velocity measurement per second, and averages those measurements to get an average velocity per measurement location. Measurements were repeated where needed when the instrument indicated too great a flow angle or fluctuation throughout the measurement interval. Unweighted mean depth and velocity were calculated for each cross section. Each measurement location depth is multiplied by the width increment between the locations to get each the flow area per measurement, which is summed to find the total cross section flow area. The discharge per increment is the product of the increment's mean velocity and its flow area, and the total cross section discharge is the sum of the incremental discharges.

1.2.2.c Field habitat assessments

Field habitat assessments for this study consisted of portions of the protocols used by Hilderbrand (*unpublished data*) and Kroll (*unpublished data*). At the paired reaches within the three PA sites, the field protocol was a modified version (Kroll, *unpublished data*; Kroll et al. 2019) of physical channel habitat assessments taken under EPA's Wadeable Streams Assessment Environmental Monitoring and Assessment Program (EMAP; Kauffman et al. 1999; US EPA 2004) and California's State Water Resources Surface Water Ambient Monitoring Program (SWAMP; Fetscher et al. 2009). Under this protocol (hereafter referred to as modified EMAP / SWAMP transects) measurements were collected from points along 11 cross-sections per each 100m reach (Figure 1.2.2). Wetted width was measured at each transect. The distance from bank and depth were measured, and flow class (riffle/glide/pool) and substrate size class (e.g., silt, sand, bedrock, etc.) were categorized, at each of 5-7 points along each transect. The number of sample points was dependent upon width of the stream. Points included left and right banks, 1m from the left and right banks (both omitted if <4m wide), the center of the stream, and the "left center" and "right center" (points between the center and respective banks). Cross section discharge measurements were not collected at the site on the day of the field surveys, so discharges at the downstream end of sites was approximated using downstream USGS gage stations. A log-linear relationship was generated between the drainage areas and discharges for each site and paired downstream gage on the dates of the aerial surveys. The slope of the log-linear line for that relationship was used to predict the discharge at each site on the day of the field survey, using the site and gage drainage areas, and the discharge at the gage on the date of the field survey.

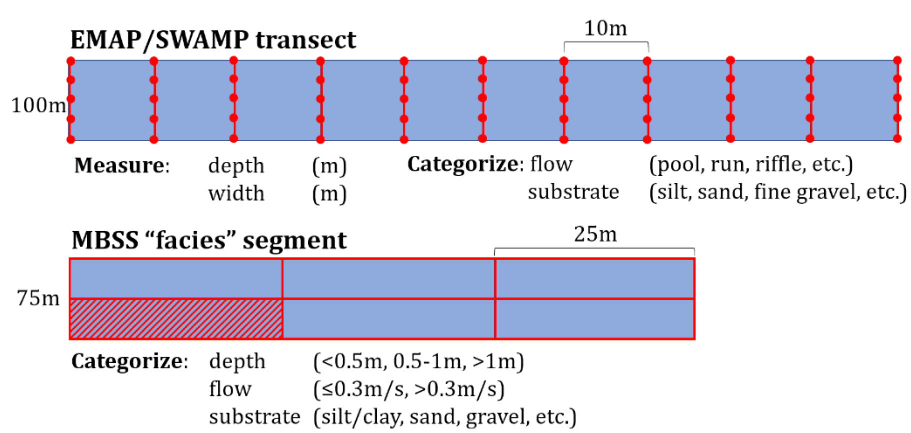


Figure 1.2.2. Field habitat assessment reach layout for EMAP/SWAMP and MBSS sites. At PA sites using EMAP/SWAMP transects, habitat variables were either measured or categorized at individual points along eleven reach transects perpendicular to flow and 10m apart. At MD sites using MBSS segments, categorical habitat variables were summarized across reach segments, which split the reach down the middle laterally and into three 25-m intervals longitudinally. Wetted width and thalweg depth and velocity are measured at MBSS transects, which occur every 25m, inclusive of the 0m-75m marks.

A portion of the Maryland Biotic Stream Survey (MBSS) physical habitat assessment field protocols were used for the triplet reaches at the four MD sites. The spring facies map portion of the MBSS protocol was completed, which involved dividing the 75m reach into 6 segments (25m long by ½ channel width wide; Figure 1.2.2), and categorizing the average depth, flow, and dominant/subdominant substrate category per segment (Stranko et al. 2015). Hereafter this facies map protocol is referred to as MBSS segments. A portion of the MBSS summer habitat assessment protocol was completed to assess field measurements of the categorical values estimated in the facies mapping process. For this protocol, hereafter referred to as MBSS transects, wetted width and thalweg depth and velocity were measured at cross-section transects at the downstream and upstream ends of each of these segments (at 0, 25, 50, and 75m). Cross section discharge measurements were taken at the downstream end of the study area per site. For comparison to the discharge estimates made with USGS gages for EMAP / SWAMP sites, the same process was used at three MBSS sites with a nearby downstream gage. Thereby measurements at each of these sites could be used to check gage predictions for each site.

1.2.3 Agisoft workflow

1.2.3.a Photo processing to digital elevation model (DEM) and orthomosaic imagery

The workflow within Agisoft PhotoScan was adapted from a selected set of guidance (Agisoft 2019; Mallison 2015; Dietrich 2015). Each flight area was processed separately (1-3 per study site). Every other photo from the nadir and oblique images were imported. This prevented distortion that can be created when there is too much overlap between the photos. A preliminary sparse point cloud was built using the common tie points of reference between photos. GCPs were imported and marked within applicable photos to georeference the point cloud. Where needed, additional markers were added to aid the alignment of adjacent photos. Gradual selection of the sparse points was performed to remove points that had high reprojection uncertainty, reprojection error or projection accuracy (Mallison 2015). The dense point cloud was generated and classified using maximum angles and distances between projected points (15 degrees and 1m, respectively) to assign them to the ground surface, rather than vegetation or buildings. A digital elevation model (DEM) was generated using only those points classified as ground. An orthomosaic image was generated using the DEM surface.

1.2.3.b Validation of Agisoft-generated DEM product

DEMs and orthomosaic images were exported from Agisoft into the geographic information system (GIS) ArcGIS as geotiff raster files. Systematic errors or distortions in the SfM surfaces were assessed using existing topographic data. LiDAR elevation data containing the respective DEM extents were imported from the National Elevation Dataset (The National Elevation Dataset (NED) 2016). The LiDAR -derived DEM was subtracted from the SfM DEM to calculate a DEM of difference (DoD) over the extent of each segment at the resolution of the respective LiDAR DEM (typically 1-2m). Resulting DoDs were assessed for systematic patterns of error. A tilted DEM was indicated by a positive or negative trend. A domed DEM was indicated by higher edges and a lower middle of the DEM, or vice versa, whereas step distortions occurred when part of the DEM was much lower or higher than another portion. Directional bias was revealed when the entire surface was much higher or lower than zero, and might be corrected by subtracting the mean absolute error (MAE) from the SfM DEM. Such systemic errors in SfM DEMs can produce inaccurate elevation profiles across the surface, leading to inaccurate representations of habitat parameters. Therefore, habitat metrics resulting from DoDs with evident systematic error were further scrutinized to assess whether the pattern of error appeared in results.

Final DEMs were resampled to the nearest 0.01m to obtain a consistent pixel resolution across sites. Where possible, DEMs from adjacent flights were stitched together to create one DEM per site, using the ArcGIS raster mosaic tool, selecting the minimum elevation in areas of pixel overlap. Resulting mosaics were assessed for severe changes in elevation along the edges of the individual panels. Where this occurred, individual survey area DEMs were used instead of the site-wide mosaics.

1.2.4 Channel habitat delineation

1.2.4.a Wetted channel edge and centerline, field reaches

The wetted channel extent was delineated manually as a shapefile using the orthomosaic images per site. Some interpolation was required where the channel edge was blocked by overhanging vegetation. Channel extent feature polygons were converted to raster polygons using maximum combined area and snapping the resulting raster to the resolution of the underlying DEM(s). Channel extent polygon features were converted to polyline features and split to represent each edge of the wetted channel individually. A channel centerline feature was calculated by generalizing wetted channel edge lines with ArcGIS Cartography Tools (ESRI 2019). Channel polylines were converted to raster lines representing the channel edge and center, again snapping to the underlying DEM. Field reaches were delineated for use in direct spatial comparisons between field and SfM data (see Chapter 2). Reaches were delineated from the coordinates representing the downstream end of the field reach to 75 or 100 meters upstream (Maryland and Pennsylvania sites, respectively), as measured along the centerline of the channel.

1.2.4.b Sediment mapping

Channel bed sediment patterns were mapped by manually outlining visible patterns of dominant sediment categories within the field reaches of the orthophotos. Categories were binned to account for visual approximation. For example, areas that appeared to be silt, but may be a mix of silt and sand, were called “silt-sand,” those that appeared to be sand or small gravel were “sand-gravel,” and so forth. Obvious demarcations between zones or major substrate types (e.g., large boulders, submerged logs) were outlined first. Where applicable, sediment that appeared to be exposed was outlined and labeled as exposed. Areas of the channel with overhanging trees or other vegetation were delineated as “exposed wood” or “exposed vegetation.” The sediment classes under these features were estimated based on the patterns evident surrounding the vegetation and throughout the reach.

1.2.4.c Distance-downstream zones

The stream centerline was used to identify spatially distinct zones approximating cross sections perpendicular to flow. The stream centerline (yellow pixels in Fig. 1.2.3) was used as a cost surface along which to calculate distance to the most downstream centerline pixel (dark yellow pixel in Fig. 1.2.3) and each successive centerline pixel. This resulted in cost distance to downstream values for each centerline pixel (left in Fig. 1.2.3). Cost distances were integerized to the nearest meter (right in Fig. 1.2.3), which created distinct, contiguous zones of pixels with the same distance to downstream (approximately 1-meter long each). Integerized values were allocated laterally across the channel to create spatially explicit zones approximately equivalent to channel cross sections (right in Fig. 1.2.3), assuming the closest orthogonal edge pixels were approximately the same distance from downstream (e.g., Carbonneau et al. 2012).

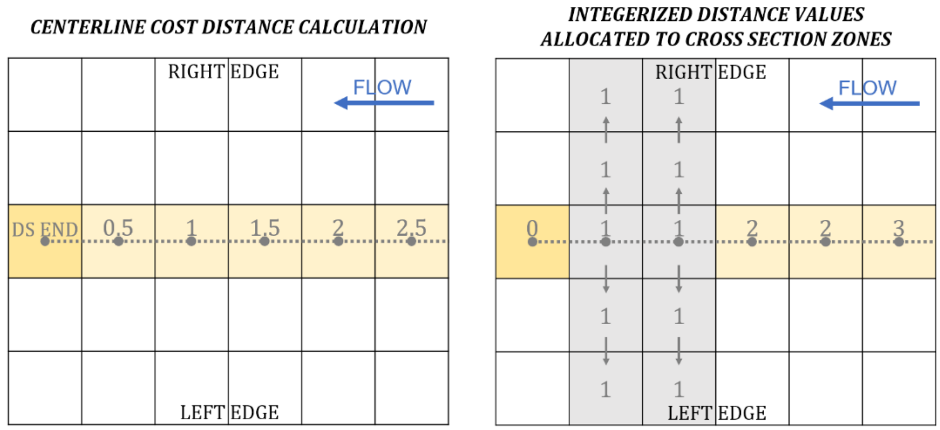


Figure 1.2.3. Depiction of cost distance to downstream if pixel resolution was $0.5m^2$. The left depicts the initial cost distance calculation per centerline pixel, and the right depicts the integerized distance values allocated to unique cross sections.

1.2.4.d Wetted width

The channel edge raster was separated into two raster lines representing the left and right bank of the stream. A Euclidean distance surface was calculated using each of these raster lines as the source cells. Figure 1.2.4 portrays blue numbers as Euclidean distance to the right edge, and orange numbers as Euclidean distance to the left edge. In any given cell, the two numbers added together give the same result of 4 cells wide (Fig. 1.2.4). While the depiction in the figure appears 5 cells wide, 4 cells is an appropriate approximation given that the centroid of edge cells should align with the wetted edge. Euclidean distance rasters were summed over the channel extent, with the resulting surface representing multiple approximations of channel width. Cross section mean width was calculated as an average across distance-downstream zones. Results were visually assessed for realistic spatial patterns and empirical distributions of values. Spatial pattern was assessed by fluctuations in width that corresponded with those evident in orthomosaic images. A cumulative frequency distribution (CFD) of width per site was plotted and the rank magnitude and range of width at each site was assessed within and among sites.

Summarized results of field surveys were also used to qualitatively assess the CFDs of widths derived from imagery.

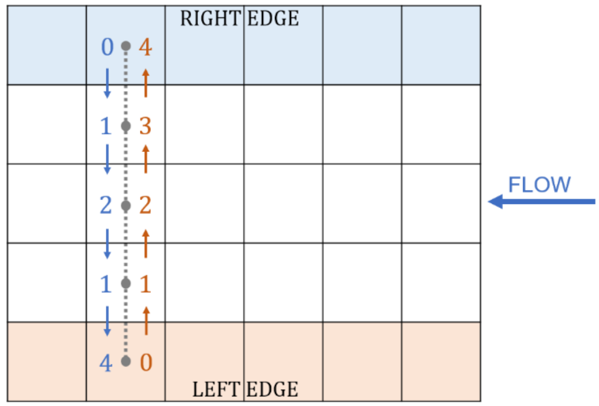


Figure 1.2.4. Depiction of Euclidean distance width calculation if pixel resolution was $1m^2$.

1.2.4.e Water surface elevation and depth

Water surface elevation was assessed using SfM DEM values extracted from each delineated channel edge raster, which assumed the water's edge fairly represented its approximate elevation across any given cross section. To account for the possibility that delineated edges might have different elevations, edge elevations were allocated across the channel and minimized over distance-to-downstream zones. The lower of the two edge elevations, if different, was expected to more accurately reflect the actual water surface, because delineated edges were considered more likely to erroneously represent exposed rocks, vegetation or steep banks than channel bottom. Nonetheless, longitudinal profiles from resulting water surfaces contained undulations (see Figure 1.3.4), so values were further smoothed using median values within an area equivalent to the median channel width. This focal radius further smoothed estimated water surface values over a limited objective extent. Where the focal analysis did not reach the edge of the channel, the surface elevation values were allocated across the remaining channel extent as a continuous estimate of the water surface.

The SfM DEM was subtracted from the water surface raster to obtain a continuous estimate of channel depth. Values <0m were set to null, which left some areas of channel depth unmapped. Cross section mean depth was calculated using the distance-to-downstream zones. Again, results were visually assessed for realistic spatial patterns and empirical distributions of values. Spatial pattern was assessed by comparison with channel features (e.g., water color changes). Again, a CFD of depth was plotted per site to assess the relative range and rank order of depths within and among sites. Summarized results of field surveys were used in conjunction with the CFDs to qualitatively assess the depth magnitudes derived from imagery.

1.2.4.f Cross section flow velocity

Cross sectional mean flow velocity was approximated assuming continuity among discharge, depth, width and flow velocity: $Q = wdv$, or, $Q=VA$ (Leopold and Maddock Jr. 1953). Using this equation assumed steady flow, and that discharge measured at the downstream outlet on the day of each aerial survey was representative of the approximate discharge throughout each associated study area. Distance-downstream zones were used to approximate cross-sectional means of width and depth, which were multiplied to obtain cross-sectional area (A). As before, results were visually assessed for realistic spatial patterns and empirical distributions. Spatial pattern was assessed by correspondence with velocity changes evident in the orthomosaic images. Only longitudinal fluctuations were assessed given that velocities were cross sectional averages, thus realistic patterns were represented by faster flow in shallow or turbulent water or slower flow in deeper water. Magnitude of mean cross-sectional velocity was assessed by plotting the cumulative frequency distribution of per site. CFD plots were qualitatively compared to summarized field values to assess velocity magnitudes relative to range and rank order.

1.2.4.g Flow velocity correction and outlier analysis

Continuity assumptions during velocity delineation were assessed by correcting discharge relative to drainage area continuously throughout each study site. Flow accumulation values were used as approximations of continuous drainage area throughout each reach. NED 10m DEMs were used to characterize flow accumulation using D8 flow direction (The National Elevation Dataset (NED) 2016). Each DEM was reconditioned using the AGREE algorithm to integrate NHD flowline (National Hydrography Dataset (NHD) 2017; ESRI 2019). Algorithm parameters were as suggested by Baker et al. (2006a), which used 150m for the reconditioning width, 10m as the smooth drop, and 1m as the sharp drop. The maximum flow accumulation that encompassed all study areas was found and used to reduce the flow accumulation surface to just an approximate channel line flow accumulation. These flow accumulation lines were multiplied by the pixel area ($100m^2$) so that each resulting pixel represented the drainage area in square meters. Because the NHD lines were not perfectly aligned with the stream lines in the study areas, the flow accumulation values were interpolated across the DEM using inverse distance weighting. Interpolated drainage area values underlying each cross-section pixel were then extracted.

A log-linear relationship between discharge (Q) and drainage area (DA) (e.g. Hack 1957) was then used to obtain a 'corrected' discharge per cross-sectional zone. As a coarse approximation, the log-linear slope between discharge and drainage area was assumed to be 1.

Table 1.3.3 depicts estimates of this slope for most sites, but these estimates were very approximate due to the distance (and drainage area difference) between gages and sites.

This process used the discharge measured at the downstream end (Q_{DS}) and the associated drainage area of that location (DA_{DS}), as well as the drainage area of a given cross section (DA_{XS}), to find the discharge at each given cross-section (Q_{XS}). First, the equation for the log linear slope could be separated, with a coarse approximation of 1:1 log-linear slope:

$$m_{Q-DA} = \frac{\log(Q_{DS}) - \log(Q_{XS})}{\log(DA_{DS}) - \log(DA_{XS})} = 1\left(\frac{Q}{DA}, \text{with units} \frac{m^3 s^{-1}}{m^2}\right) \tag{1}$$

$$\left(m_{Q-DA} * \log(DA_{DS})\right) - \left(m_{Q-DA} * \log(DA_{XS})\right) = \log(Q_{DS}) - \log(Q_{XS}) \tag{2}$$

Drainage area units (m^2) were eliminated, because each drainage area component was multiplied by the slope, which have drainage area units in the denominator:

Units of eq.(2): $\left(\frac{\log(m^3 s^{-1})}{\log(m^2)} * \log(m^2)\right) - \left(\frac{\log(m^3 s^{-1})}{\log(m^2)} * \log(m^2)\right) = \log(m^3 s^{-1}) - \log(m^3 s^{-1})$

Finally, the discharge of a given cross-section was separated from the rest of the equation, and the log removed by taking the exponential of each side of the equation:

$$\log(DA_{DS}) - \log(DA_{XS}) - \log(Q_{DS}) = -\log(Q_{XS}) \tag{3}$$

$$\log(Q_{DS}) - \log(DA_{DS}) + \log(DA_{XS}) = \log(Q_{XS}) \tag{4}$$

$$Q_{XS} = 10^{\log(Q_{DS}) - \log(DA_{DS}) + \log(DA_{XS})} \tag{5}$$

After cross-sectional mean velocity values were corrected for discharge relative to drainage area, extreme outliers were assessed and removed. Extreme outliers were identified per site as those that were less than the first quantile minus three times the interquartile range (IQR) or were greater than the third quartile plus three times the IQR. Before removing these outliers from the data, they were assessed relative to pixelated depth, and the proportion of depth pixels that contributed to each cross-sectional mean depth estimate.

1.3 Results

To effectively display the progression of the workflow, most figures throughout this section focus on one site, with the entire restored extent of site 144 portrayed in Figure 1.3.1 and 1.3.10, and the field reach within that restored extent portrayed in Figures 1.3.4, 1.3.6-8, and 1.3.11-13. However, it should be noted that these results are available for all 7 sites across the study extent, from upstream through to downstream of the restored extents. Other sites are presented where needed to display phenomena not present within the focal extent of site 144 (e.g., Figures 1.3.2-3 and 5).

1.3.1 Study design

Although field surveys were completed initially at all 8 sites included in the study, channel habitat metrics were only delineated for 7 sites. Therefore, the results of field surveys are only presented for the 7 included sites. The reference site ("RKGR") was removed from the study due to failure to generate a point cloud from orthoimagery. Remaining sites were mapped throughout their extents.

All sites were within the Piedmont physiographic province and represent first to fourth order using Strahler ordering, with upstream drainage areas ranging from approximately 3 - 26 km^2. Land cover in the upstream drainage areas ranged from 11 - 37% forested land and 12 - 74% developed land, with 2 - 18% impervious surface (Table 1.2.1).

1.3.2 Field Surveys

Cross-sectional discharge collected at the downstream end of each study area indicated sites 144 and EBCH had the smallest and largest discharges, respectively, on the day of the aerial survey and site 144 the lowest. (Table 1.3.1).

Table 1.3.1. Cross section (XS) discharge measurement characteristics, with upstream drainage area.

	Date Time	XS mean depth (m)	XS width (m)	XS flow area (m^2)	XS mean flow velocity (ms^{-1})	XS discharge (m^3s^{-1})	Drainage Area (km^2)
144	2/21/2018 14:13	0.253	4.877	1.146	0.016	0.0184	2.9
253	2/28/2018 18:16	0.132	6.706	0.883	0.0645	0.0569	8.8
CRAB	3/4/2018 15:31	0.208	3.962	0.825	0.106	0.0874	2.8
EBCH	3/4/2018 13:27	0.156	6.401	1.000	0.3434	0.3433	15
PLUM	3/30/2018 14:33	0.099	3.658	0.362	0.1773	0.0642	3.2
ULP2	1/29/2018 14:09	0.217	7.315	1.585	0.1125	0.1782	26.2
WL	2/28/2018 17:39	0.14	3.353	0.470	0.0896	0.0421	6.6

EMAP/SWAMP transects were assessed at PA sites between June 2016 and July 2018, where habitat parameters were measured (depth and width) or categorized (flow and sediment). Modified-MBSS field habitat assessments were completed at Maryland sites in August 2018 or January 2019, where habitat parameters were categorized within MBSS segments (depth, velocity and sediment) or measured at MBSS transects, at the upstream and downstream ends of the segments (depth, velocity and width). The range and mean per reach of the measured variables, and the proportion per reach of the categorized variables, are presented in Tables 1.3.2 (PA sites) and 1.3.4 (MD sites). The downstream (DS) reach at EBCH displayed the largest maximum and mean depth values, and the US reach displayed the largest maximum and mean width values (Table 1.3.2). CRAB was dominated by pool and riffle flows, while EBCH and PLUM were dominated by pools and glides (Table 1.3.2). Sites 144, CRAB, EBCH and PLUM had lower discharges during field surveys than during aerial surveys, while sites 253, ULP2 and WL had larger discharges during field surveys (Table 1.3.3). Maryland sites were dominated by the lowest depth and flow velocity categories, but sites 253 and WL also showed larger depths and velocities (Table 1.3.4). Sand and gravel were the most dominant sediment categories across Maryland sites (Table 1.3.4).

Table 1.3.2. EMAP / SWAMP transect measured results summarized as ranges and means (top), and categorized summarized as proportions per reach (bottom).

		CRAB		EBCH		PLUM	
		DS	US	DS	US	DS	US
Depth (m)	Min.	0.00	0.00	0.00	0.00	0.00	0.00
	Max.	0.50	0.27	0.54	0.49	0.42	0.19
	Mean	0.08	0.05	0.22	0.10	0.10	0.04
Width (m)	Min.	2.30	1.70	4.00	7.30	2.00	1.20
	Max.	4.20	7.10	6.10	12.60	7.10	4.90
	Mean	2.91	4.13	4.80	10.15	4.35	2.75
Flow Category (prop.)	Pool	0.39	0.49	0.42	0.32	0.37	0.28
	Glide	0.02	0.02	0.47	0.45	0.45	0.47
	Riffle	0.51	0.35	0.08	0.16	0.12	0.18
Sediment Category (prop.)	Silt					0.11	
	Sand	0.09	0.05	0.27	0.16	0.22	0.12
	Gravel	0.42	0.58	0.42	0.29	0.32	0.37
	Cobble	0.07	0.15	0.16	0.30	0.26	0.44
	Boulder/bedrock	0.30	0.18	0.03	0.17	0.08	0.07
	Wood/other	0.12	0.03	0.04	0.05	0.02	

Table 1.3.3. Discharge measurements (Q_m) and estimates (Q_e), field vs. aerial survey.

	144	253	CRAB	EBCH	PLUM, Rest	PLUM, US	ULP2	WL
Site Drn. Area (km^2)	2.9	8.8	2.8	15	3.2	2	26.2	6.6
Aerial Survey Date (ASD)	2/21/2018	2/28/2018	3/4/2018	3/4/2018	3/30/2018	3/30/2018	1/29/2018	2/28/2018
Site Q_m (m^3s^{-1}), ASD	0.018	0.057	0.087	0.343	0.064	0.064	0.178	0.042
Field Survey Date (FSD)	8/27/2018	1/2/2019	6/28/2017	6/30/2017	7/16/2018	6/29/2016	8/15/2018	1/3/2019
Site Q_m (m^3s^{-1}), FSD	0.006	0.114	-	-	-	-	0.281	0.116
USGS gage (GS) code	1650500	1650500	1473169	1481000	1481000	1481000	-	1650500
Gage Drn. Area (km^2)	54.4	54.4	54.1	743.3	743.3	743.3	-	54.4
Gage Q_m (m^3s^{-1}), ASD	0.337	0.337	1.209	15.376	11.723	11.723	-	0.337
Gage Q_m (m^3s^{-1}), FSD	0.170	1.045	0.643	5.154	7.844	5.437	-	0.977
Log-linear DrnArea-Q slope	0.992	0.977	0.887	0.974	0.956	0.956	-	0.986
Site Q_e (m^3s^{-1}), FSD	0.009	0.176	0.046	0.115	0.043	0.019	-	0.131
Field / aerial Q	0.337	2.011	0.532	0.335	0.669	0.302	1.579	2.762

Table 1.3.4. MBSS segment categorized results summarized as proportions per reach (top), and MBSS transect measured results summarized as ranges and means per reach.

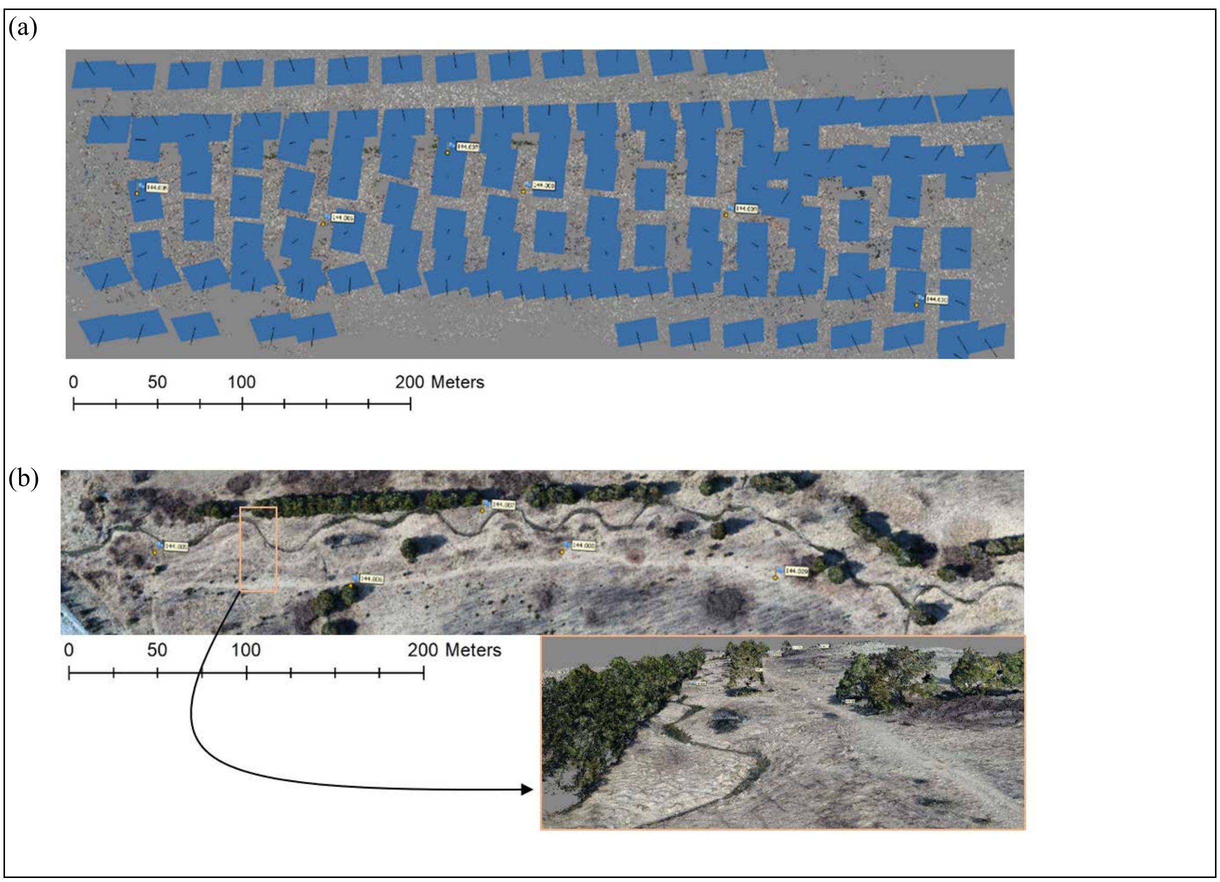
		144			253			ULP2			WL		
		DS	Rest	US	DS	Rest	US	DS	Rest	US	DS	Rest	US
Depth Category	1 (0-0.5m)	1.00	0.83	1.00	0.67	0.67	0.50	1.00	1.00	1.00	1.00	0.33	0.50
	2 (0.5-1.0m)		0.17		0.33	0.33	0.33					0.50	0.50
	3 (>1.0m)						0.17					0.17	
Flow Category	1 (0-0.3m/s)	1.00	1.00	1.00	0.67	0.50	0.67	1.00	1.00	1.00	0.33	0.50	0.50
	2 (>0.3m/s)				0.33	0.50	0.33				0.67	0.50	0.50
Dominant Sediment Category	Cobble		0.33		0.33		0.17						
	Gravel	0.83	0.33		0.50	0.67	0.50			0.17	0.67	0.17	0.67
	Sand	0.17	0.33	1.00	0.17	0.33	0.33	1.00	1.00	0.83	0.33	0.83	0.33
Subdominant Sediment Category	Boulder		0.33										
	Cobble	0.50				0.50	0.50		0.33			0.50	0.83
	Gravel	0.17	0.33	0.67	0.50	0.33	0.33	0.67	0.67	0.50	0.33	0.33	0.17
	Sand	0.33	0.33		0.50	0.17				0.17	0.67		
	Silt			0.33			0.17	0.33		0.33		0.17	
Width (m)	Min.	2.20	1.50	0.50	3.63	5.18	4.69	4.10	5.10	6.90	3.20	3.35	2.70
	Max.	5.40	2.70	2.70	5.67	7.80	9.60	8.80	6.70	10.50	5.85	5.90	4.70
	Mean	3.83	2.08	1.73	4.63	6.16	7.13	6.63	5.95	8.20	4.06	4.50	3.53
Thalweg depth (m)	Min.	0.06	0.06	0.09	0.003	0.002	0.004	0.26	0.31	0.45	0.14	0.18	0.18
	Max.	0.47	0.47	0.22	0.49	0.64	0.34	0.78	0.33	0.51	0.35	0.70	0.41
	Mean	0.29	0.24	0.18	0.18	0.24	0.12	0.50	0.32	0.48	0.23	0.39	0.24
Thalweg velocity (ms^{-1})	Min.	0.01	0.01	0.02	0.00	0.05	0.01	0.21	0.25	0.09	0.12	0.06	0.05
	Max.	0.18	0.24	0.20	0.36	0.66	0.48	0.40	0.30	0.85	0.53	0.14	0.23
	Mean	0.07	0.11	0.07	0.18	0.29	0.15	0.29	0.27	0.31	0.30	0.10	0.14

1.3.3 Agisoft workflow and product validation

The Agisoft workflow was completed for the 7 restored study sites. The reference MBSS sentinel site remained unstudied because alignment of the point cloud was not possible. Each study site had 1-3 separate flight areas, which were processed as separate chunks in the Agisoft environment. Figure 1.3.1 displays examples of Agisoft-generated products for the middle portion of site 144. For the sparse point cloud, the position and orientation of the orthophotos are portrayed by the blue boxes (Figure 1.3.1a). To create this point cloud, these photos were preliminarily stitched together based on common visual tie points. The sparse point cloud was then enhanced to a dense point cloud after adding common georeferenced ground control points (GCPs), indicated spatially by the labeled blue flags, into stitched images (Figure 1.3.1b).

Overlapping imagery was stitched together to create a continuous orthomosaic image for each study segment (Figure 1.3.1c). Interpolation between dense points generated an elevation surface (DEM; Figure 1.3.1d), which was used in subsequent analysis.

Patterns of difference in elevation were assessed within the DoD for signs of systematic error in the SfM DEM. For the site 144 example, there was little to no systematic error evident, as most of the difference was ±0.1m (yellow in Figure 1.3.1e). Lighter green areas of the DoD, where the SfM DEM was higher than the LiDAR DEM, included trees and other high vegetation that were not removed from the SfM DEM, but were not included in the LiDAR DEM. Darker green areas in the DoD, where the SfM DEM was lower, were areas near or within the channel, either where the channel extent may have changed between the time the two DEMs were created, or where the SfM picked up bathymetry that LiDAR represented as a flat surface.



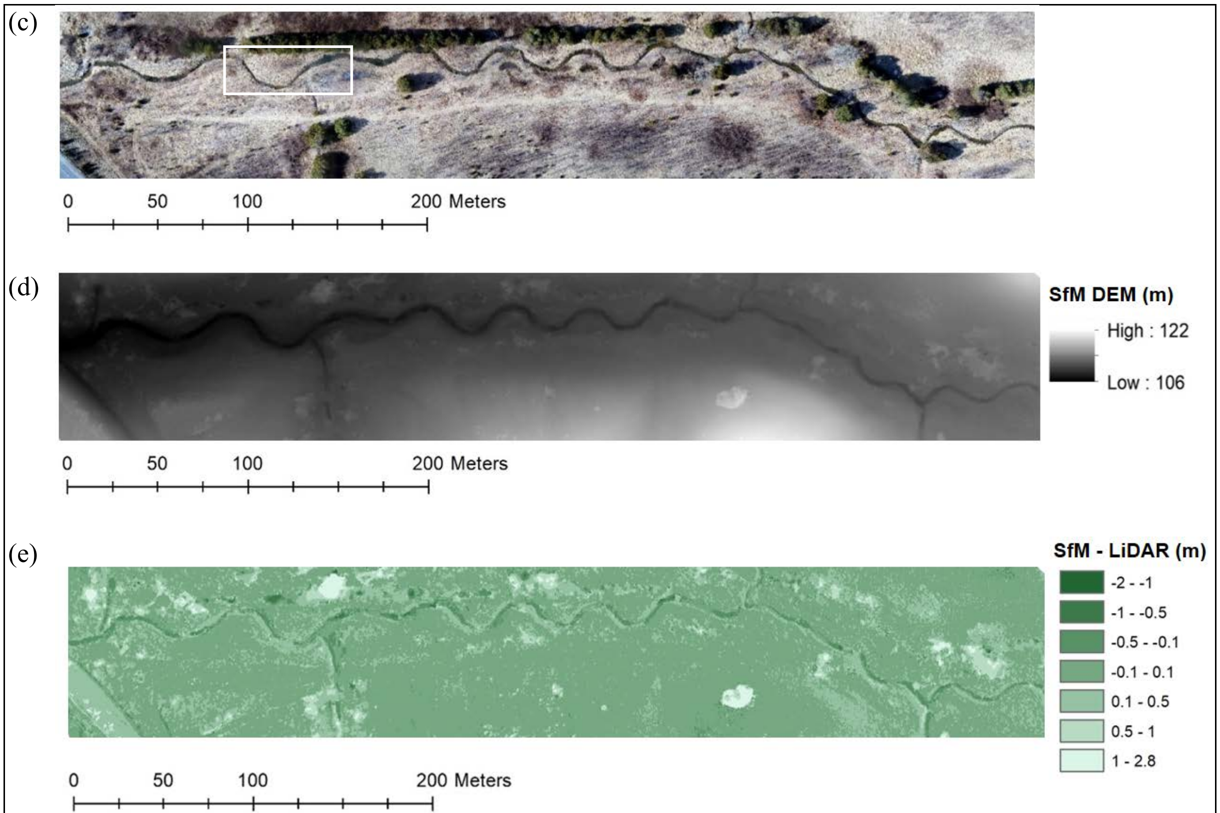


Figure 1.3.1. Agisoft products and validation for the middle section of site 144: (a) sparse point cloud with depiction of photo location and orientation, and marked GCP locations; (b) dense point cloud, with depiction of marked GCP locations, and three-dimensional perspective of the point cloud; (c) orthomosaic image, with the field reach extent used Figures 1.3.4-7, 9-10, 12 outlined in white; (d) DEM; and (e) DoD.

For some flight areas, a combination of UAV GPS issues, and heavily forested canopy complicated the point cloud generation process. For one site, the "unstudied reference site," these issues were so extensive, that a point cloud and successive DEM and orthomosaic photos could not be created. At one site, #253 in Maryland, UAV GPS issues led to difficulty in point cloud alignment. The UAV geotags each photo it takes with GPS information. However, calibration issues or damage to the UAV can lead to imprecise GPS information associated with the photos. Photos were "unchecked" prior to point cloud generation to try to circumvent this issue. Unchecked orthophotos within Agisoft should remove the influence of any GPS errors attached to the photo metadata and create the point cloud using GCP elevation and the relative orientation and position of the photos. Similarly, manual tie points were added into overlapping photos to aid in alignment. Despite these efforts to improve the alignment of photos and projection of points, the DEM resulting from the point cloud still displayed extensive warping and distortion, evident by the DoD (Figure 1.3.2). Although results were nonetheless derived for this site, they were assessed with caution given this error.

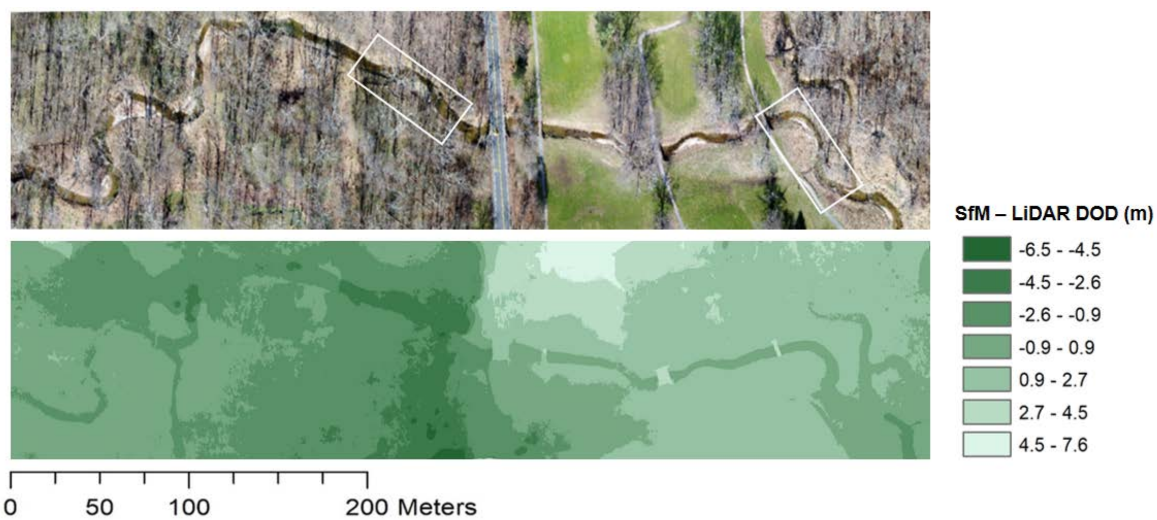


Figure 1.3.2. Orthomosaic image and DoD over the downstream extent of site 253. The DoD between the SfM and LiDAR DEMs indicated a warped SfM DEM. The restored and downstream field reaches are indicated with white boxes within the orthoimage.

Systematic error was evident within the DoDs at a few other study sites, but not to the extent or severity as site 253. As evident in Figure 1.3.3, the DEM for the upstream area of PLUM showed evidence of a slight N-S tilt. This type of error indicated the need for additional GCPs. However, error around the channel was between -0.5 and 0.5m different, so the DEM and orthoimagery were used to derive results.

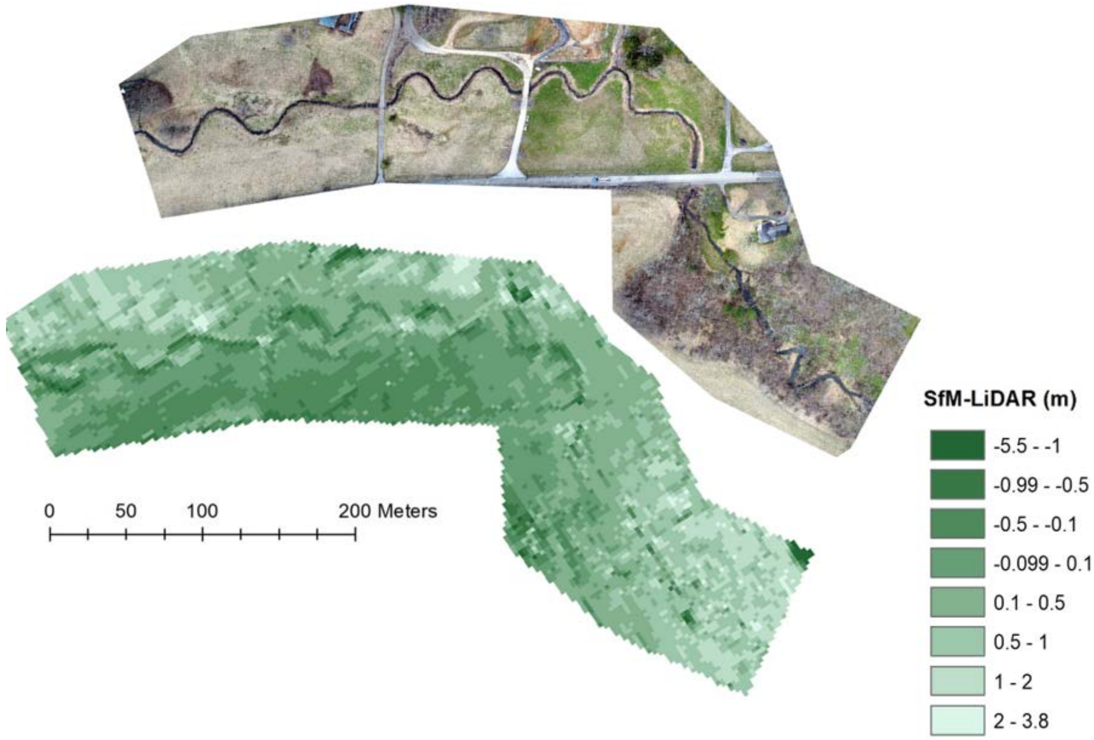


Figure 1.3.3. Orthomosaic image and DoD over the upstream extent of site PLUM. The DoD between the SfM and LiDAR DEMs indicated a tilt in the SfM DEM.

1.3.4 Channel habitat delineation

1.3.4.a Wetted channel edge and centerline, field reaches

The wetted channel edge captured the boundary between the wetted channel and the bank, which enabled the delineation of the channel centerline, field reaches and successive channel extent rasters. These features are depicted for the field reach within the restored area at site 144 in Figure 1.3.4.

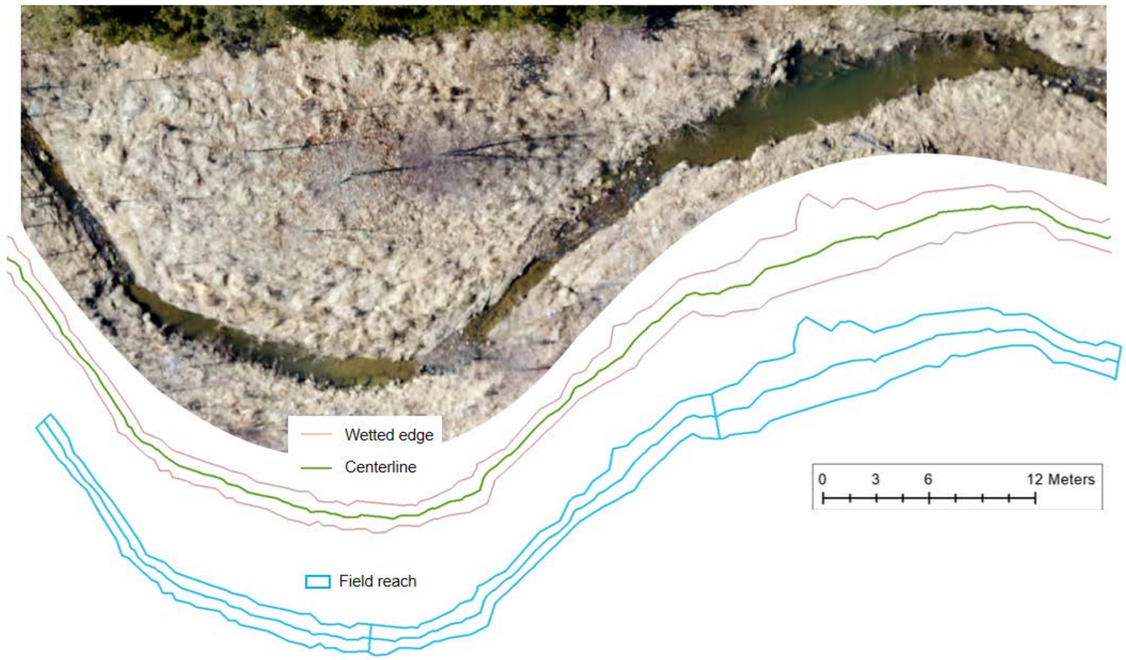
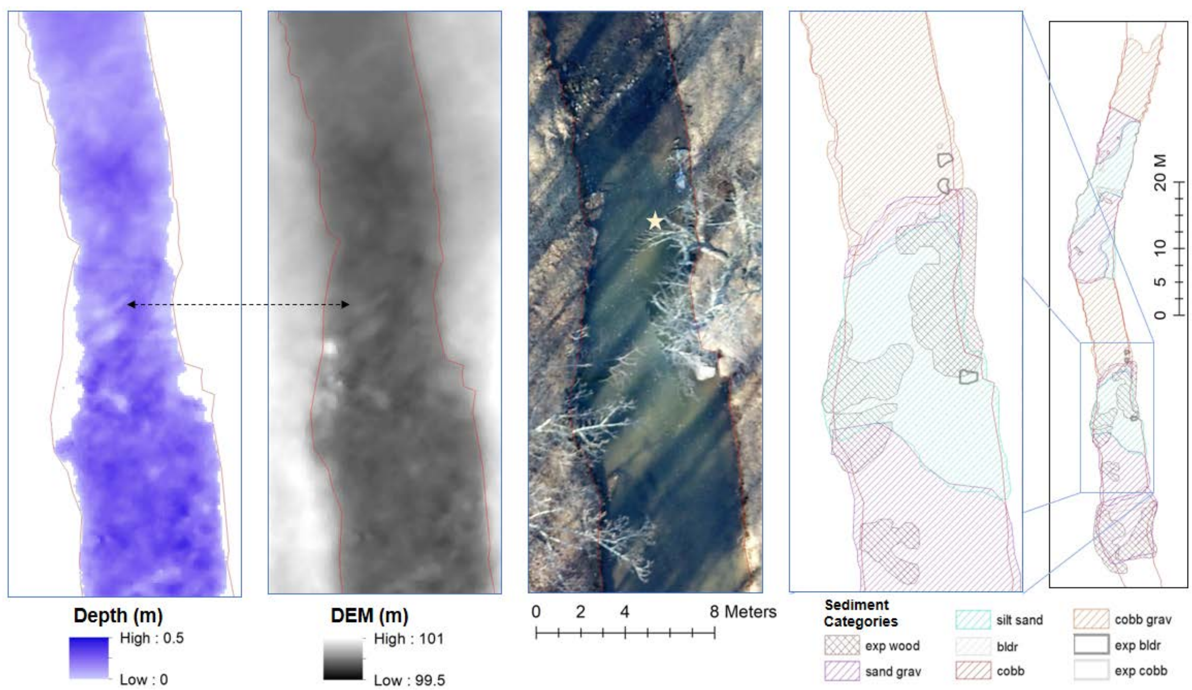


Figure 1.3.4. Orthomosaic image at site 144 over the extent of the field reach within the restored area. Over the same extent, the extracted wetted channel edge, centerline, and field reach segments (MBSS site), are depicted below the orthoimage.

Certain areas of orthomosaic images revealed that, even overlapping imagery from multiple view angles, factors such as vegetation, shadows, deep water, or blur could nonetheless block the view of the channel edges and/or bottom (e.g., Figure 1.3.5c). Best judgement interpolation was used to delineate channel margins in these areas. In particular, where vegetation obstructed view of the channel in the orthoimagery, it may also be reflected in the water surface and interpreted as ground in the DEM, which could lead to overestimation of channel bed elevation. With areas of error in both orthoimages and DEMs, resulting habitat metrics may be misrepresented. However, as evident from the DEM in Figure 1.3.5b and the depth map in Figure 1.3.5a, some variation in channel bed elevation was detected even where orthoimages lacked obvious visual representation of that variation. The orthoimage in Figure 1.3.5c contained a diagonal pattern of shadows, which was not propagated to fluctuations in elevation throughout the DEM in Figure 1.3.5b. However, there was at least one place where shadows appeared to influence relative elevations, and therefore depths. In the middle of the extent shown in Figure 1.3.5a-c (indicated with dashed arrow), a portion of the diagonal striped pattern appeared in the depth and DEM map associated with two bright streaks surrounding a slightly darker streak in the orthoimage.



(a) (b) (c) (d) (e)

Figure 1.3.5. Channel detail within the restored field reach of ULP2, displaying vegetation overhang, image stitching error and blur, and shadows, each of which complicate automated wetted area and sediment extraction. Manual channel edge extraction is represented by red lines. (a)-(d) display the depth, DEM, orthoimage, and sediment categories, respectively, within the lower portion of the reach. (e) displays the whole extent of the field reach with mapped sediment categories, for comparison. The double-headed arrow between (a) and (b) points to potential shadow interference in the DEM and depth maps. The star in (c) lies on a straight edge within the image showing possible image stitching error.

1.3.4.b Sediment mapping

Approximate patterns of major sediment categories were mapped within reaches using orthomosaic imagery. The mapped sediment for the restored field reach at site 144 is depicted in

Figure 1.3.6. There is some subjectivity in the placement of major sediment category edges. However, the delineation captured both the pattern of sediment apparent in the orthoimagery (e.g., coarser sediment is evident in shallower water; Figure 1.3.6), and the categories found in the field survey (e.g., Table 1.3.4 shows site 144-Rest had dominant sediment categories of sand, gravel and cobble).

As with wetted edge delineation, the accuracy of sediment mapping was limited by the quality of the orthomosaic imagery. In some channel areas, shadows, image stitching error, blur and/or overhanging vegetation affected the ability to see the channel sediment. This was evident in Figure 1.3.5, where the pool in the middle of the orthoimage (Fig. 1.3.5c) is obscured by overhanging tree branches and blurred by the way the images were stitched together (an edge of this phenomenon is denoted with a yellow star in Fig. 1.3.5c). However, some sediment was still visible (e.g., large boulders on the boundaries). If the obscured area was relatively small, it was assumed that the sediment category could be interpolated from surrounding visible features. For instance, as displayed in the middle of Figure 1.3.5c-d, the silt-sand was visible in the other deep pool portion of the reach (Fig. 1.3.5e), and the approximate boundaries of the sediment category could be seen on either side of a small blurred area, so the sediment category ("silt sand" in this case) was assumed to persist through the blurred area.

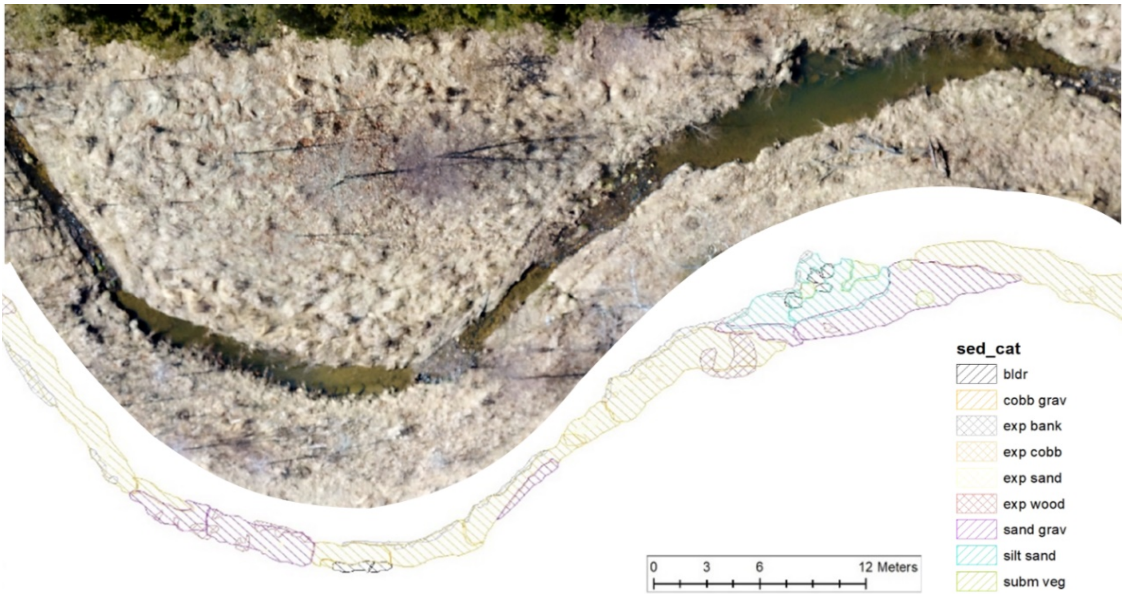


Figure 1.3.6. Major sediment categories mapped for the extent of the field reach within the restored area at site 144.

1.3.4.c Distance-downstream zones

Integerized cost-distance to the downstream end of each study reach along the channel centerline was allocated across the channel to distinguish distinct zones approximately equivalent to channel cross sections. These zones were not perfectly perpendicular to flow in all areas of the channel, especially where the channel edges were not perfectly parallel (e.g., the top right of Figure 1.3.7). However, the zones did seem to adequately represent distinct, approximately 1-meter wide sections of the channel that were representative of its lateral width at a given location.

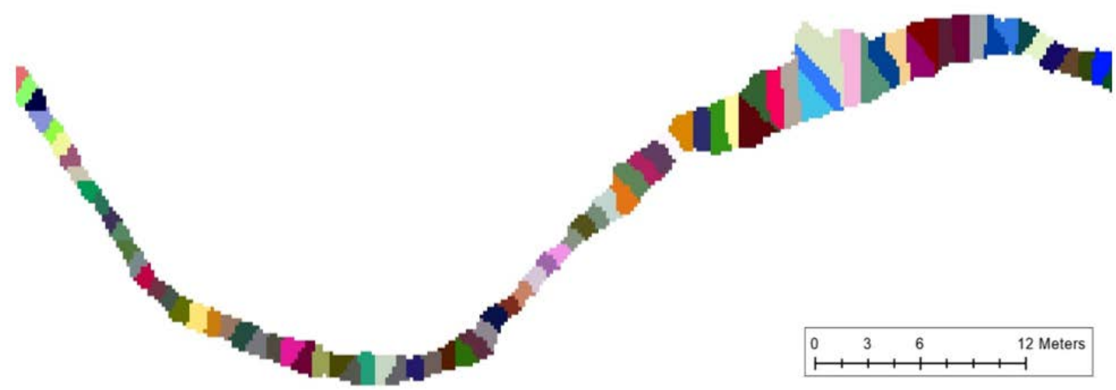


Figure 1.3.7. Integerized cost-distance-to-downstream zones for the extent of the field reach within the restored area at site 144.

1.3.4.d Wetted width

Wetted channel width was approximated by summing the Euclidean distances to each edge for every channel pixel. This resulted in a continuous map of wetted width that appropriately fluctuates with the constriction and expansion of the channel, as evident in the example portrayed in Figure 1.3.8b. Again, there is some evidence that non-parallel edges lead to approximations of width that slightly differ laterally across the channel. This would lead to slightly different measures of channel width than might be approximated in the field. Cross-sectional mean width (e.g., Figure 1.3.8c) might more closely match what is measured in the field at a given location, as it is more consistent laterally across the channel.

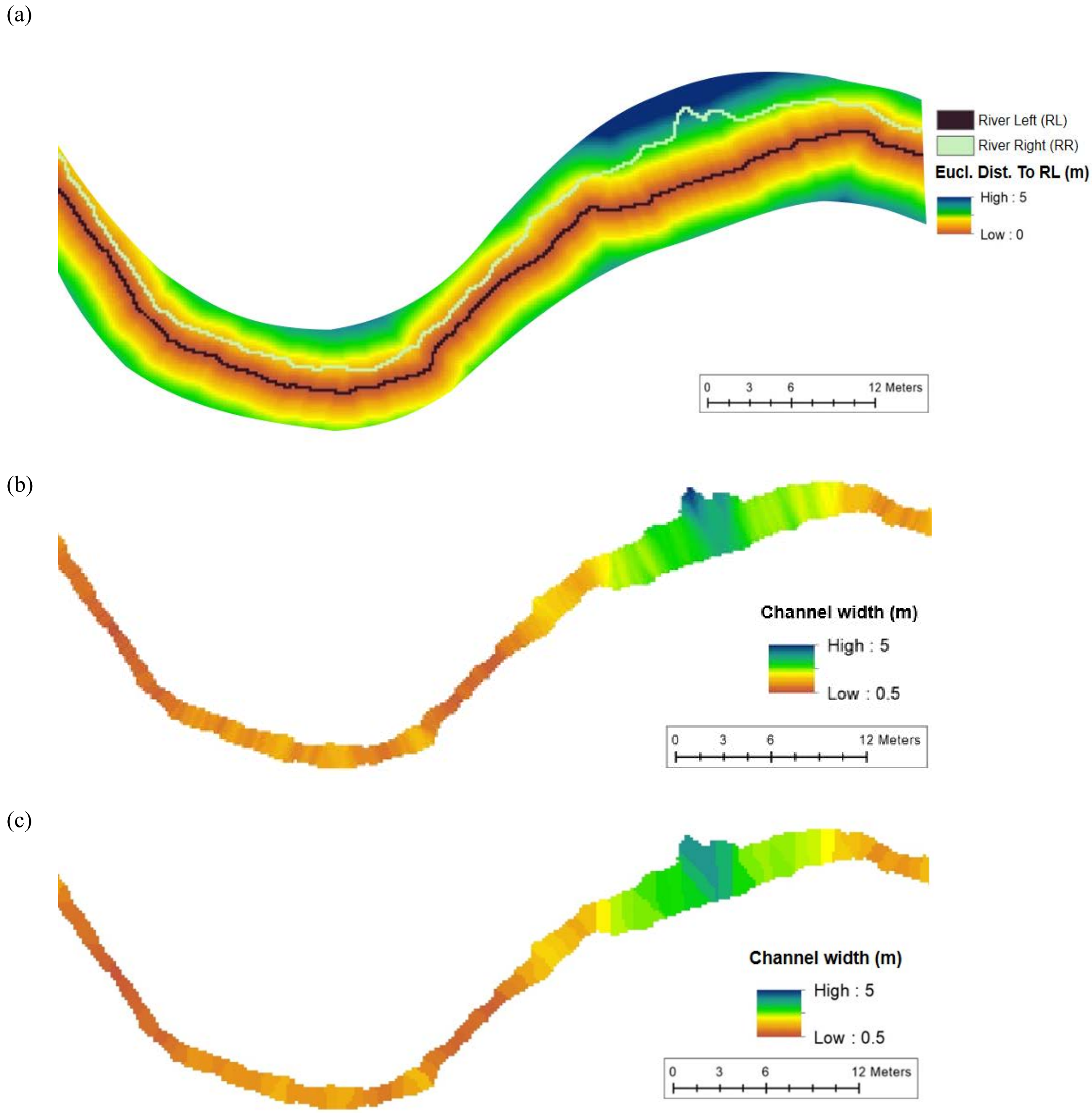


Figure 1.3.8. Continuous channel width approximations for the field reach extent within the restored area of site 144. Pixelated channel width (b) was approximated by adding the Euclidean distance to each edge pixel (e.g., distance to the left edge portrayed by (a)), across the channel extent. Distance-to-downstream cross section zones were used to calculate cross-sectional average width (c).

Magnitudes of width at each site met expectations relative to field surveys. Figure 1.3.9 displays CFDs of continuous (a) and cross-sectional mean width (b). Width varied between approximately 0.5-13m. This range and approximate rank order was corroborated by the field data, where site 144-US had a minimum field width of 0.5m (Table 1.3.4), and EBCH-US had a maximum width of 12.6m (Table 1.3.2). The CFDs also showed the distribution of widths at site 144 was consistently lower than all others, with almost 100% of continuous and cross-section mean values <5m (red line in Figure 1.3.9). This is consistent with the field width values in Table 1.3.1 and 1.3.4, and in that 144 had the lowest mean widths per reach of the sites in Tables 1.3.2 and 1.3.3. Similarly, the distribution of widths at EBCH was higher than others, with approximately 80% of continuous and cross-section mean values >5m (green line in Figure 1.3.9). This pattern was again consistent with field values, in that EBCH had the highest mean width per reach of the sites in Table 1.3.2 and 1.3.4.

A noteworthy feature of the CFDs for EBCH and ULP2 were the slight breaks in the distributions near 7m (green and purple lines, respectively, in Figure 1.3.9). For EBCH and ULP2, these breaks likely occured due to differences between reach extents within the sites. In field data for EBCH, the mean width for the restored field reach was 4.8m, whereas the US field reach was 10.2m (Table 1.3.2). Breaks in the CFD for EBCH around 7.5m probably indicated that the majority of values within the restored reach were between 2-7.5m, whereas the majority of values within the US reach were between 7.5-13m.

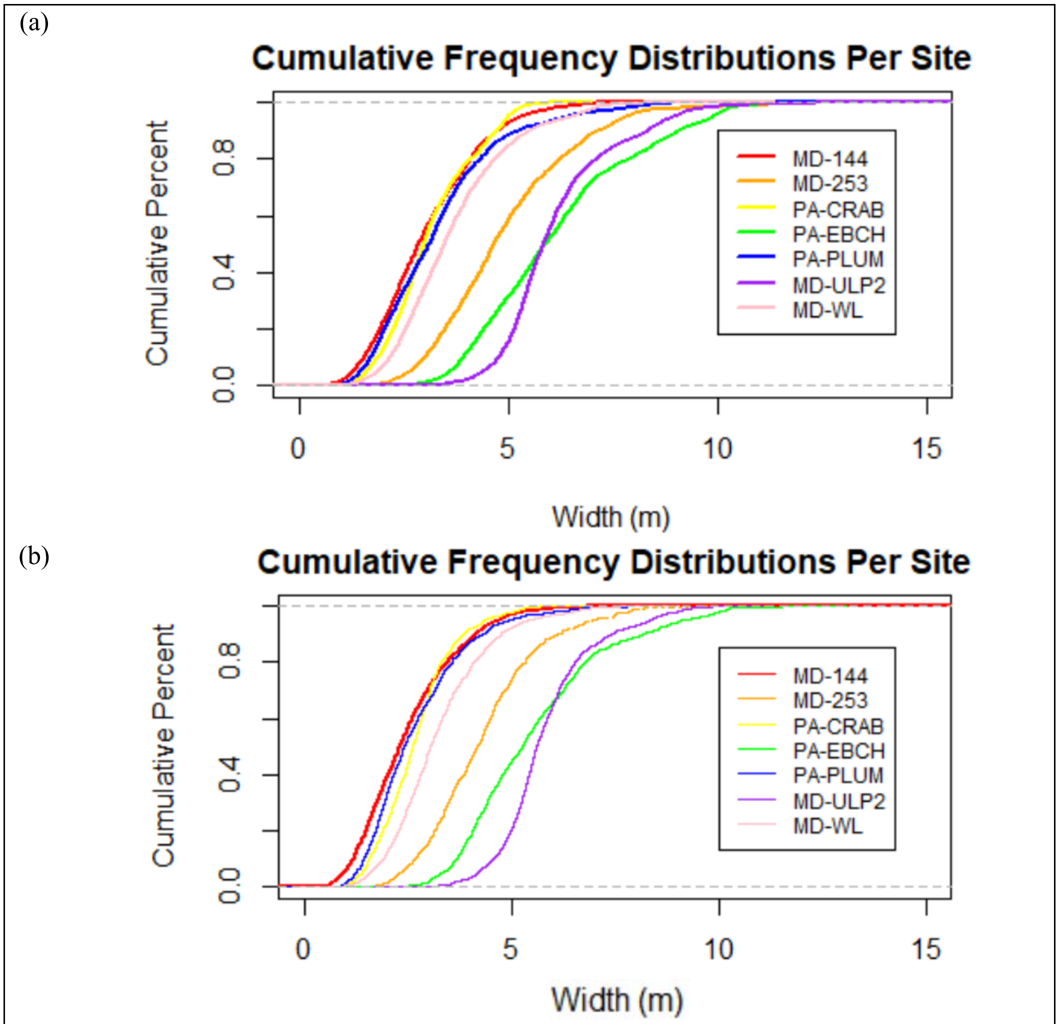


Figure 1.3.9. Cumulative frequency distributions of (a) continuous width values and (b) cross-section mean values at study sites.

1.3.4.e Water surface elevation and depth

A continuous water surface elevation was derived for each study site. Figure 1.3.10a displays an example of the derivation via minimization and focal median smoothing of wetted edge elevation values, and Figure 1.3.10b depicts a portion of this profile in planform. Flow was from right to left in the image, so the water surface elevation was expected to be higher on the right and lower on the left. This planform segment shows that the water surface elevation varies approximately 0.5m vertically over approximately 50m of the channel, which represents a slightly higher slope (*c.* 0.01) than the approximate slope over the entire extent displayed in Figure 1.3.10a, (which varies approximately 3.3m vertically over 400m, or *c.* 0.00825 slope). In Figure 1.3.10b, blue arrows point to the relatively flat water surface in pools. Red-orange arrows point to the transition from pool to riffle, and yellow arrows to elevation decreases through riffles. Black arrows point to erroneous changes in elevation. For instance, the black arrow on the left side of the image points to where the delineated channel edge may have erroneously been placed on the bank or captured exposed overhanging vegetation, and even through smoothing, the erroneous elevation increase remained. Similarly, the black arrow on the right side of the image points to where overhanging vegetation was likely the cause of an increase in elevation, right of a perhaps erroneous decrease in elevation where the channel edge may have been placed in the wetted area (Fig. 1.3.10b). However, the planform showed an expected pattern of higher elevation in the transition from pool to riffle (red arrows), decreasing elevation throughout the riffle (yellow arrows), and then lower elevation in the next pool (blue arrows; Fig. 1.3.10b).

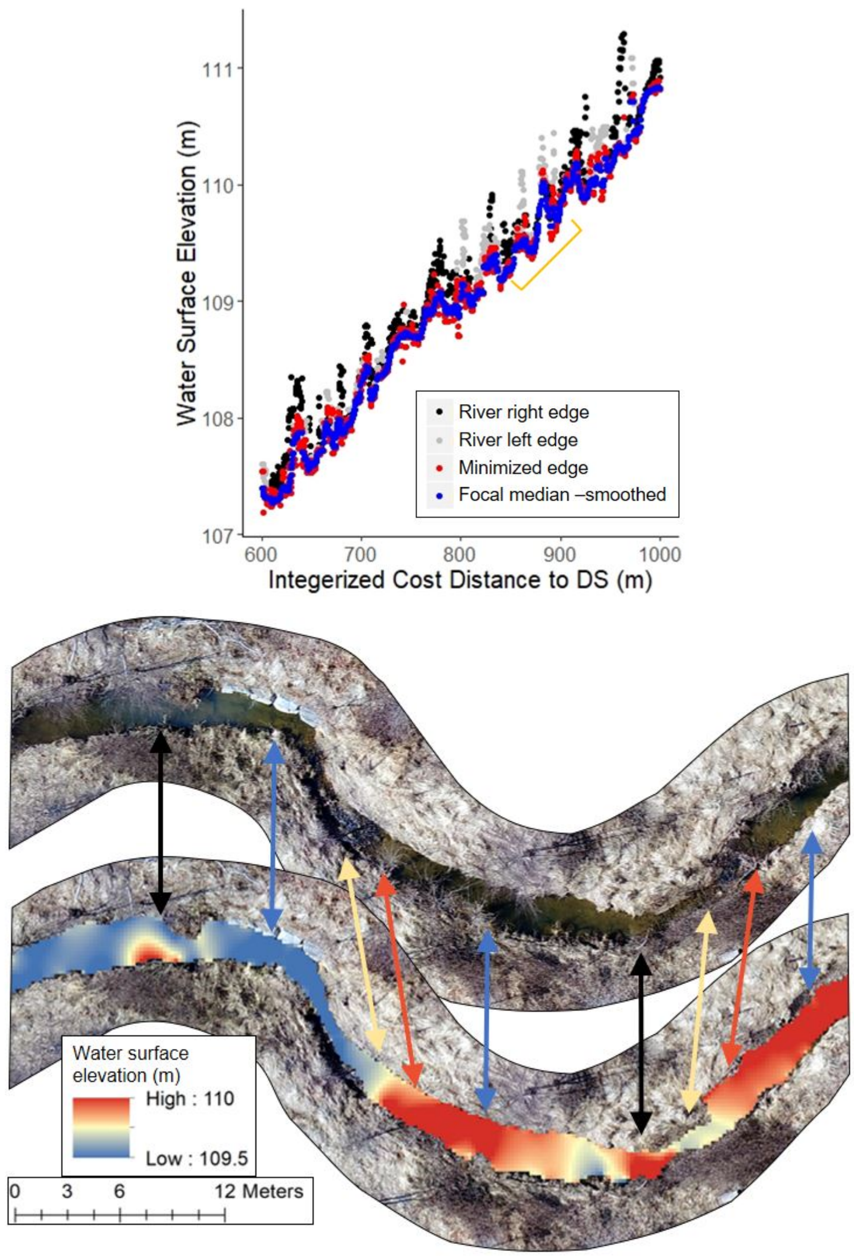


Figure 1.3.10. Water surface elevation. The longitudinal profile between 600-1000m to downstream at site 144 (a), displays the effects of smoothing, Between approximately 850-900m to downstream (indicated by the orange bracket in (a)), the smoothed water surface is portrayed in planform (b), with the choropleth stretched to represent only the range of values within that area, and with arrows indicating changes in elevation associated with visible features.

The DEM within the channel area was subtracted from the water surface elevation to approximate continuous water depth. Figure 1.3.11b shows a portion of a channel DEM. The elevations throughout this portion appeared to be appropriately lower than the elevations throughout the water surface mapped above it (Fig. 1.3.11a-b). However, it was evident that there was some noise in upper right of the channel DEM, indicated by the lighter areas, which represent higher elevations (Fig. 1.3.11b). These areas – where the channel elevation was higher than the water surface elevation – were interpreted as null values in the depth map, as shown in Figure 1.3.11c. The depth delineation method properly captured the alternating shallow-deep patterns of depth visible in the orthoimagery. Patterns of varying water depth were similarly represented by cross-sectional means (Fig. 1.3.11d), which portrayed fluctuations that mimicked patterns evident in the orthoimagery (e.g., darker water indicating larger depths; Fig. 1.3.11d).

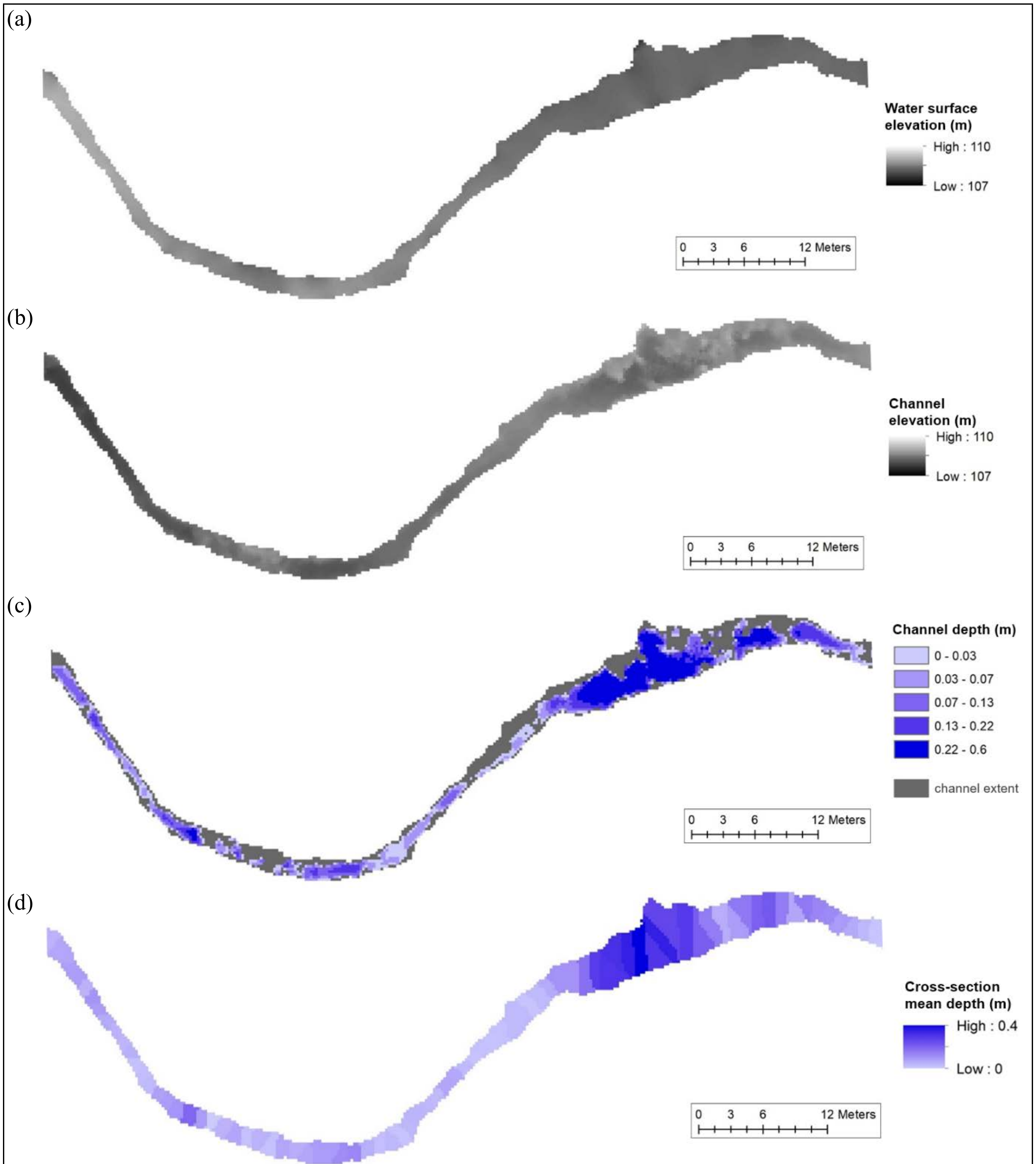


Figure 1.3.11. Water surface elevation (a), channel elevation (b), pixelated water depth, with null (<0m depth) values within the channel indicated in dark grey, and cross-section mean water depth (d) for the field reach extent within the restored area at site 144.

Magnitudes of depth at each site were similar to field surveys. Figure 1.3.12 displays the cumulative frequency distributions (CFDs) of pixelated (a) and cross-section mean (b) depth. Depth varied continuously between approximately 0-1m. This range was reflected by the field data. Zero depth within the wetted channel was present at all sites, as evidenced by the 0m minimum values at all Pennsylvania reaches (Table 1.3.2), and the prevalence of depth category 1 (0-0.5m) at Maryland sites (Table 1.3.4). The CFDs for Maryland sites 253, ULP2, WL, and Pennsylvania site EBCH had greater proportions (25-50%) of their depth distributions above 0.2m, as evidenced by the orange, purple,pink and green CFD lines, respectively (Fig. 1.3.12a). This was corroborated by the field data as these three Maryland sites also had large portions of their reaches categorized as >0.5m (Table 1.3.4), whereas EBCH was the deepest of the three Pennsylvania sites (Table 1.3.2). The CFDs of Pennsylvania sites CRAB and PLUM and Maryland site 144 show that approximately 80-90% of their depth distributions were <0.2m (yellow, blue and red lines, respectively, in Figure 1.3.12). This was matched by field observations, where CRAB and PLUM both show mean depths for both of their respective reaches ≤0.1m (Tables 1.3.2), and where 144 was predominantly assigned to depth category 1 (0-0.5m) for all reaches (Table 1.3.4). The CFDs for pixelated depth values (Fig. 1.3.12a) were smoother than those of cross-sectional mean depth (Fig. 1.3.12b) because each curve represented a step function of cross-sectional means rather than a continuously varying estimate.

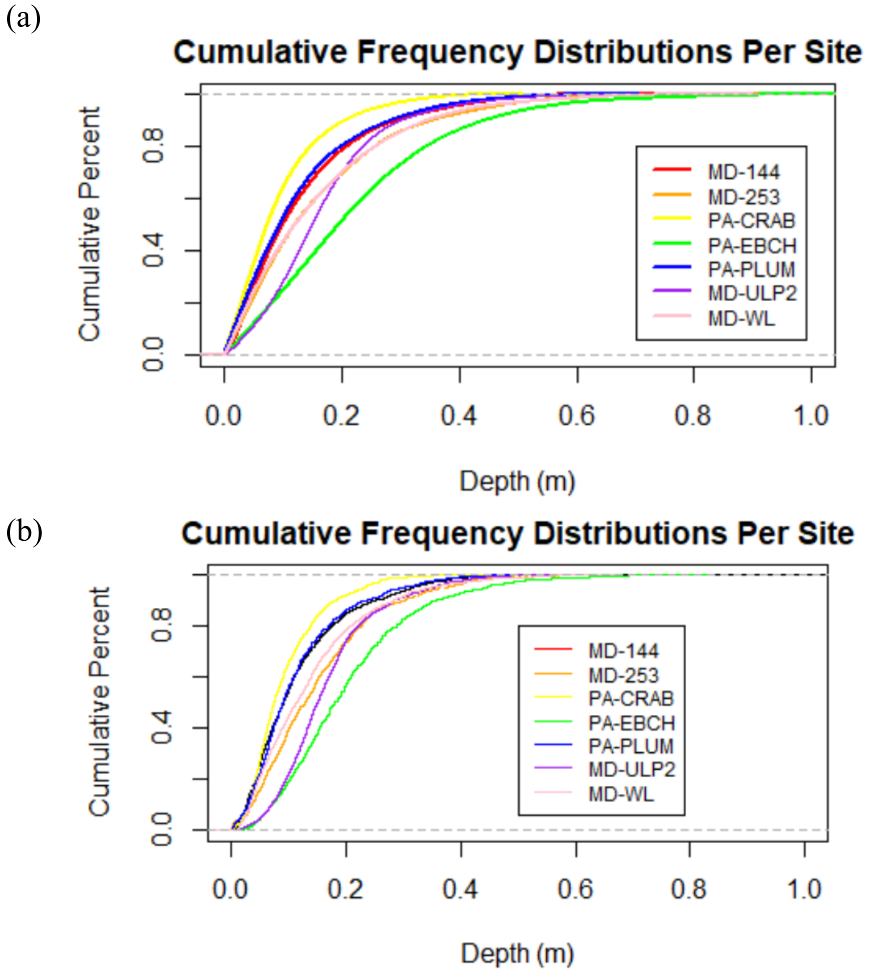


Figure 1.3.12. Cumulative frequency distributions of pixelated depth values (a) and cross-section mean values (b).

1.3.4.f Cross-section flow velocity

Cross-section mean flow velocity (V) at each study site was calculated as the quotient of the cross-section discharge (Q) at the downstream end of the study area and the continuous approximation of cross-section flow area (A; e.g., Q/A = V; Leopold and Maddock Jr. 1953). The spatial pattern of the resulting velocity maps appropriately displayed patterns of increasing velocity in areas of channel constriction and decreasing in areas of channel expansion, as displayed for a portion of site 144 in Figure 1.3.13. These patterns were also congruent with the pattern of pools and riffles visible in the orthoimagery, where slow velocities would be expected in the pools in the upper right and lower left portions of the channel in Figure 1.3.13, and fast velocities would be expected in the riffle in the middle and end portions of channel in Figure 1.3.13.

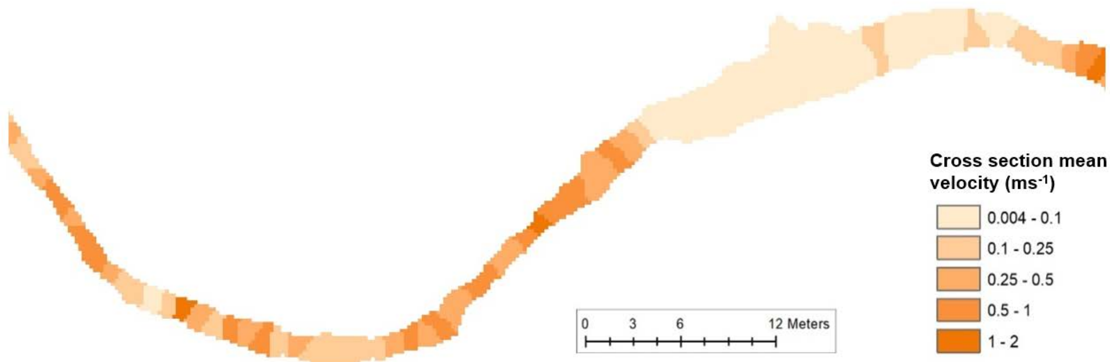


Figure 1.3.13. Cross section mean velocity over the extent of the field reach within the restored area of site 144.

Figure 1.3.14a-b display the CFDs of cross-section mean velocity. These CFDs displayed clear outliers in the maxima of the site ranges (Fig. 1.3.14a). However, central tendencies and site-based rank order of cross-section mean flow velocities were as expected relative to what was found in field surveys (Fig. 1.3.14b). Slow velocities were present at all sites, as evidenced by the assignment of the pool category to approximately 30-50% of reaches at Pennsylvania sites (Table 1.3.2), and the prevalence of flow category 1 (0-0.3ms^{-1}) at Maryland sites (Table 1.3.4). The CFDs for Maryland sites 144 and 253 had higher percentages of their distributions (70-90%) below approximately 0.3ms^{-1}, as evidenced by the red and orange CFD lines being higher than others below this value (Fig. 1.3.14a-b). The CFDs for Pennsylvania sites CRAB and EBCH had higher percentages (~50%) of their distributions above approximately 0.5ms^{-1}, as evidenced by the green and yellow CFD lines being lower than others (Fig. 1.3.14). This was matched by the field data where CRAB had 30-50% of its reaches categorized as riffle (Table 1.3.2), and EBCH had the highest measured mean velocity (Table 1.3.1). However, up to 10% of the PLUM CFD was >2ms^{-1} (Fig. 1.3.14b), which likely indicated outliers and/or error in the assumption of continuity (e.g., that discharge was unlikely to be the same throughout the study extent). With approximately 80% of values <1ms^{-1}, and 90% <1.5 ms^{-1}, the distribution seemed somewhat appropriate relative to field observations. However, drainage area correction of velocity, as well as outlier removal, improved distribution ranges, and made them closer to what was expected relative to field values (Fig. 1.3.14c, and see section 1.3.4.g).

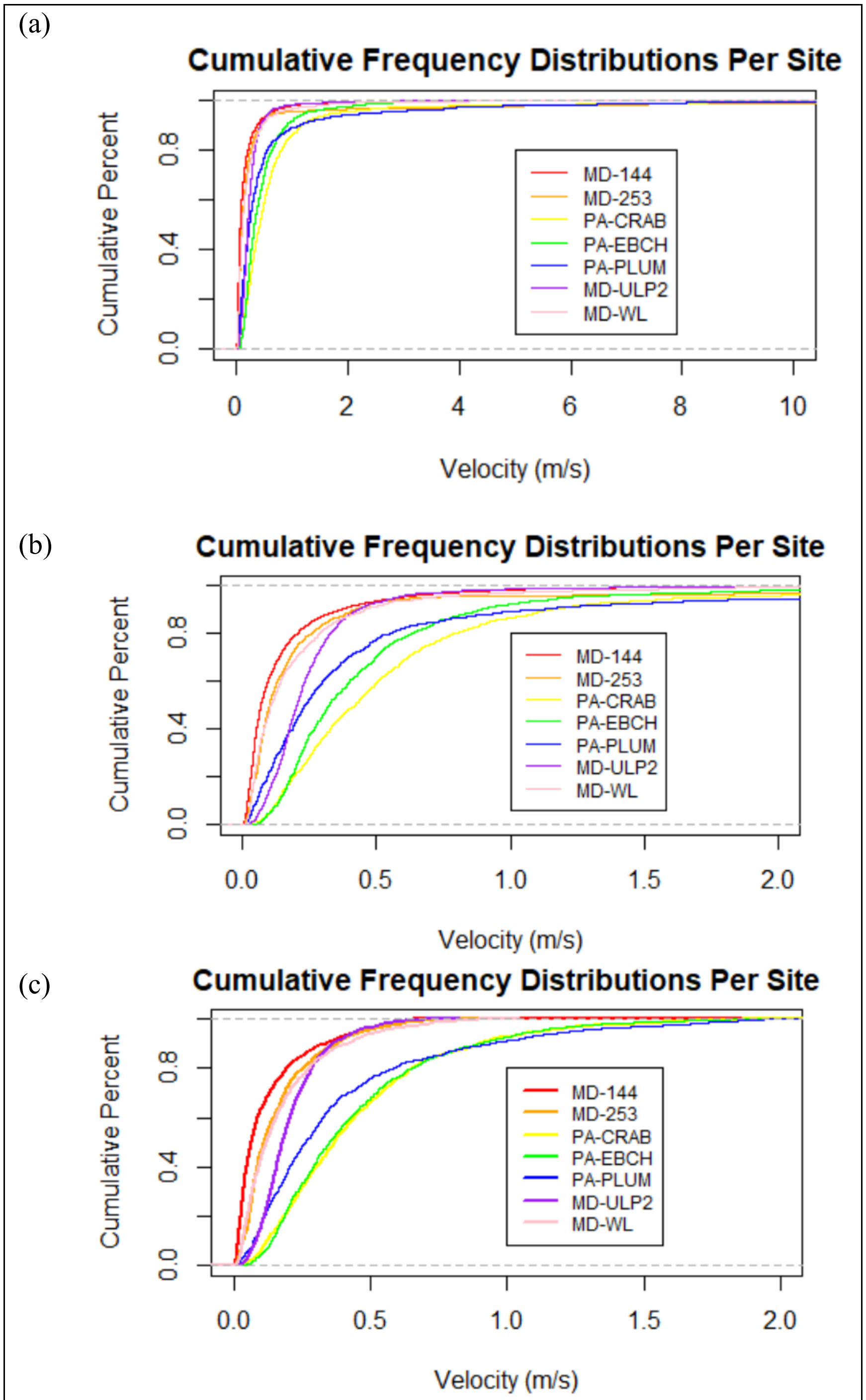


Figure 1.3.14. (a) Cumulative frequency distributions per site of cross section mean velocity. (b) CFD of cross-sectional mean velocity between 0-2 m/s. (c) Cross-sectional mean velocity CFD after discharge-drainage area correction and outlier removal.

1.3.4.g Flow velocity correction and outlier analysis

Differences between upstream and downstream drainage areas and measured discharge (Q_m) at the downstream end on the date of aerial surveys (ASD) were used to estimate upstream discharge (Q_e; assuming a log-linear relationship between discharge and drainage area; Table 1.3.5). Analysis revealed the ratio of upstream to downstream discharge ranged from near 1:1 (0.96 US:DS at 253) to 1:3 (0.36 US:DS at site 144). It was expected that sites with smaller US/DS fractions would be most altered by applying a discharge-drainage area correction rather than a single discharge when calculating cross-sectional mean velocity.

Table 1.3.5. Estimated discharge-drainage area relationships per site. Q_m is measured discharge, and Q_e is discharge estimated using a log-linear relationship between discharge and drainage area.

	DS end Drn. Area (m^2)	US end Drn. Area (m^2)	Aprx. cost dist. (m) US-DS	DS Q_m (m^3s^{-1}) ASD	US Q_e (m^3s^{-1}) ASD	US / DS Q
144	2.9	1.0	1905	0.0184	0.007	0.36
253	9.0	8.7	857	0.0569	0.055	0.96
CRAB	2.7	1.9	1091	0.0874	0.061	0.70
EBCH	15.0	13.8	1200	0.3433	0.317	0.92
PLUM	3.2	2.0	1356	0.0642	0.040	0.63
ULP2	26.5	18.5	2161	0.1782	0.124	0.70
WL	6.4	2.9	1253	0.0421	0.019	0.44

Velocities derived from drainage area-corrected discharge were reduced relative to uncorrected velocities (Fig. 1.3.15). As expected due to ratios between US and DS drainage area (Table 1.3.5), it was evident that site 144 was most altered by the correction, and especially for US cross-sections (Fig. 1.3.15, where site 144 is red and square points represent US cross-sections). However, it was evident from Figure 1.3.15 that there were still many outliers in cross-sectional mean velocity after discharge-drainage area correction.

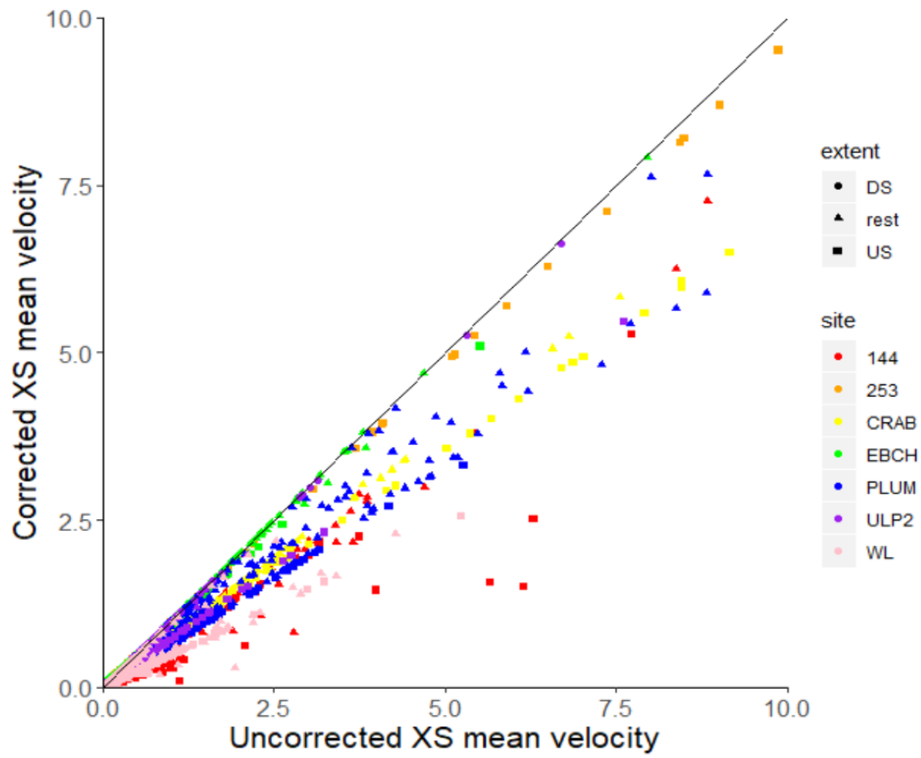


Figure 1.3.15. Comparison of cross-sectional mean velocities before and after correction relative to drainage area. Point color represented the 7 study sites, and point shape identified the extent of the cross-section represented, and the black line represents a 1:1 relationship.

There was visual evidence in mapped values that extreme outliers in velocity were related to a combination of low depth values and low percentage of depth pixels contributing to the cross-sectional mean depth (Fig. 1.3.16). Therefore, after drainage area correction, velocity outliers were examined per site relative to both the proportion of pixels within a cross section with depth values that contributed to the cross sectional mean depth, and the cross-sectional mean depth value itself.

After drainage area correction, outliers represented between 2.7%—7.4% of cross-sectional mean velocities per site. The site with the greatest proportion of outliers was PLUM. Figure 1.3.16 displays pixelated depths mapped on top of cross-sectional mean flow velocities that were below (white) or above (black) the cutoff for extreme outliers (~1.4 m/s) at PLUM. It was evident that black cross sections were associated with very low depth values and/or very few depth pixels within the cross section (low proportion of depth pixels contributing to the cross-sectional mean depth calculation; Fig. 1.3.16). Purple cross sections in Figure 1.3.16 show where there were no pixels of depth within the cross section, so neither mean depth nor mean velocity could be calculated for that cross section. All depth values within the depicted black (outlier velocity) cross sections were below 0.1m (light blue), and many were below 0.01m (orange), which inevitably yielded high velocity values when width and discharge were held relatively constant within the depicted area (Fig. 1.3.16).

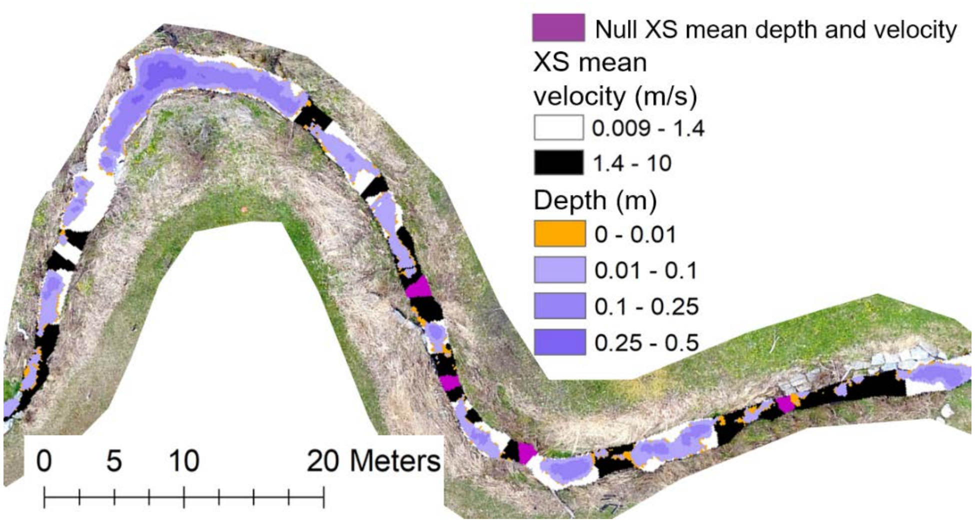


Figure 1.3.16. Mapped cross-sectional mean velocity underlying pixelated depth estimates within the restored extent of PLUM to depict where outliers occur. Purple cross-sections show where there were no pixels of depth, and therefore neither cross-sectional mean depth nor velocity estimated. White areas depict realistic (non-outlier) cross-sectional mean velocity estimates, while black areas depict extreme outliers in cross-sectional mean velocity for this site.

In plotting cross-sectional mean depth against cross-sectional mean velocity across sites, it was evident that very low depth values were what drove very high velocity values (Fig. 1.3.17). The inset on Figure 1.3.17 displays the relationship between cross-sectional mean velocity and mean depth values below 0.1m. The approximate outlier threshold across sites, which ranged between approximately 1.2 and 1.4m, is portrayed by the black dashed vertical line (Fig. 1.3.17). It was evident from the inset that very low values of cross-sectional mean depth were what drove the majority of outliers in cross-sectional mean velocity, particularly mean depth values between 0-0.02m (Fig. 1.3.15 inset). Very low proportions of depth pixels also yielded high cross-sectional velocities, but the pattern was not as clear as with depth, as there were high proportions of depth pixels that contributing to a cross sectional mean depth that still yielded outlier velocity values. Therefore, the low depth values are more likely what was the major contributor to outliers in velocity.

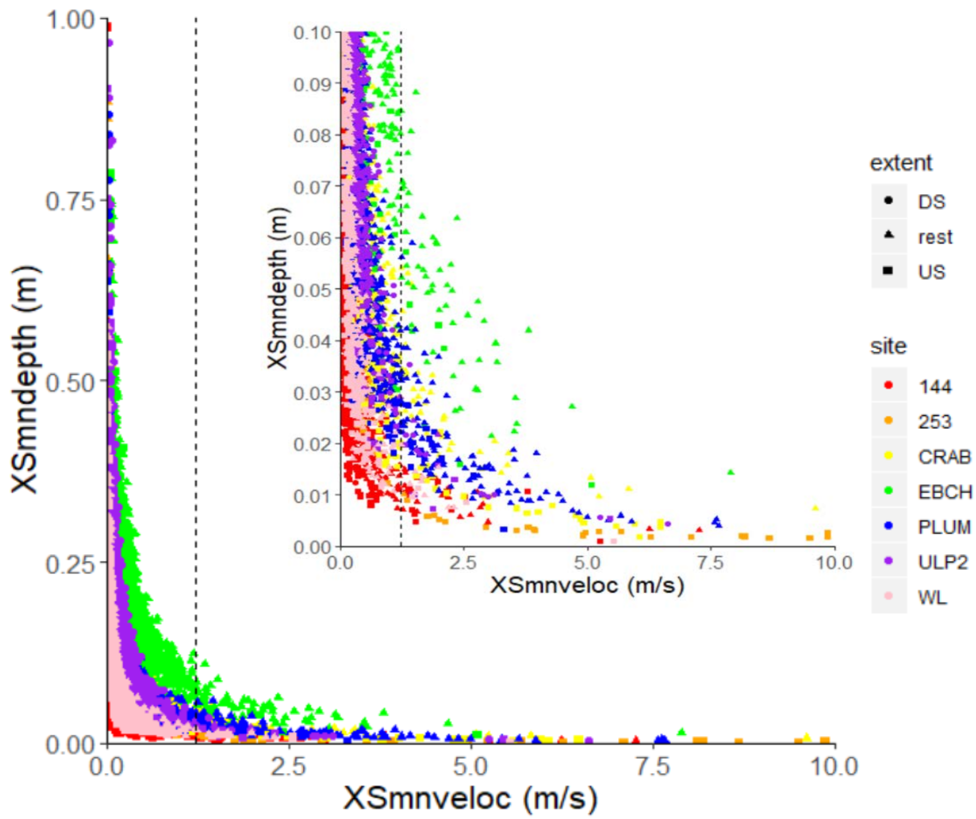


Figure 1.3.17. Comparison of cross sectional mean depth (m) to drainage area-corrected cross-sectional mean velocities per site. The inset zooms to just 0-0.1m depth. Both plots display the approximate outlier threshold with the black dashed line. (No points in the main graph were obscured by the inset.)

After assessment and removal of outliers separately from each site, the cross-sectional mean velocities were much more realistic in magnitude (Fig. 1.3.14c; Fig. 1.3.18). CRAB and PLUM (green and blue in Fig. 1.3.18, respectively) had the highest basin slopes of the study sites (6.5 and 4.3 degrees, respectively; Table 1.2.1). Similarly, EBCH (blue-green in Fig. 1.3.18) had the highest discharge on the day of the aerial survey of all study sites (0.34 m^3/s, whereas the rest of sites ranged between 0.02 – 0.18 m^3/s). Therefore, it was unsurprising CRAB, EBCH and PLUM had higher velocity values than the rest of the study sites (green, blue-green and blue, respectively, in Fig. 1.3.18).

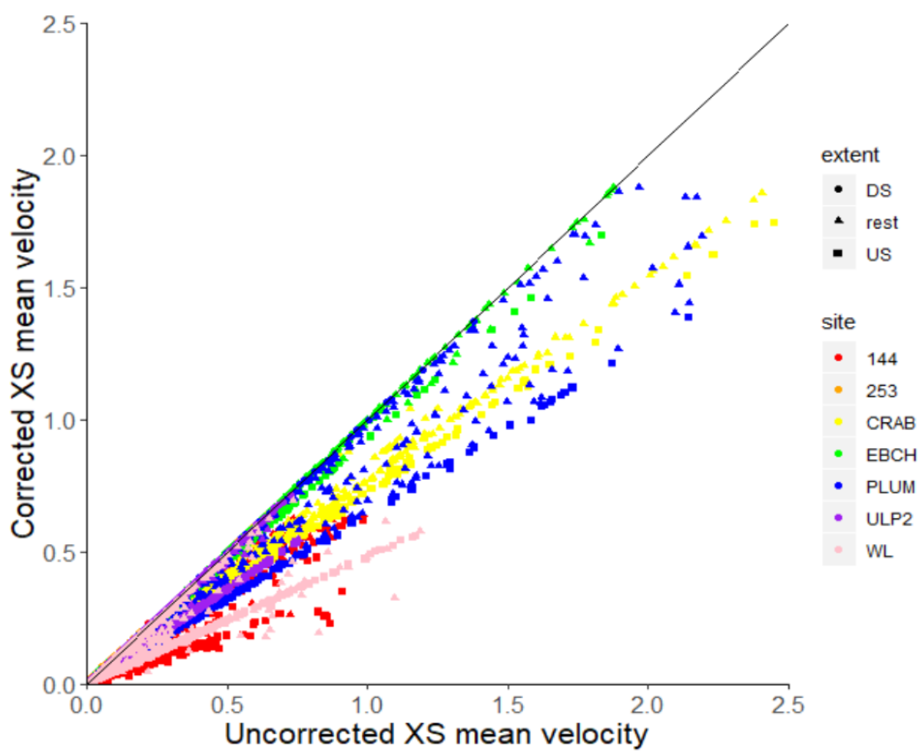


Figure 1.3.18. Comparison of cross-sectional mean velocities before and after correction relative to drainage area (extreme outliers removed from both datasets).

1.4 Discussion

Aerial SfM imagery and surveyed ground control points successfully produced DEMs and orthomosaic images for the study sites. These products were used in a workflow created to extract measures of physical channel habitat within the GIS environment using repeatable methods. Measures of channel dimension (depth and width), flow and sediment were realistic in their pattern, extent and magnitude. Interpolating, smoothing, binning and correcting results was often successful in addressing anomalous estimates. Ultimately, the workflow generated reasonable results that can be directly spatially compared to field data within delineated reaches and used more broadly to assess the effects of in-stream restoration on physical channel habitat.

Patterns of habitat variables generally matched expectations based on interpretations of orthoimagery and theoretical relationships. For instance, sinuous riffle-pool sequences with visible rocks and breakwaters aligned well with measures of shallower, faster water. Water color was darker in pools, and associated with deeper, slower measurements. Measured wetted widths appeared narrower in riffles and wider in pools. Finally, sediment appeared finer within pools and coarser within riffles. These phenomena were reflected in the results, with a pattern of coarse sediment, narrow widths, and shallow, fast water in what was interpreted as riffles, and finer sediment, broader widths, and deeper, slower water in what was interpreted as pools. Such patterns were also generally consistent across units at each site. For areas of channel with non-parallel banks, cross-sectional approximations and width measurements were less uniform laterally than might be expected from field surveys. However, end points of cross-sections placed in the field could easily be subjective in such circumstances. Typical cross sections are meant to be placed perpendicular to flow, but that can be difficult when flow is directionally variable, as around bends. Representing a cross section as more of a triangle in areas of divergent banks may better represent the range of choices available in the field. This is particularly true in terms of cross-sectional velocity, where orthogonal lines to flow are difficult to determine. Cross-sectional mean velocity also generally corresponded with field values, and estimates were improved by drainage area-discharge correction and outlier removal. CFDs per site displayed central tendencies and rank orders among sites that were corroborated by field surveys.

Alignment within Agisoft and validation of the resulting DEMs was successful across most of the study areas within sites. However, as mentioned, habitat measurements were not delineated for the reference (unrestored) site due to issues with alignment of the point cloud within Agisoft. Similarly, one of the sites' DoDs indicated distortion. Systematic error via patterns of difference between the SfM-generated DEMs and LiDAR DEMs was not a major problem at most sites. At the sites where this error was present, or even dominant, habitat measures were nonetheless derived, and exhibited similar correspondence with expected magnitudes and patterns. This is not surprising for wetted width, the delineation of which was not dependent on elevation values. For depth and velocity, this is likely due to the local and relative nature of the approximations.

Figure 1.4.1 displays the habitat metrics dependent on elevation values (e.g., depth and velocity) for the upstream extent of PLUM, the site with tilted systematic error shown in Figure 1.3.3. If the tilt were severe enough at this site, minimization of edge elevations would lead to large extents with no depth values, as the lower edge elevation would be below the elevation of much of the channel. However, this phenomenon was not present in the results, nor was there a pattern of depth or velocity reflecting the pattern of systematic error (Fig. 1.4.1). Instead, areas of null depth were associated with road crossings as expected, channel islands and vegetative cover. The distribution of null depths did lead to some cross-sections with null means and, accordingly, no velocity estimate. Alternatively, where just a few pixels with low depth remained within zones dominated by null depths, cross-sections had very small mean depths, and unusually high mean velocities. Correcting velocities using a discharge-drainage area relationship lowered some of the high velocities at PLUM. High values remaining after discharge corrections were removed as outliers, which led to an improved range of cross-sectional mean velocity.

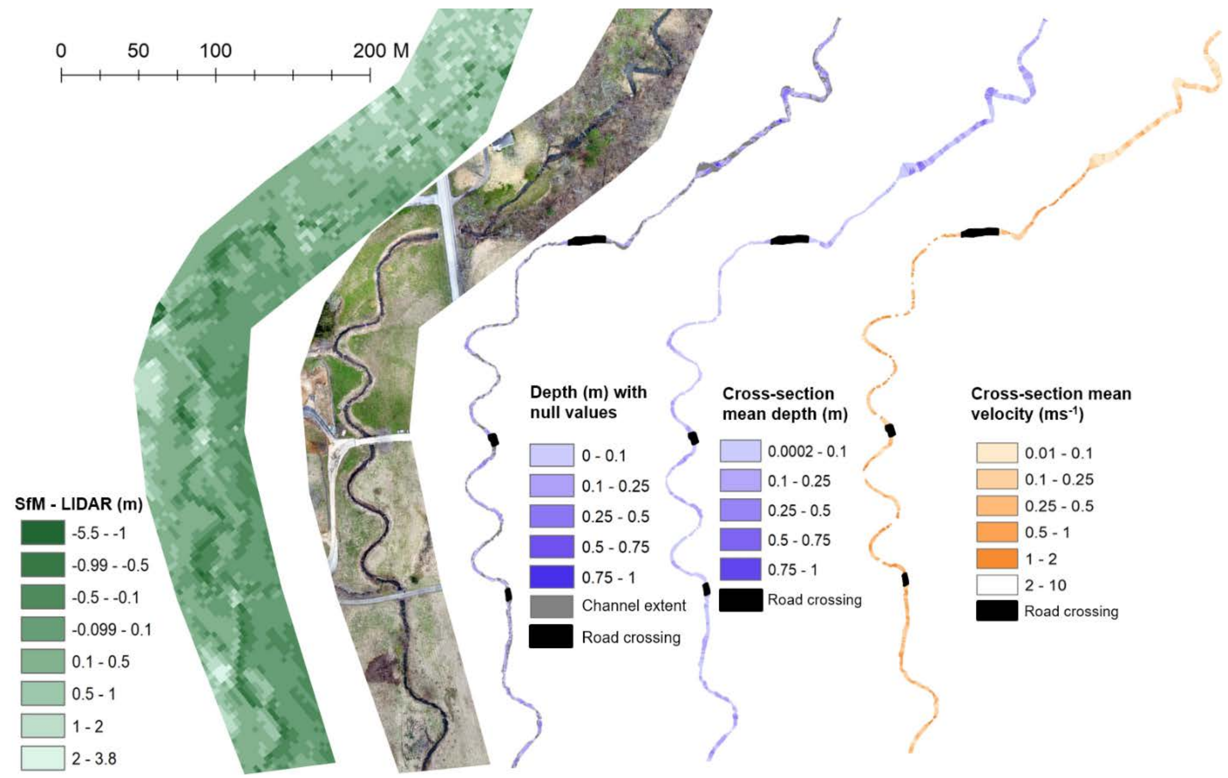
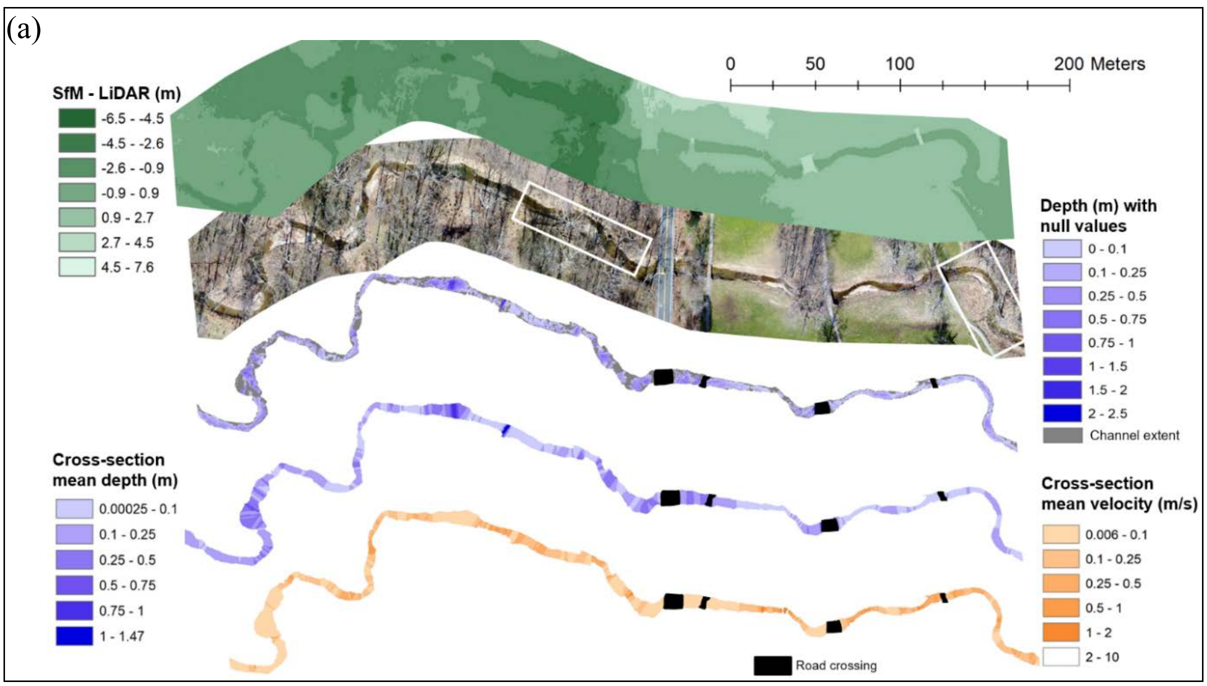


Figure 1.4.1. Upstream extent of PLUM, showing that the habitat measurements dependent on elevation do not seem affected by the systematic error in the DEM.

For the highly distorted DEM, it was possible that local portions of the channel were either similarly offset in elevation or equivalently tilted, allowing for success in the local and relative nature of the depth delineation, where absolute measures over broader extents may have failed. Figure 1.4.2 displays the habitat metrics dependent on elevation values (e.g., depth and velocity) for this site. As with PLUM, null depth values were associated with road crossings (e.g., right side of Fig. 1.4.2a, top of 1.4.2c), vegetation (e.g., left side Fig. 1.4.2a, throughout

1.4.2b), and an island (e.g., bottom of orthoimage in Fig. 1.4.2c). Similarly, patterns of depth and velocity were not associated with the pattern of systematic error (Fig. 1.4.2a). As evident in Figure 1.4.2b-c, habitat measures adequately captured the patterns expected within the orthoimages: higher depths with lower velocities in the areas that appear to be pools (e.g., left and middle of Fig. 1.4.2b; middle of Fig. 1.4.2c), and lower depths with higher velocities in areas appearing as riffles (e.g., right, middle and far left of Fig. 1.4.2b; top, middle and bottom of Fig. 1.4.2c). The right side of the pixelated depth map in Fig. 1.4.2b provides further evidence that a combination of very low depth values and few depth pixels contributing to the cross-sectional mean depth yielded a high velocity value (1 - 1.5m/s; right side of velocity map in Fig. 1.4.2b).



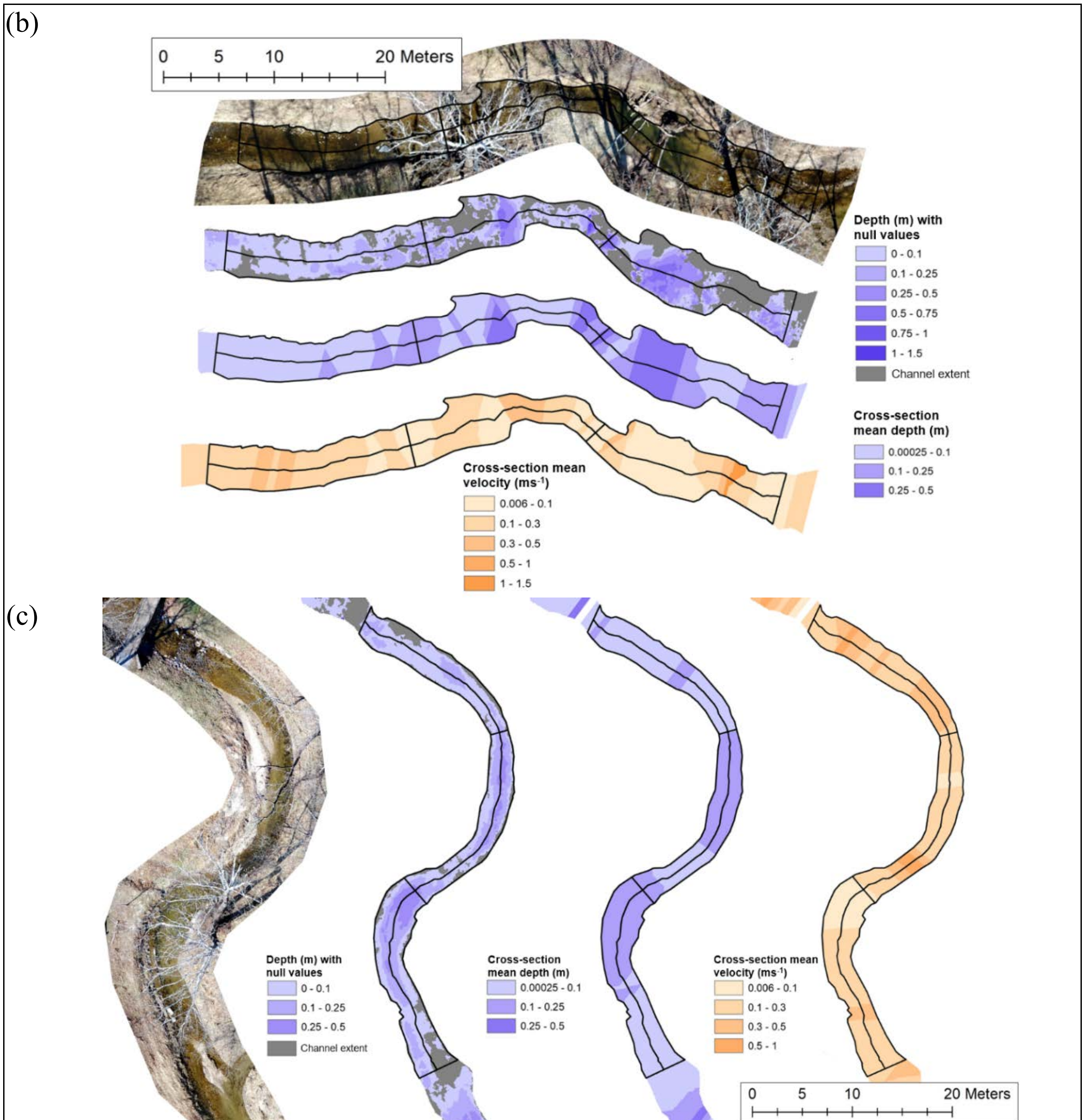


Figure 1.4.2. (a) Downstream extent of site 253, showing that the habitat measurements dependent on elevation do not seem to be severely affected by the systematic error in the DEM, with inset white boxes showing the extents of (b) and (c). Orthoimagery, depth and velocity values are displayed for the restored (b) and downstream (c) field reaches.

Despite the ability to capture habitat metrics, systematic error present in any form was obviously not ideal. The DEM displayed in Figures 1.3.2 and 1.4.2 might not be useful for other terrain applications, such as flow routing models. Further investigation of systematic errors might assess whether they could be fixed via detrending, whether greater attention to point cloud quality would improve results, or whether another aerial survey of these sites would be necessary. Future flights could involve additional GPS points and unique ground features that might aid in alignment and accuracy of the point cloud, as recommended by previous UAV-based SfM work (e.g., Fonstad et al. 2013; Woodget et al. 2017; Dietrich 2017; Carbonneau and Dietrich 2017).

Portions of the workflow accomplished manually were based on best judgement relative to study area conditions, and where possible, the workflow attempted to account for error associated with manual delineation. Manually delineated channel edges were subject to placement on the bank (higher than the actual water surface elevation) or within the wetted area (lower than the actual water surface elevation), which likely led to error in the estimations of water surface and depth. Figures 1.3.5 and 1.4.2 also display where interpolation was necessary for wetted channel edge delineation due to obstructions (e.g., vegetation) within the orthoimagery. Minimizing and smoothing the water surface attempted to address these errors. Zonally minimizing the channel edge elevation per cross-section area removed some noise in elevation likely caused by including non-wetted channel features (e.g., exposed rocks, vegetation). This meant areas mapped within the channel but actually representative of exposed vegetation (e.g., bright spot within the middle left of the DEM in Fig. 1.3.5b), were excluded from the depth map, and therefore were at least partially removed from analysis. The decision to use the median elevation value within a focal radius of ½ the median channel width per site to smooth the water surface could have resulted in over-smoothing fluctuations in the water surface, particularly in transition zones between pools and riffles, where water surface elevation is expected to change more rapidly. The use ½ the median channel width may also over smooth the water surface in areas where the channel is much narrower than the median width, or under smooth the water surface in areas where the channel is much wider. However, as seen in Figure 1.3.10b, despite areas of vegetation picked up in the water surface, fluctuations in channel elevation can be largely attributed to the channel form (e.g., pool-riffle sequence).

Potential error associated with manually delineating sediment categories was addressed by interpolating and binning categories. Delineated sediment represented binned categories to account for the possibility that, for example, areas that appear to be dominantly cobble were likely in fact a mix of cobbles and gravels. Results pointed to vegetation, image stitching errors and shadows within the orthoimagery as sources of increased subjectivity in the manual delineation of wetted channel edge and sediment categories. As described earlier, sediment categories mapped using visible edges of sediment patches. However, this required the assumption that sediment patches were relatively homogenous throughout (e.g., there was not a cobble patch in the middle of a silt patch). An automated process for extracting wetted areas and sediment from imagery was developed by Carbonneau et al. (2004). Their procedure was developed for gravel bed rivers, that were larger and more exposed, with less forested banks. The channel extents and sediment patterns were manually delineated here to account for areas where channels were obscured.

A comparison of manual to automatic delineation would be an interesting area of further study, particularly in the context of habitat measurement, as it could be argued that obscured areas not mapped by automated methods would be suitable to remove from analysis anyway, given the potential associated errors in sediment, depth and velocity delineation. However, a goal of the current analysis was to approximate habitat measures that could be compared with measures made in the field (chapter 2), so continuity of results, particularly within field reaches, was necessary.

Shadows and vegetation visible within the orthoimagery do appear to at least partially affect the channel bathymetry and resulting habitat measures. Patterns of depth were not clearly correlated with patterns of brightness throughout the orthoimagery. It is not surprising that SfM algorithms might interpret darker areas as lower elevations. Previous work has mapped channel depth using a spectral depth relationship, which correlated water depth with image brightness (e.g., Marcus et al. 2003; Legleiter and Fonstad 2012; Woodget 2015). An interesting area of further research might involve brightness correction (polarization) in the imagery and comparing SfM DEM to spectral image-based extraction.

Refraction correction is another alternative used for depth mapping, but not employed in this workflow. Refraction correction is based on the idea that actual elevation of submerged topography may be underestimated due to the refraction of light through water. Woodget and collaborators (2015) described a method of refraction correction that requires capturing the SfM photogrammetry at nadir, and using a simplified version of Snell's Law, $\left(\frac{\sin r}{\sin i} = \frac{h}{h_A} = \frac{n_1}{n_2}\right)$, to correct each point's apparent coordinates to true coordinates in the cloud. Others have successfully utilized multi-view off-nadir SfM photogrammetry to produce a point cloud with improved precision and accuracy in comparison to the single-view used by Woodget and collaborators (Figure 1.4.3). Such corrections were not performed in this workflow due to the predominance of shallow bathymetry at the sites, with maximum depths of approximately 1m (e.g., Tables 1.3.1 – 1.3.4), which was captured by the uncorrected depth estimates (Figure 1.3.12). Similarly, Dietrich (2017) mapped points along the water surface elevation in the field, which was not done here. Further study might involve mapping the water surface elevation for validation of the surface delineated from SfM (e.g., Figures 1.3.10-11), and comparing uncorrected to refraction corrected depths.

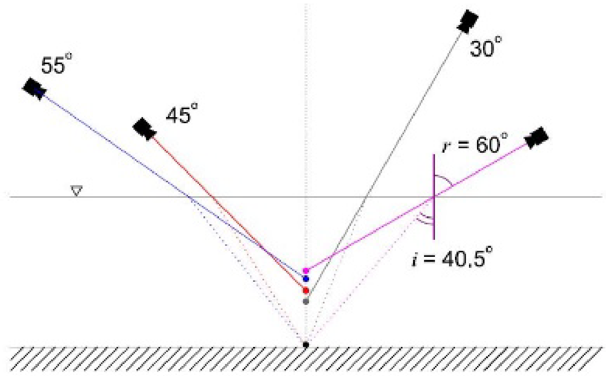


Figure 2. Refraction angles and apparent depth locations from four different camera angles/locations. The differences in the apparent depths illustrate one of the challenges of using multiple camera locations (as in Structure-from-Motion) to resolve depth.

Figure 1.4.3. Multi-view off-nadir camera angles used to capture depth (reprinted figure from Dietrich 2017).

The correction relative to drainage area did improve velocity values (e.g., Fig. 1.3.15), particularly for those sites with large changes in in drainage area study area. However, unrealistic values of velocity. Outlier analysis revealed that unrealistic values were closely associated with low depth values. Removing extreme outliers, which represented approximately 2-7% of velocity estimates per site, greatly improved the range of velocity values for all study sites. Given the close connection between velocity outliers and very small depth values, it is apparent that further work should investigate improving depth delineation. For instance, given the current method of water surface elevation, it would be interesting to see whether errors may be reduced by ensuring the wetted edge that was more visible in the imagery was accurately delineated, and then making sure the other edge was on the bank. Therefore, minimization of the edges would only consider the most accurately delineated edge elevation. Similarly, masking out areas of the channel covered with vegetation may improve the delineation by preventing elevation peaks associated with overhanging vegetation (e.g., Fig. 1.3.10). Given the broad extent of the data, such a mask should not prevent the methods from capturing the pattern of physical habitat throughout a given study area.

Although estimated cross-sectional flow velocities matched expectations regarding pattern and rank order magnitude per site, underestimation of depth due to image obstruction or improper edge delineation may have resulted in overestimations. Another potential source of error was the assumption of continuity throughout sites. Discharge correction relative to drainage area did improve velocity values, particularly for those sites with large changes in drainage area. remained and removing such values (2-7% of velocity estimates per site) greatly improved velocity ranges. Given the close connection between velocity outliers and very small depth values, it is apparent that further work should invest in improving depth delineation. For instance, given the current method of water surface elevation, it would be interesting to see whether errors may be reduced by ensuring that any visible wetted edge was accurately delineated, and ensuring obscured edges were higher on the bank. Minimization of the edges would therefore only consider the more accurately delineated elevation. Similarly, masking out areas of the channel covered with vegetation might improve delineations by eliminating elevation anomalies associated with overhanging vegetation. Given the broad extent of the data, such a mask would not prevent capture of physical habitat patterns throughout a given study area.

Even though manual delineation introduced some subjectivity into the procedure, the described workflow primarily consists of objective methods, which benefit from replicability. Future work with reliably mapped wetted channel extents could allow other researchers and monitoring agencies to gain continuous, objective measures of channel habitat from SfM imagery. However, the current methods still require training in SfM processing and an ability to scrutinize DEM distortion, and therefore might not be appropriate for broad use.

In addition to testing other methods of derivation, additional habitat measures commonly included in stream habitat assessments could also be derived from this data. Given the derivation of the water surface elevation profile across space, the water surface slope can be easily extracted at various spatial scales. Similarly, the bankfull height and angle to the water surface could be mapped for the channel extent using DEM analysis. Riparian vegetation could be classified and mapped from imagery and/or point clouds. Channel roughness could be approximated by DEM variation (e.g., pixelated slope calculation used by Marteau et al. 2017). These possibilities highlight one of the great advantages of SfM data: it can be revisited and utilized to support multiple objectives.

This chapter has shown the benefit of SfM data in its broad spatial extent and continuity of data. The workflow successfully used SfM data to extract habitat measures that were strong in spatial extent, pattern, and magnitude. This allows for an efficient means to analyze channel habitat precisely and explicitly over broader extents (e.g., closer to mesoscales, as suggested by Fausch et al 2002), than what was previously possible with conventional habitat assessment methods. The ability to gain spatially explicit, precise and extensive habitat data that can be directly compared to field measurements has been sparsely studied (e.g., Marcus et al. 2003; Woodget 2015; Woodget et al. 2016, 2017), particularly in the context of restored habitat (Carrivick and Smith 2019). The continuous habitat measures delineated in this chapter will allow for direct, spatially paired comparison with field values (chapter 2) and analysis of heterogeneity in restored physical channel habitat over variable spatial scales (chapter 3).

Chapter 2 Assessing equivalence between field and aerial representations of physical stream channel habitat

2.1 Introduction

Given the continuity, precision and extent of structure-from-motion (SfM) data, it could be especially useful in re-examining questions about aquatic habitat left vague by conventional field data, and in asking questions on scales beyond what is possible with field data. However, before SfM data can be reliably used for these kinds of questions, the representation of field conditions by SfM should be examined. Chapter 1 explored methods to extract physical habitat data from SfM imagery, and how realistic results were in pattern, extent, and magnitude. Chapter 2 will explore whether SfM data and field data equivalently represent components of aquatic habitat. To do so, habitat data must be extracted from SfM and the field at equivalent, spatially paired intervals to ensure comparable sampling effort.

As articulated in Chapter 1, this study employed portions of two sampling regimes used to monitor and assess streams in the Mid-Atlantic: one was a combination of modified versions of EPA's Environmental Monitoring and Assessment Program (EMAP; Kauffman et al. 1999; US EPA 2004) and of California's Surface Water Ambient Monitoring Program (SWAMP; Fetscher et al. 2009) protocols; and the other was the Maryland Biotic Stream Survey (MBSS) protocol (Stranko et al. 2015). These protocols were used for compatibility with previous studies at the sites (e.g., Kroll *unpublished data* used modified EMAP / SWAMP protocols, and Hilderbrand *unpublished data* used MBSS protocols). Each protocol represents a different approach to assessing physical habitat conditions, namely through their different scales of summary. For the portion of the modified EMAP / SWAMP protocols employed in this study, many points throughout the reach are used to measure (depth and width) and categorize (flow and substrate) features (Fig. 2.2.1). Alternatively, for the portion of the MBSS protocols employed in this study, each reach was divided into six segments (called a “facies” map in the MBSS protocols), which are used to summarize dominant habitat categories (depth, flow and substrate; Fig. 2.2.2).

Each protocol requires different levels of interpolation to summarize habitat measures at the reach scale. The portion of the modified EMAP / SWAMP protocol used in this study provides measurements and estimates at many points placed systematically throughout the field reach (Kauffman et al. 1999; US EPA 2004), while the portion of the MBSS protocol used in this study estimates conditions across broader extents (segments) to summarize the entire area of the field reach (Stranko et al. 2015). Each method was thus susceptible to subjectivity and imprecision in different ways. Although the EMAP protocol was perhaps more quantitative, discrete transects are vulnerable to missing features between successive cross sections that will not be included in interpolated reach-wide summary (e.g., average depth). The MBSS protocol assesses the entirety of the reach area, however in a much more qualitative manner, and may be susceptible to subjective interpretation of variability within each segment. In discussing potential biases in data collection, Barbour et al. (1999) noted that increased observer training (Hannaford et al. 1997) and moving from qualitative to semi-quantitative or quantitative data collection (Kauffman et al. 1999) can increase the precision of resulting estimates.

Researchers have directly tested the success of different habitat assessment techniques. Simonson (1993) compared the precision and accuracy of estimating channel habitats, substrate composition and dominance, bank erosion, and bank land use at two spatial scales: transect (stream cross section) and segment (tens to hundreds of meters longitudinally). Estimates were overall strongly correlated, with low mean differences, and ultimately Simonson (1993) suggested utilizing a combination between the two scales to gain the precision of the transect scale for most measurements, but the benefit of the compositional view of substrate, as well as the acknowledgement of the heterogeneous nature of habitat at the segment scale. This study represented a direct test of the ability of different habitat assessment techniques to adequately represent channel habitat. Similarly, Wang et al. (1996) explored how visual estimates of channel habitat can introduce bias and other error. While researchers didn't find subjectivity to be a major source of overall error in their study, they outline how subjectivity has increasing influence on results with increasing heterogeneity of the channel (Wang et al. 1996). They similarly advocate that clear habitat definition reduces ambiguity and therefore leads to lower observer error (Wang et al. 1996). These studies point to the need to scrutinize the scale, resolution, variation and bias in data to adequately understand how it represents local habitat conditions.

Each of the sampling protocols employed in this study were also developed to summarize physical habitat at representative scales. For instance, sampling reach lengths are typically delineated proportionally to channel width, and to capture habitat units systematically (e.g., Karr et al. 1986; Kauffman et al. 1999). This sampling design attempts to represent habitat features at scales and resolutions proportional to what shapes the habitat, such as hydrologic and geomorphic factors (e.g., Leopold and Wolman 1957; Karr et al. 1986; Kauffman et al. 1999; Fetscher et al. 2009). As introduced in Chapter 1, habitat classification schemes also aid in capturing representative habitat, defining various spatial scales on which habitat is formed and affected by environmental factors, and from which it may be important to get representative measures (e.g., Fig. 1.1.2; Frissell et al. 1986; Hawkins et al. 1993). Further, Fausch et al. (2002)

argued for greater continuity over spatial scales that capture larger extents (e.g., to the segment scale: $10^3 – 10^5$m), to better represent understand the processes affecting aquatic biota. They also propose that managers work at these larger scales (Fausch et al. 2002), a view furthered by skepticism with the scales, and therefore processes, missed by stream restoration projects (e.g., Hilderbrand et al. 2005; Simon et al. 2007; Bernhardt and Palmer 2011). These criticisms have led researchers to assess stream degradation and its management on broader (catchment) scales (Walsh et al. 2005; Swan and Brown 2017; Kroll et al. 2019). The use of remote sensing data to capture broader extents has supported the effort to capture physical characteristics of streams on larger scales (Marcus and Fonstad 2008; Bergeron and Carbonneau 2012; Carbonneau et al. 2012; Marcus et al. 2012; Dietrich 2016; Woodget et al. 2017).

Unlike field data, there is not a tradeoff between precision, continuity and extent of what is represented in SfM data. SfM data can also be summarized at many different scales given its extent and continuity, while retaining its precision. As with field assessments, it is necessary to address measurement goals when planning the extents on which SfM data is collected. Given the extent of the datasets resulting from SfM, it can be a challenge to find the right extent over which to summarize SfM data. Researchers have used a variety of remote sensing platforms to assess their utility at varying scales. Woodget et al. (Woodget 2015; Woodget et al. 2016) used SfM data to map flow types in a 50m stretch of river for comparison to conventional bankside mapping, and to more precise field validation. Marcus et al. (2003) similarly mapped flow units for comparison to field efforts, but used hyperspectral imagery over larger (3-5km) reaches. Light Detection and Ranging (LiDAR) data has been used to perform an EMAP-like habitat assessment over 500m (Hall et al. 2009), and SfM data has been used to map hydraulic fish habitat (Tamminga et al. 2015), though neither habitat mapping effort was compared to field-measured habitat. Researchers have also used SfM data over broader extents (10^4 m) to understand larger scale processes with more continuity than are typically possible with field surveys (e.g., Marcus and Fonstad 2008; Carbonneau et al. 2012; Dietrich 2016). The extent, continuity and precision of SfM allow for flexibility in the scale of assessment, which may help to assess representativeness. As discussed, conventional field measurements and estimations are subject to error, so may not be the best assessment of the accuracy of remote sensing data (Marcus et al. 2003, 2012). However, it is useful to compare SfM data to field efforts for some validation that the two data sources can represent the system in a similar manner. Therefore, the objective of this chapter is to assess whether field and SfM data can equivalently represent spatially paired reaches when data is extracted at equivalent intervals.

2.2 Methods

To compare field and SfM -measured habitat data, equivalent points within the data were assessed. For direct spatial comparison between field and SfM measures of habitat, modified EMAP / SWAMP protocols were completed at Pennsylvania study sites, and modified MBSS protocols were completed at Maryland study sites (Chapter 1). To execute these protocols within the geographic information system (GIS) environment on the SfM data, the field reaches were delineated in approximately the same locations, as described in Chapter 1. The samples within each reach were then aggregated to compare with aggregate field sample values.

2.2.1 Paired sampling unit metric comparisons

2.2.1.a EMAP / SWAMP transects at Pennsylvania sites

At PA sites, field reaches upstream of and within the restoration projects were delineated to match the modified EMAP/ SWAMP protocol used in the field (Kroll, *unpublished data*; Kauffman et al. 1999; US EPA 2004; Kroll et al. 2019). For the paired SfM reach, a cross-section transect polyline feature was placed perpendicular to the channel boundaries at the approximate coordinates of the downstream end of the field reach in ArcGIS (Chapter 1; Fig. 2.2.1). Successive transect lines were then placed 10m apart in the upstream direction, as measured along the channel centerline (11 total transects, 100m total reach length; Chapter 1). A label field was added to the attribute table for transect line features, with each transect labeled with the spatially paired field reach name (e.g., transect A-K at 0-100m from the downstream end of the transect). Along each transect, 5-7 points were placed to match the field sampling points. A label field was added within the point attribute table to label each point with its position, which was spatially paired to field values. (The label for left or right bank was "LB" or "RB", for 1m from left or right bank was "1L" or "1R", for the center point between the left or right bank and the center was "LC" or "RC", and for center was "C;" Fig. 2.2.1). Hereafter these are referred to as EMAP / SWAMP transects.

To assess potential sampling error, different methods of transect placement and measurement, and different ways to replace null depth values, were assessed. The transect placement described above was considered "automated" placement. Alternative "manual" placement of transects were not exactly 10m apart, but were instead located and oriented closer to where and how they were placed in the field (e.g., Fig. 2.2.1). This was performed to assess the degree that differences in transect placement during field surveys and during SfM data extraction may have affected comparisons. Automatic and manual transect placement is from here forth abbreviated as AP and MP, respectively, for all EMAP / SWAMP transect metrics. MP transects were compared to AP transects to assess the potential sampling error associated with transect placement differences between the field and SfM data. Associated sampling points were placed along MP transects as described above for the AP transects. Transects were "manually" measured using the Calculate Geometry tool within the attribute table of the transect line feature (ESRI 2019), which included both AP and MP transects. Widths were manually measured in this way to assess the accuracy of the more "automated," Euclidean distance-based measurement method. Euclidean distance and manual width measurements are from here forth abbreviated as EM and MM, respectively. Null depth values were replaced with the underlying cross-section mean depth, which was calculated using the cross-section zone approximations (see Chapter 1), for comparison to no null value replacement. Comparisons with no depth value replacement are abbreviated as NR and those with replacement are abbreviated as WR from here forth.

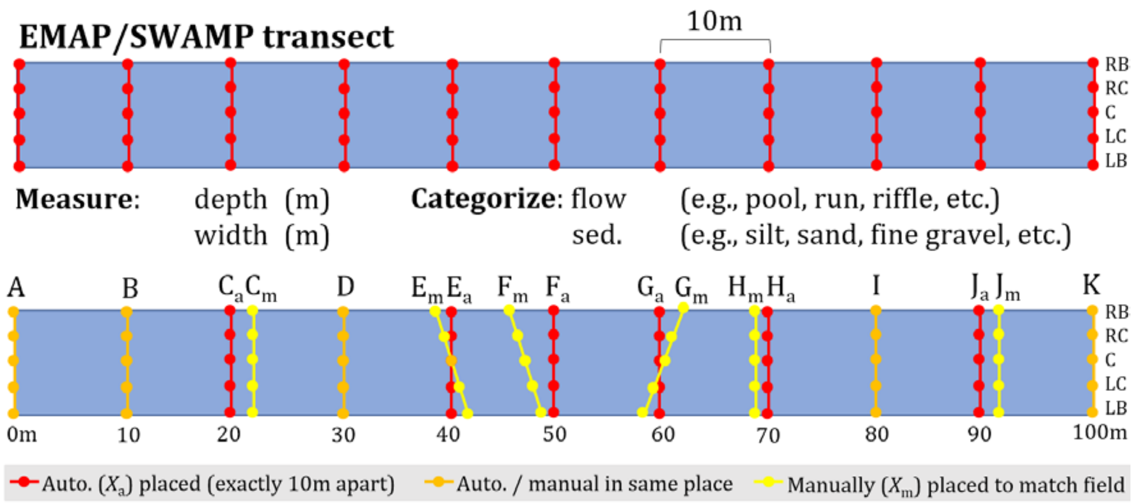


Figure 2.2.1. EMAP / SWAMP transect and sampling point placement within PA reaches, with labels and what is extracted per point. Manual vs. automatic placement of transects is depicted, with red indicating automatic placement (exactly 10m apart), yellow indicating manual placement to match field transect placement, and orange indicating the automatic and manual placement was in the same place.

The value of the depth, width, velocity and sediment underlying all sampling points were extracted per point, and the resulting tables were exported from the GIS. Field transect widths were compared via correlation to paired SfM transects for each method of extraction (AP-EM, AP-MM, MP-EM and MP-MM). Similarly, field depths were compared via correlation to SfM depths at paired transect points for each method of SfM extraction (AP-NR, AP-WR, MP-NR and MP-WR). The distribution of differences between paired sampling units for each of the methods were displayed via boxplot. The root mean square error (RMSE) was calculated for each comparison and displayed below the boxplot. Correlation strength and significance was tested with a two-sided Pearson's product moment correlation coefficient (R Core Team 2018).

SfM-derived velocity values were binned into equivalent categories as the field data to enable sampling unit comparison. To establish these categories, a combination of flow definitions was used. Fetscher et al. (2009) define flow classes relative to velocity and depths, respectively, for riffles as >0.3ms^{-1} and <0.5m, glides as <0.3ms^{-1} and <0.5m, runs as >0.3ms^{-1} and >0.5m, and pools as <0.3ms^{-1} and >0.5m. Barbour et al. (1999), also separate “fast” from “slow” water at 0.3 ms^{-1} (pg. 5-15), but determined riffles and runs to be between 0.1 – 0.5 ms^{-1} (pg. 6-5). The latter definition better captures shallow pools, but the former better captures the general longitudinal relationship between depth and velocity, so a combination of the two definitions were used here. Glides and runs were combined as one category to match field methodology (Kroll, *unpublished data*; Kroll et al. 2019). A null velocity value was considered exposed (“EXP”); a value between 0 – 0.1 ms^{-1} regardless of depth, or <0.3 ms^{-1} with depth >0.5m, was considered a pool (“P”); a value between 0.1 and 0.3 ms^{-1} with depth <0.5m, or >0.3 ms^{-1} with depth >0.5m, was considered a glide/run (“GL”); and a value greater than 0.3 ms^{-1} where depth was <0.5m was considered riffle (“RI”). (Rapids were not present at these sites). The categorized SfM velocity was then compared to the field category at each spatially paired point using a linear-by-linear association test (R Core Team 2018; Hothorn et al. 2019), and displayed in a contingency table. The linear-by-linear test assessed the null hypothesis that the two categorical datasets have no association, which can be rejected given a significant p-value (α = 0.05). The comparison between field and SfM categorical flow data was displayed and tested separately for AP and MP transects.

Field sediment categories were binned into six, ordered categories:

Table 2.2.1. EMAP/SWAMP sediment categories

FN (1)	SA (2)	GR (3)	CB (4)	B (5)	R (6)
Fine or hardpan	Sand	Fine or coarse gravel	Cobble	Small or large boulders	Bedrock

Other field categories, such as wood and "other" (e.g., cement) were removed from the analyses. For the binned sediment categories (e.g., sand-gravel, gravel-cobble, etc.) mapped within SfM reaches, only the finer of the two in the bin were kept in the analysis (e.g., sand-gravel became just sand) for comparison to field categories. Field and SfM categories were displayed in contingency tables and compared via a linear-by-linear association test (R Core Team 2018; Hothorn et al. 2019), which was completed for AP and MP transects separately.

2.2.1.b MBSS segments at Maryland sites

For MD sites, field reaches within, and upstream and downstream of the restoration projects were delineated within the SfM data to match the modified MBSS protocols used in the field (Hilderbrand, *unpublished data*; Stranko et al. 2015). Coordinates representing the downstream end of each reach were used to measure 75m upstream along the centerline. The resulting reach polygon was split perpendicular to channel edges at 25-meter increments from the downstream end along the centerline (Chapter 1; Fig. 2.2.2). The reach polygons were then split longitudinally using the centerline, which left 6 segments that approximated the 25m long x ½ channel wide "facies" (Stranko et al. 2015). Hereafter these are referred to as MBSS segments.

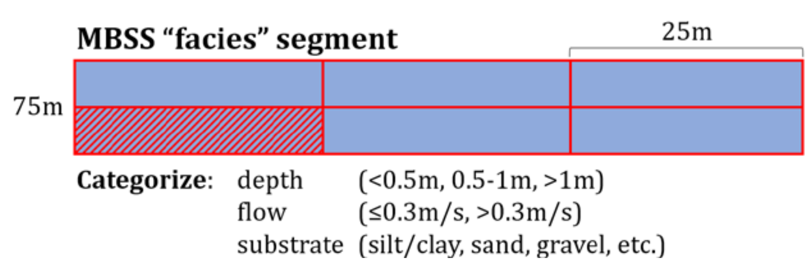


Figure 2.2.2. MBSS segment sampling regime.

MBSS segments were used as zonal areas with which to summarize mean and median depth and velocity, as well as the dominant sediment type. Depth data were binned to correspond with MBSS depth categories 1-3 (Table 2.2.2; Stranko et al. 2015). Velocity data were binned for comparison to the field flow velocity categories 1-2 (Table 2.2.2; Stranko et al. 2015). Only the comparison between mean and median velocity categories were described here, as the median depth per segment did not change any segment depth categories. Dominant sediment per segment was calculated by finding the delineated sediment category with the highest area per segment (Table 2.2.2; Stranko et al. 2015). Resulting categories of depth, flow, and sediment were compared to respective categories identified in the field. Segment-to-segment paired categorical comparisons were displayed in separate contingency tables and analyzed using a linear-by-linear association test (R Core Team 2018; Hothorn et al. 2019).

Table 2.2.2. MBSS "facies" map categories.

<table>
<tr><td rowspan="2">Average depth</td><td>1</td><td colspan="2">2</td><td>3</td></tr>
<tr><td><0.5m</td><td colspan="2">0.5 – 1.0m</td><td>> 1.0m</td></tr>
<tr><td rowspan="2">Average flow velocity</td><td colspan="2">1</td><td colspan="2">2</td></tr>
<tr><td colspan="2">≤0.3m/s</td><td colspan="2">>0.3m/s</td></tr>
<tr><td rowspan="2">Dominant sediment</td><td>Y (1)</td><td>S (2)</td><td>G (3)</td><td>C (4)</td></tr>
<tr><td>Silt / clay</td><td>Sand</td><td>Gravel</td><td>Cobble</td></tr>
</table>

2.2.2 Paired reach-aggregated metric comparisons

2.2.2.a EMAP / SWAMP transects, PA sites

All sampling units were averaged across each reach for field and SfM derived estimates of depth and width. Transect placement, measurement and null replacement methods were again assessed separately for comparison, and the strength and significance of each correlation was tested with a two-sided Pearson's product moment correlation coefficient (R Core Team 2018). Slopes and intercepts of linear regressions were computed for each comparison only to assess how closely relationships matched a 1:1 line.

Reach-level velocity and sediment category proportions were assessed as the ratio of sampling unit instances relative to the total number of sampling units per reach. Differences between field and SfM proportions of each velocity and sediment class were displayed as a histogram and cumulative frequency distribution (CFD). Transect placement methods were again assessed separately. Equivalence between SfM and field data was assessed using the null hypothesis that the mean difference in proportions was different from zero (Robinson 2016). A standard t-test was inappropriate in this case, because failing to reject the null would not imply equivalence, just the inability to reject, which could occur as a result of low power or confounding variables (Robinson and Froese 2004). Distributions of differences in proportion per category per reach were tested for normality using the Shapiro-Wilk normality test (R Core Team 2018), then a paired t-test of equivalence (PTTE; Wellek 2003, 2010; Robinson and Froese 2004; Robinson 2016) was used when the distribution was normal. The alpha level was set to 0.05 to obtain a 95% confidence interval. The equivalence value is set subjectively but based on recommendations by Wellek (2003, 2010) for a strict test, using ±25% of each distribution's standard deviation as the bounds on either side of 0 (Wellek 2003,2010; Robinson 2016). Significant results of the PTTE occurs when the calculated t-statistic is less than the "critical value." This "critical value" is a function of the F-distribution quantile for the square of the equivalence value (Wellek 2003,2010; Robinson and Froese 2004).

When distributions did not meet the assumption of normality, a robust two one-sided test (RTOST; Robinson 2016) was used instead of the PTTE. The RTOST is a combination of a two one-sided test of equivalence (TOST; Robinson and Froese 2004; Robinson 2016) and Yuen's trimmed means t-test, and is robust to nonnormality, outliers and heteroscedasticity by using a trimmed mean and winsorized variances (Yuen and Dixon 1973; Yuen 1974). A TOST is a more basic version of the PTTE, with two one-sided confidence intervals calculated relative to the α-level, and significance assessed through a standard student's t-test (and significance in test results occurs when these confidence intervals lie completely within the equivalence interval, again set as a function of the distribution's standard deviation; Robinson and Froese 2004). The RTOST uses Yuen's trimmed t-test as an alternative, more robust approach to the Student's t-test. If both one-sided tests are rejected using Yuen's t-test, the RTOST is significant, e.g., the null of dissimilarity can be rejected. The α was set to 0.05 to gain a 95% confidence interval for each test determining the interval. The trimming proportion was set to 0.1 after some experimentation of how trimming affected results (authors recommend less than 0.25; Robinson 2016).

2.2.2.b MBSS segments, MD sites

Proportions per reach of each unique category was assessed as the ratio of the number of segments to six (total segments per MBSS reach; e.g., proportions per reach ranged between 0/6 and 6/6). Differences between field and SfM proportions per reach for depth, flow and sediment categories were calculated. The distribution of differences between field and SfM proportions per category per reach were displayed in histograms and CFDs. Distributions of mean differences in proportion were tested for equivalence using a PTTE (if normally distributed) or a RTOST (if not normally distributed), as above.

2.2.3 MBSS sampling unit comparisons with refraction and discharge difference correction

2.2.3.a Refraction correction

A simple refraction correction was applied to depths to assess the degree of improvement of MBSS sampling unit depth and/or velocity categories. A refraction correction of 1.34 (Woodget et al. 2015; Dietrich 2017) was used as a multiplier of pixelated depth values. Corrected values were then averaged across segments, (as described in section 2.2.1.b).

2.2.3.b Field versus aerial discharge difference correction

For MBSS segment comparisons, depth was corrected relative to discharge differences between field and aerial survey dates as a coarse assessment of whether correcting for these differences improved correlations and associations. The ratio between the discharges on the two survey dates (Table 1.3.3) was applied to hydraulic geometry relationships between discharge (Q) and width (w), depth (d) and velocity (v), respectively:

$$w = aQ^b \; ; \; d = \; cQ^f \; ; v = \; kQ^m$$

Average values for these exponents from Leopold and Maddock's (1953) work (b = 0.5, f = 0.4 and m = 0.1), were used as a coarse representation of hydraulic geometry relationships in the study streams. These relationships were used to solve for correction factors for width, depth and velocity. The depth relationship is used below as an example of this derivation, with F and A subscripts denoting field and aerial values, respectively. The direct comparison of values between field and aerial surveys suggest the values should be the same (e.g., the goal is to find a correction factor for d_A relative to the difference in Q_A and Q_F so that $d_{Acorr} = d_F$):

$\mathrm{d_F} = \mathrm{cQ_F}^{0.4}, \mathrm{d_A} = \mathrm{cQ_A}^{0.4}$, and $\mathrm{d_F}\ (\text{should}) = \mathrm{d_{Acorr}}$, If Q_F (should) $= Q_A$, but $Q_F = R * Q_A$,

then $d_F = c{Q_F}^{0.4} \rightarrow c[R * Q_A]^{0.4} = d_{Acorr}$.

So $d_{Acorr} = c[R * Q_A]^{0.4} \rightarrow R^{0.4} * c * Q_A^{0.4} \rightarrow R^{0.4} * [d_A] \rightarrow R^{0.4} d_A$

And ultimately, $d_{Acorr} = d_A * R^{0.4}$.

This derivation shows that, if the discharge-corrected aerially-derived depth (d_{Acorr}) should be equal to the field measured depth (d_F), the relationship between depth and discharge (Leopold and Maddock Jr. 1953) could be used in conjunction with the ratio between the field and aerial survey discharges to find a correction for d_A.

Similar relationships were established for width and velocity, so that corrected values of each were a function of the original SfM values and site discharge ratios:

$$w_{Acorr} = w_A * R^{0.5};\ d_{Acorr} = d_A * R^{0.4};\ v_{Acorr} = v_A * R^{0.1}$$

Corrected depth values were then summarized across segments to find the average depth per segment, which was converted to ordered categories, as described in section 2.2.1.b.

2.2.3.c Combined refraction and discharge difference correction

Refraction-corrected depths were corrected for discharge using the depth equation in section 2.2.3.b. Cross-sectional mean depth was calculated from the refraction and discharge-corrected depths. The average values of corrected depths per MBSS segment were calculated, converted to ordered categories, and compared to field values, as in section 2.2.1.b.

Combined refraction and discharge difference corrections were also performed to assess improvements to flow velocity comparisons at MBSS segments. Cross-sectional mean width values corrected for discharge using the width equation in section 2.2.3.b were multiplied by the refraction and discharge-corrected cross-sectional mean depth. The resulting corrected cross-sectional area was used with the aerial survey date discharge (as in section 1.2.4.f) to find the refraction and discharge-corrected cross-sectional mean velocity. Average values of corrected velocities per MBSS segment were calculated, converted to ordered categories, and compared to field values, as in section 2.2.1.b.

2.3 Results

2.3.1 Paired sampling unit metric comparisons

2.3.1.a EMAP / SWAMP transects, PA sites

Strong correlations were evident between field and SfM estimates of paired EMAP / SWAMP transect widths (Fig. 2.3.1). Correlation coefficients and RMSE values are displayed below each width comparison to represent trends in correspondence and error between the paired transects across SfM measurement methods (Fig. 2.3.1). Discrepancies between field-measured and SfM-derived channel widths showed distinct effects of SfM measurement methods for both transect placement (AP vs MP) and measurement (EM vs MM; Fig. 2.3.1). AP transects resulted in greater variation in differences between field and SfM values, as evidenced by the larger ranges in boxplots for AP transects relative to those of MP transects. For both measurement types, AP transects had lower correspondence than MP transects, with correlation coefficients of 0.85 versus 0.93, respectively, for EM widths, and 0.83 versus 0.98, respectively, for MM widths (Fig. 2.3.1).

These differences portrayed the sensitivity of transect width results relative to transect placement (location and orientation). EM displayed a negative bias of lower widths in SfM measurements relative to field measurements. This is evident by the lower mean and median values of SfM-field difference for EM widths relative to MM widths for both placement methods. There was strong evidence that the median difference in width for EM of both placement methods was different than zero, as evidenced by the notched portion of the boxplots lying completely below the zero-difference line (Chambers et al. 1983). Alternatively, both placement methods for MM transects suggested the median difference between SfM and field widths did not differ from zero, given the inclusion of zero in the notched boxplot. For AP transects, the EM width is better correlated with field measurements, but with slightly lower (though comparable) error (Fig. 2.3.1). Alternatively, the SfM AP-EM combination lead to both mean and median differences from field widths that were the farthest from zero of the paired comparisons (Fig. 2.3.1). The strongest positive correlation between field and SfM width with the lowest error and with the mean and median difference the closest to 0 was for MP-MM transects with manual measurement of width (Pearson's $r = 0.98$; RMSE = 0.61m; Fig. 2.3.1).

The differences between EM and MM likely resulted from the way EM was calculated in the GIS environment. As indicated in Chapter 1 (section 1.2.4.d), the Euclidean distance-based width relied on the assumption that the centroid of edge pixels approximately overlaid the wetted edge. However, during conversion between shapefile and raster, the edge pixel may have snapped to a pixel within the channel. This would decrease the Euclidean measured width relative to manually measuring the distance between the original polyline shapefiles. The bias was not likely majorly a result of discharge differences between field and aerial survey dates, given the bias was present for EM comparisons, but not for MM comparisons. Similarly, the bias indicated SfM widths were lower than field widths for EM comparisons, but the discharge during the aerial survey was greater than the discharge during the field surveys at these sites (e.g., 1.3.3, sites CRAB, EBCH and PLUM). Widths would not be lower because of a higher discharge.

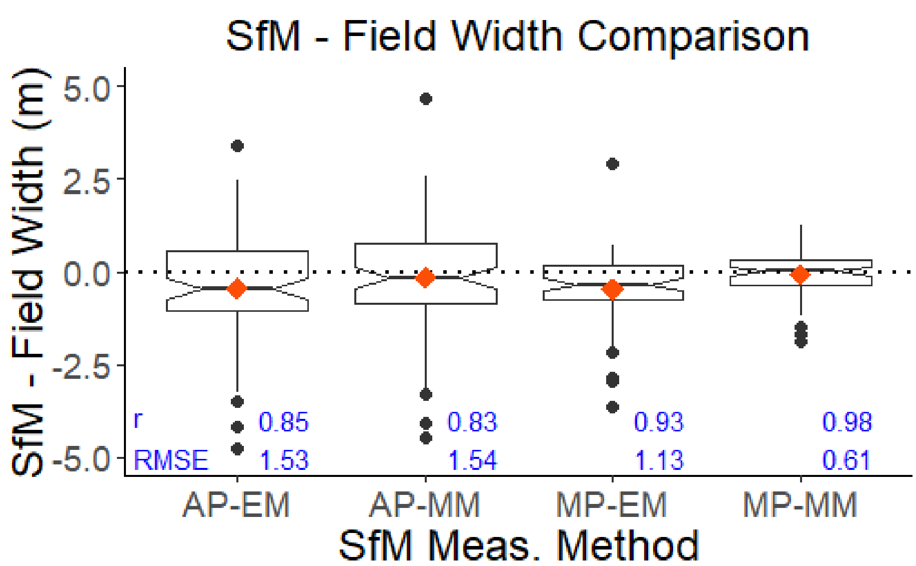


Figure 2.3.1. Width difference between SfM and field measurements of paired transects at Pennsylvania sites. Mean differences per distribution are represented by red points.

Field-measured depths were subtracted from SfM-derived depths to display the distribution of difference between the paired cross-section transect sampling points for each extraction method. Figure 2.3.2 displays these distributions, comparing the different methods of transect placement (AP vs MP) and of null value replacement (NR vs WR). The median differences between SfM and field depths for WR values strongly differed from zero, as evidenced by the notched portion of the boxplots lying completely below the zero-difference line (Chambers et al. 1983). The median differences in depth for points along automatically placed transects were closer to zero than those of MP transects for both null replacement methods. However, there was more variation around the mean for the AP transects, as evidenced by the higher RMSE values. For AP transects, WR reduced the correlation strength (r = 0.24) and increased error (RMSE = 0.27m) relative to the NR distribution (r = 0.33, RMSE = 0.14m; Fig.

2.3.2b). Alternatively, for MP transects, WR improved the strength of the correlation (r=0.78) and reduced error (RMSE = 0.08m) relative to NR (r = 0.74 and RMSE = 0.09; Fig. 2.3.2b).

Replacing null depth values with the underlying cross-sectional mean depth creates a negative bias in Figure 2.3.2 (e.g., SfM WR depths < paired field depths). The differences between NR and WR likely resulted from the location of null values within SfM. Null depth was common where vegetation covered the edge of the channel, which itself was common where the thalweg was at or near the channel edge (Fig. 2.3.3). Therefore, if these large values were replaced by the cross-sectional mean depth, which would be inherently less than the thalweg depth, the resulting SfM depth would be less than the field depth at that point. However, the negative bias was not only present in the WR comparisons. The MPNR comparison also showed a negative bias (e.g., SfM depths < field depths). Therefore, the bias could also have resulted from refraction, where the channel bed elevation was modeled as higher than reality (and therefore depth would be lower than in reality) due to light refraction. Again, the discharge differences between aerial and field survey dates were not likely to have created the negative biases in depth. Aerial survey discharges were greater than field survey discharges at these sites (Table 1.3.3, sites CRAB, EBCH and PLUM), so on average, aerially surveyed depths should not have been lower than field depths.

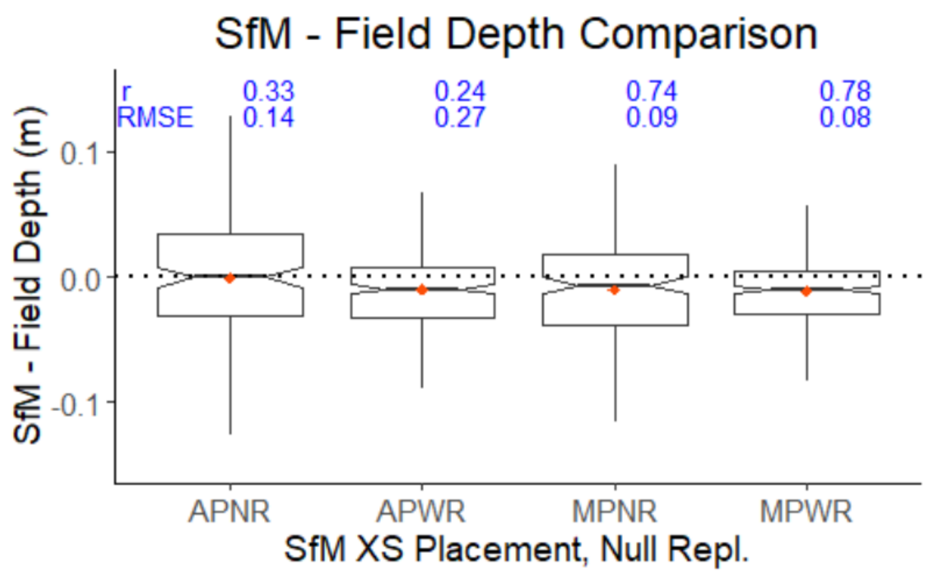


Figure 2.3.2. Depth difference between SfM and field measurements of paired transect sampling points at Pennsylvania sites with outliers excluded from the display, but not the analyses. Mean differences per distribution are represented by red points.

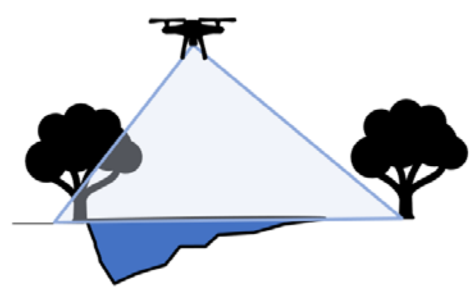


Figure 2.3.3. Depiction of edges that may have large values missed by SfM due to vegetation covering the edge of the channel.

Flow velocities categorized in the field were compared to SfM categories for points along AP and MP transects. Table 2.3.1 shows the contingency tables for these comparisons. Results of linear-by-linear association tests for both methods of transect placement indicated the null hypothesis of independence could be rejected at $\alpha = 0.05$. Given that the categories are ordinal (e.g., RI is faster than GL), there was evidence of higher velocities in the SfM than in the field. This was indicated by larger numbers in the upper right of the two contingency tables than in the lower left, which indicated that many points had higher velocity categories in the SfM than in the field. This was likely a result of both the nature of the velocity calculation for SfM values, and the discharge differences. SfM velocities were cross-sectional averages, which meant lateral velocity variation was not captured. Laterally, depth and velocity have more of a positive relationship, where low depths have lower velocities than high depths (e.g., the thalweg). Similarly, differences in discharge would more directly impact SfM velocities, given the reliance on discharge in the cross-sectional mean velocity calculation (e.g., $v = Q / wd$). Lastly, drainage area correction of discharge (and therefore of cross-sectional mean velocities) was not performed for these comparisons, so larger SfM velocity values could have resulted from the assumption of continuity of discharge throughout the sites.

Table 2.3.1. Contingency tables comparing categorical cross-section flow velocities at paired points, with SfM transects placed automatically (left) or manually (right).

		SfM, AP				
		EXP (0)	P (1)	GL (2)	RI (3)	Sum
Field	EXP (0)	10	1	7	11	29
	P (1)	0	14	51	85	150
	GL (2)	0	13	51	63	127
	RI (3)	2	7	23	56	88
	Sum	12	35	132	215	394
Linear-by-linear test $p = 2.65 \times 10^{-4}$						

		SfM, MP				
		EXP (0)	P (1)	GL (2)	RI (3)	Sum
Field	EXP (0)	21	0	5	3	29
	P (1)	0	6	49	95	150
	GL (2)	0	6	49	72	127
	RI (3)	0	1	23	64	88
	Sum	21	13	126	234	394
Linear-by-linear test $p = 1.42 \times 10^{-8}$						

Sediment categorized at transect points was compared between field and SfM for different methods of transect placement (Table 2.3.2). Again, results of linear-by-linear association tests for both methods of transect placement indicated the null hypothesis could be rejected at $\alpha = 0.05$. Most points where sediment did not exactly match were within one size category. Differences between the sediment categories more than one size category different were likely a result of the nature of sediment delineation within SfM. Sediment patches within SfM were delineated to represent the dominant sediment within patches. Therefore, small sediment between larger sediment that is locally dominant might be picked up in the field survey but not by the SfM sediment delineation methods.

Table 2.3.2. Contingency tables comparing sediment categorized in the field versus within SfM. Sampling point along AP transects (top) are compared to field data separately from points along MP transects (bottom).

		SfM, AP						
		FN (1)	SA (2)	GR (3)	CB (4)	B (5)	RS (6)	Sum
Field	FN (1)	0	4	5	4	0	0	13
	SA (2)	0	16	27	16	2	0	61
	GR (3)	0	24	73	55	4	0	156
	CB (4)	0	11	44	33	2	0	90
	B (5)	0	1	25	25	2	0	53
	RS (6)	0	0	2	0	0	0	2
	Sum	0	56	176	133	10	0	375
Linear-by-linear test p-value = 1.6×10^{-4}								

		SfM, MP						
		FN (1)	SA (2)	GR (3)	CB (4)	B (5)	RS (6)	Sum
Field	FN (1)	0	5	2	3	0	0	10
	SA (2)	0	18	28	14	0	0	60
	GR (3)	0	24	81	50	2	0	157
	CB (4)	0	15	38	38	0	0	91
	B (5)	0	2	18	30	3	0	53
	RS (6)	0	0	2	0	0	0	2
	Sum	0	64	169	135	5	0	373
Linear-by-linear test p-value = 1.5×10^{-6}								

2.3.1.b MBSS segments, MD sites

Categorized mean depths per segment were displayed in a contingency table comparing field to SfM results at paired segments (Table 2.3.3). This comparison could not be assessed with a linear-by-linear association test, as all SfM depths were categorized the same (only had one level). However, approximately 79% (57/72) segments were categorized the same as paired field segments.

Table 2.3.3. Contingency table comparing depth categories between SfM and field data in paired MBSS segments.

		SfM			
		1	2	3	Sum
Field	1	57	0	0	57
	2	13	0	0	13
	3	2	0	0	2
	Sum	72	0	0	72

Categorized flow velocities were compared between field to SfM results at paired segments (Table 2.3.4). Within the SfM data, each segment was summarized using the mean or median cross-sectional velocity (left or right, respectively, in Table 2.3.4). Results of each linear-by-linear association test indicated that the null hypothesis of independence could not be rejected ($p > 0.05$). Approximately 53% (38/72) of mean-summarized segments, and approximately 64% (46/72) of median-summarized segments, were categorized the same as paired field segments.

Table 2.3.4. Contingency tables comparing flow categories between SfM (left is mean per segment, right is median) and field data in paired segments.

		SfM, mn.					SfM, med.		
		1	2	Sum			1	2	Sum
Field	1	33	22	55	Field	1	44	11	55
	2	12	5	17		2	15	2	17
	Sum	45	27	72		Sum	59	13	72
Linear-by-linear p = 0.43					Linear-by-linear p = 0.44				

Table 2.3.5 displays the paired segment comparison between dominant sediment categorized in the SfM and field data. Results of the linear-by-linear association test indicated that the null hypothesis of independence could be rejected ($p = 0.017$). Approximately 57% (41/72) of segments were categorized the same as paired field segments.

Table 2.3.5. Contingency table comparing sediment categories between SfM and field data in paired segments.

		SfM				
		Y (1)	S (2)	G (3)	C (4)	Sum
Field	Y (1)	0	0	0	0	0
	S (2)	4	19	15	2	40
	G (3)	0	4	22	1	27
	C (4)	0	2	3	0	5
	Sum	4	25	40	3	72
Linear-by-linear p = 0.017						

2.3.2 Paired reach-aggregated metric comparisons

2.3.2.a EMAP / SWAMP transects, PA sites

The mean width per reach was calculated from the 11 transects comprising each sampling reach. Figure 2.3.4 displays the correlation between the SfM and field mean widths per reach for the four SfM sampling methods: automatically placed transects that were Euclidean distance measured ("APEM"), automatically placed transects that were manually measured ("APMM"), manually placed transects that were Euclidean distance measured ("MPEM"), and manually placed transects that were manually measured ("MPMM").

The correlation coefficient (r) and significance level (p), along with the slope and intercept of a linear regression, is tabulated on the plot for each comparison of mean width per reach. Correlation was used to analyze the relationships, and regression was just used for qualitative analysis of how close the relationships were to a 1:1 line. All four comparisons had significant strong positive correlations ($r = 0.98 - 0.99$, $p < 0.001$). However, the comparison per reach between mean field width and mean SfM MPMM width was the closest to a 1:1 relationship. This relationship indicated that at a given reach, the mean SfM MPMM width was equal to 0.09m plus 0.97 times the mean field width (Fig. 2.3.4). Error bars represent the standard error of measurements per reach (field is horizontal and SfM is vertical; Fig. 2.3.4). These error bars indicate that most reaches had similar magnitude of standard error between field and SfM measurements, but for a few reaches, there was slightly more variation in the field measured mean width per reach than in the SfM measured mean width per reach.

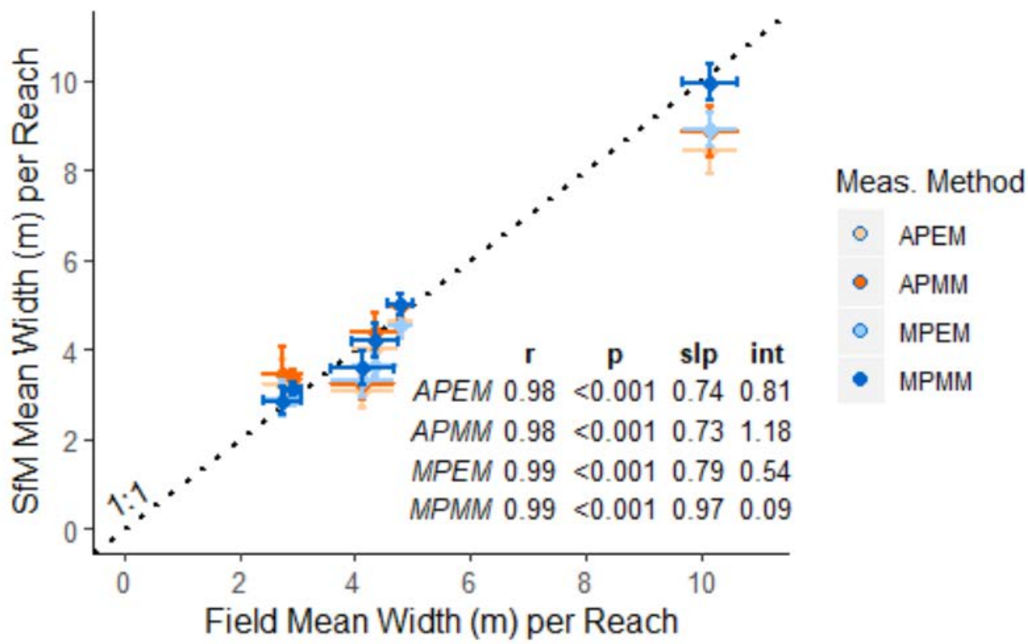


Figure 2.3.4. Linear correlation between field and SfM-derived reach-averaged width at Pennsylvania sites (n=6). The correlation coefficient (r) and significance (p), as well as the linear regression slopes and intercepts, are tabulated per comparison.

The mean depth per reach was calculated from all sampling points within each reach. Figure 2.3.5 displays the correlation between paired SfM-field reach comparisons of mean depth values for each SfM sampling method. The correlation coefficient (r) and significance (p) (for quantitative analysis), along with the slope and intercept of the linear regression (for qualitative analysis), were tabulated on the plot for each comparison of mean depth per reach. All four comparisons had significant strong positive correlations ($r = 0.88 - 0.99$, $p < 0.05$). However, the comparison between reach-averaged field depth and SfM MPWR depth was the closest to a 1:1 regression relationship. This relationship indicated that at a given reach, the mean SfM MPWR depth was equal to the mean field depth less 0.02m (Fig. 2.3.5). Error bars represent the standard error of measurements per reach (field is horizontal and SfM is vertical; Fig. 2.3.5). From these error bars, it appeared that most reaches had similar magnitudes of standard error between field and SfM measurements. However, there were a few reaches with slightly more variation in the field measured depths than in the SfM measured depths, and one reach (with the largest mean depth) where the SfM APWR depths had much more variation than the paired field depths.

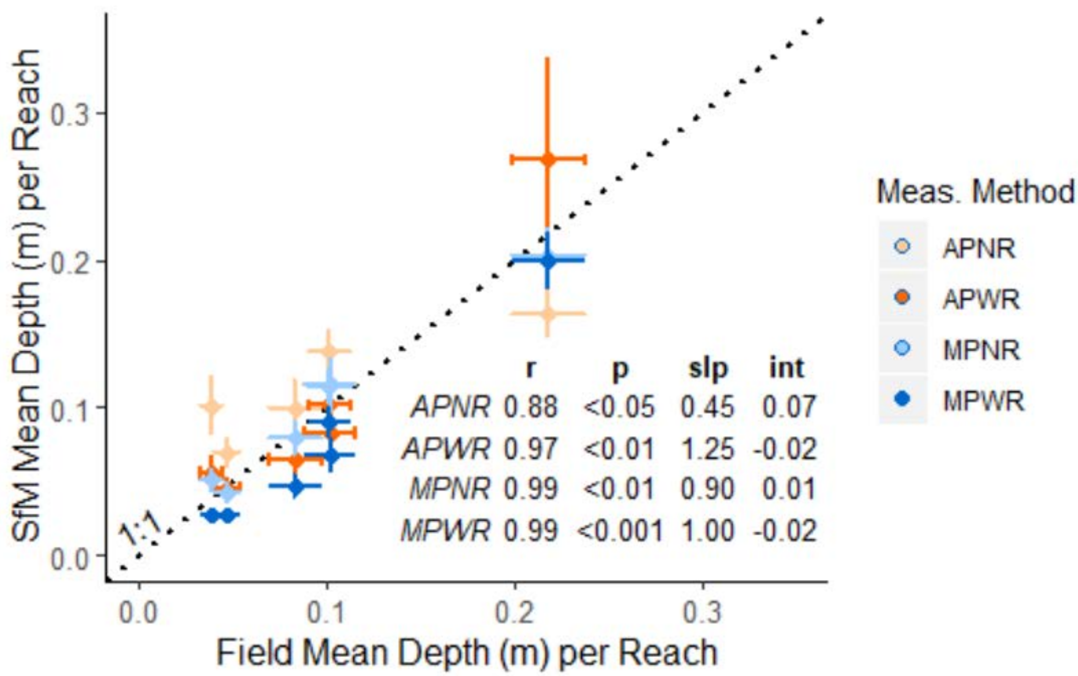
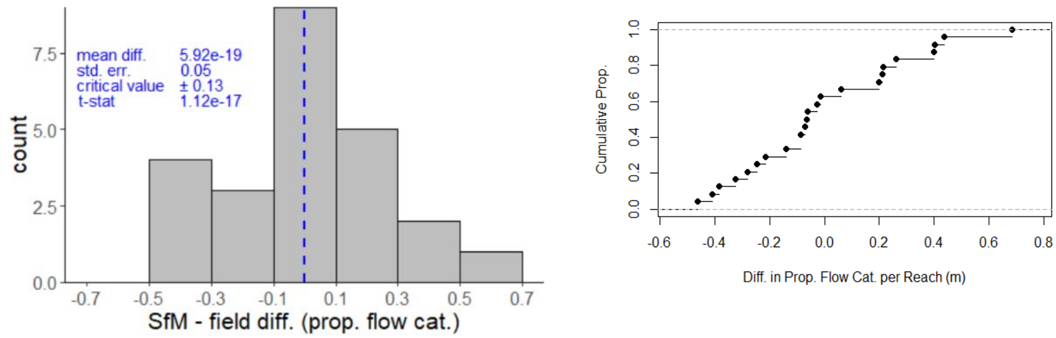


Figure 2.3.5. Linear correlation between field and SfM-derived reach-averaged depth at Pennsylvania sites. AP = automatically placed transects, MP = manually placed transects, NR = no null value replacement, and WR = null values replaced with cross-sectional mean depth. The correlation coefficient (r) and significance (p), as well as the linear regression slopes and intercepts, are tabulated per comparison.

Each flow velocity category was summarized as a proportion per reach, and the difference in proportions per reach between SfM data and field data were compared (4 flow categories for 6 reaches, so the maximum n of the comparison was 24, if all flow categories were found within each reach). Figure 2.3.6 displays the distribution of differences between field and

SfM proportions per category per reach as a histogram and cumulative frequency distribution (CFD) separately for automatically versus manually placed transects. The mean difference, as well as results of the paired t-test for equivalence (PTTE) are displayed on each histogram. The t-statistic was less than the critical value in both tests: t-statistic of 1.13×10^{-17} is less than a critical value of 0.13 for the comparison between field data and automatically placed transects, and a t-statistic of -3.01×10^{-17} is less than a critical value of 0.13 for the comparison between field data and manually placed transects. In each case, this indicated the null hypothesis of dissimilarity between paired field and SfM flow category proportions per reach could be rejected (Wellek 2003, 2010; Robinson and Froese 2004).

(a)



(b)

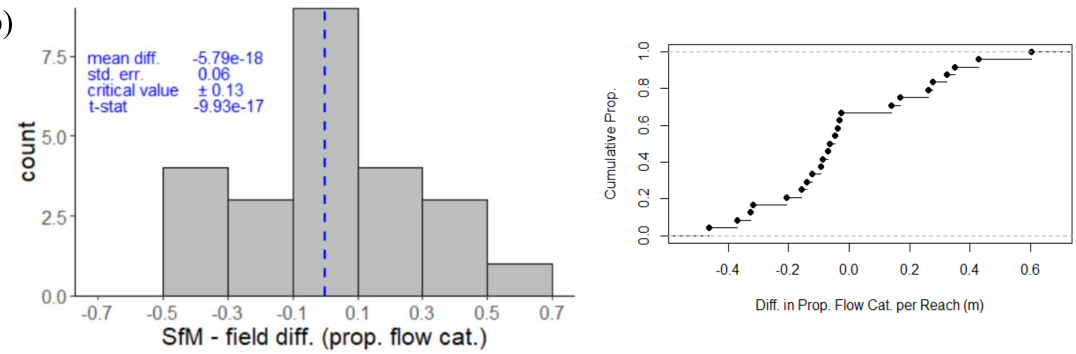


Figure 2.3.6. Distributions (histograms and cumulative frequency distributions) of the difference in the reach-wide proportions of each flow category between categorized SfM and field data at Pennsylvania sites. (a) Automatically placed transects. (b) Manually placed transects. The means and standard deviations of the differences, as well as results of PTTEs, are displayed on each histogram. The mean difference in proportions is also portrayed by the vertical blue dashed lines.

Each sediment category was summarized as a proportion per reach, and the proportions per category for SfM data was compared to those for field data at paired reaches (8 sediment categories and 6 reaches, so the maximum n of the comparison 48, if all sediment categories are found within each reach). Figure 2.3.7 displays the distribution of differences between field and SfM proportions per category per reach as a histogram and cumulative frequency distribution (CFD) separately for automatically versus manually placed transects. Neither distribution was normal, as indicated by Shapiro-Wilk normality tests resulting in p-values of 6.02×10^{-5} and 2.38×10^{-5} for automatically and manually placed SfM comparisons, respectively). Therefore, robust two one-sided tests (RTOSTs) were run instead of PTTEs, the results of which are displayed on each histogram, along with the mean differences. The p-value of each RTOST was < 0.05, suggesting that the null hypothesis of dissimilarity could be rejected.

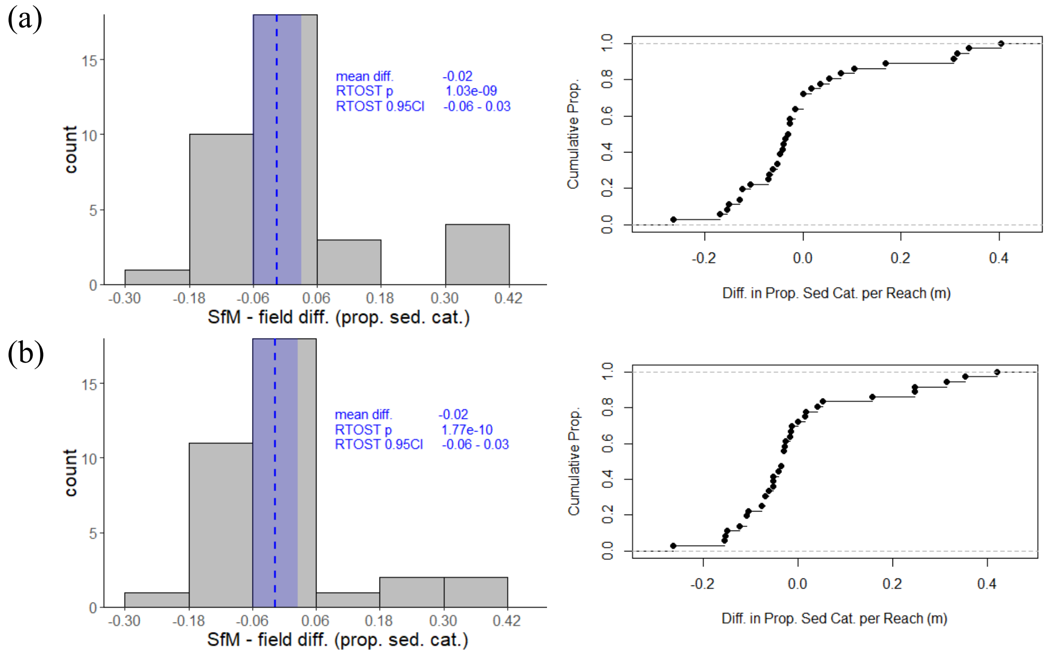


Figure 2.3.7. Distributions of the difference in the reach-wide proportions of each sediment category between SfM-derived sediment and field-measured sediment at Pennsylvania sites. (a) Automatically placed transects. (b) Manually placed transects. Mean differences, as well as results of the RTOSTs (p-value and 95% CI) are displayed on each histogram. The mean difference in proportions is also portrayed by the vertical blue dashed lines, and the confidence intervals are indicated by the blue regions.

2.3.2.b MBSS segments, MD sites

For the MBSS segments, each depth category was summarized as a proportion per reach by finding the quotient of the number of segments categorized the same and the 6 segments per reach. These proportions per category of SfM data were compared to those for field data at paired reaches (3 depth categories and 12 reaches, so the maximum n of the comparison was 36 if all depth categories are found within each reach). Figure 2.3.8 displays the histogram and cumulative frequency distribution of differences in proportions, with field proportion subtracted from SfM proportion per depth category per reach. The mean difference, as well as results of the paired t-test for equivalence (PTTE) are displayed on the histogram. The t-statistic (1.3×10^{-16}) was less than the critical value (0.12), which indicated the null hypothesis of dissimilarity between paired field and SfM depth category proportions per reach could be rejected (Wellek 2003, 2010; Robinson and Froese 2004).

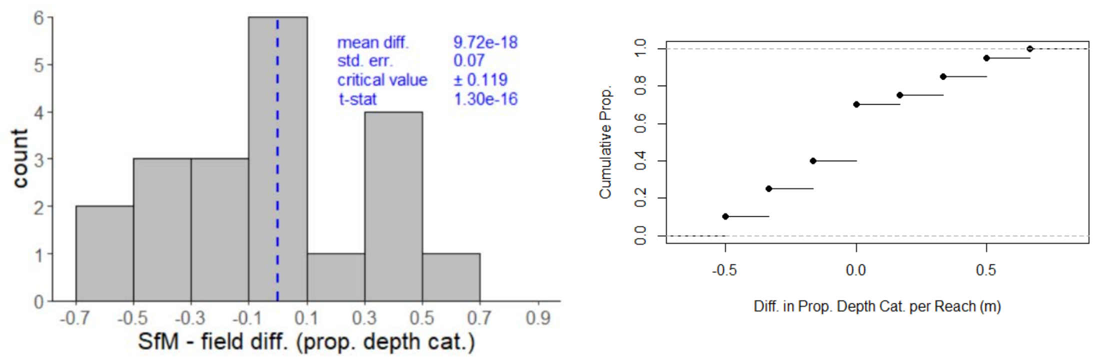


Figure 2.3.8. Distributions of the difference in the reach-wide proportions of depth categories between SfM- and field-measured depth at Maryland sites. The mean and standard deviation of the differences, as well as results of PTTEs, are displayed on each histogram. The mean difference in proportions is also portrayed by the vertical blue dashed line.

Each segment flow category was summarized as a proportion per reach by finding the quotient of the number of segments categorized the same and the 6 segments per reach. These proportions per category of SfM data were compared to those for field data at paired reaches (2 flow categories and 12 reaches, so the maximum n of the comparison was 24, if all flow categories were found within each reach). This was completed separately for segments categorized from mean cross-sectional flow velocity per segment and those categorized from median cross-sectional flow velocity per segment. Figure 2.3.9 displays the histogram and cumulative frequency distribution for each comparison of differences in proportions, with field proportion subtracted from SfM proportion per flow category per reach. The mean difference, as well as results of the paired t-test for equivalence (PTTE) are displayed on each histogram. The t-statistic was less than the critical value in both tests: t-statistic of -4.27×10^{-17} is less than a critical value of 0.13 for the comparison between field proportions and SfM proportions from mean-summarized SfM segments, and a t-statistic of -8.9×10^{-17} is less than a critical value of 0.13 for the comparison between field proportions and SfM proportions from median-summarized SfM segments. In each case, this indicated the null hypothesis of dissimilarity between paired field and SfM flow category proportions per reach could be rejected (Wellek 2003, 2010; Robinson and Froese 2004).

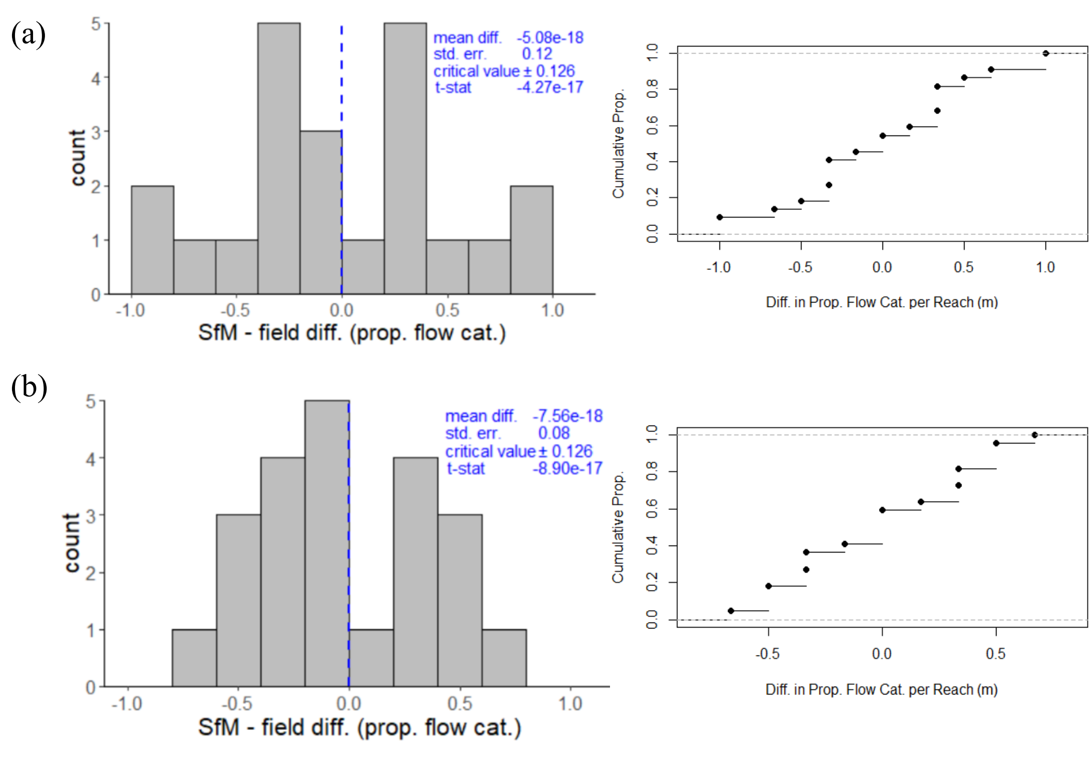


Figure 2.3.9. Distributions of the difference in the reach-wide proportions of flow categories between SfM- and field-measured flow velocity at Maryland sites. (a) SfM mean velocity per segment. (b) SfM median velocity per segment. The means and standard deviations of the differences, as well as results of PTTEs, are displayed on each histogram. The mean difference in proportions is also portrayed by the vertical blue dashed lines.

Each sediment category from MBSS "facies" maps was summarized as a proportion per reach by finding the quotient of the number of segments categorized the same and the 6 segments per reach. These proportions per category of SfM data were compared to those for field data at paired reaches (5 sediment categories and 12 reaches, so the maximum n of the comparison was

60, if all sediment categories were found within each reach). Figure 2.3.10 displays the histogram and cumulative frequency distribution of differences in proportions, with field proportion subtracted from SfM proportion per sediment category per reach. A PTTE indicates the null hypothesis of dissimilarity could be rejected (t-statistic ~0 < critical value 0.156, Fig. 2.3.10).

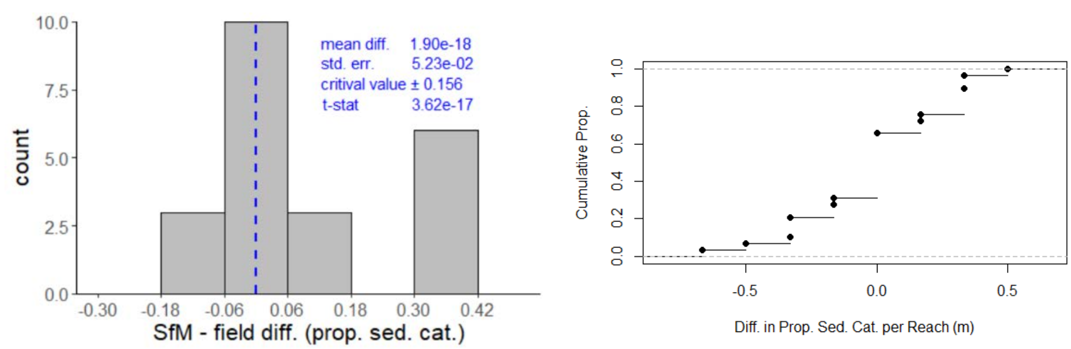


Figure 2.3.10. Distributions of the difference in the reach-wide proportions of sediment categories between SfM- and field-estimated dominant sediment at Maryland sites. The means and standard deviations of the differences, as well as results of PTTEs, are displayed on each histogram. The mean difference in proportions is also portrayed by the vertical blue dashed lines.

2.3.3 MBSS sampling unit comparisons with refraction and discharge difference correction

The middle of the restored reach at site WL was used to display changes in pixelated depth after corrections (Table 2.3.6; Fig. 2.3.11). This area was chosen for display, as the river right segment (displayed on the left in Fig. 2.3.11) was identified as a "1" depth category by uncorrected SfM values, but a "2" depth category in the field survey (Table 2.3.6). Similarly, WL had the highest ratio between field and aerial survey date discharges, as the field survey date discharge was approximately 2.8 times the aerial survey date discharge (Table 2.3.8). During the field survey, the pool in the right segment upper left of the orthoimage in Fig. 2.3.11 was not wadable, with areas over 1.0m in depth. However, the highest values in the pool for uncorrected aerial survey depths were approximately 0.5-0.6m. Therefore, it was evident that aerially derived depth needed correction for adequate comparison to field values.

Table 2.3.6. Depth values within the restored 25-50m right and left MBSS segments at site WL.

site	reach	seg.	side	Field category	(a) Uncorr. depth mean (m)	(b) Refrac. corr. depth mean (m)	(c) Q diff. corr. depth mean (m)	(d) Refrac. and Q diff. corr. depth mean (m)
WL	REST	25	L	1	0.20	0.27	0.31	0.41
WL	REST	25	R	2	0.28	0.38	0.42	0.57

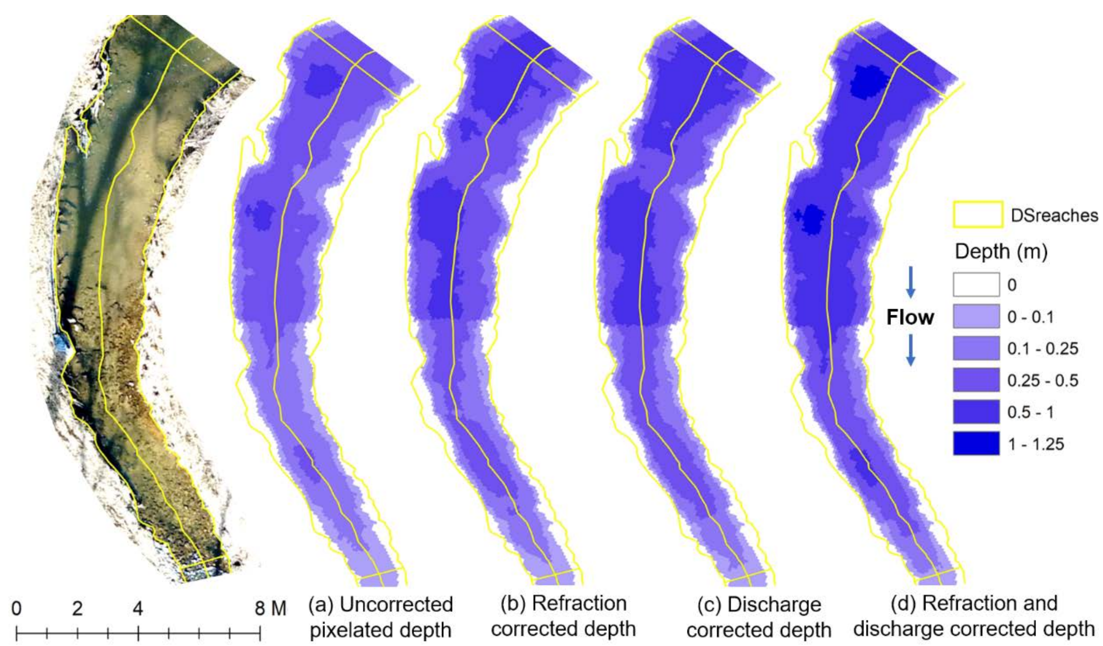


Figure 2.3.11. Depth correction within the restored reach of site WL. Orthophoto was brightened for display purposes. Flow is from top to bottom of the images, so river right is the segment on the left in the figure.

2.3.3.a Refraction correction

The comparison between Figure 2.3.11 (a) and (b) display that refraction correction increased point depth estimates within segments. As evident from Table 2.3.6, the mean refraction-corrected depth per segment did increase relative to the mean uncorrected depth per segment (Table 2.3.6, a-b). However, all aerially derived depth categories remained "1" at site WL (e.g., mean depth remained <0.5m; Table 2.3.6), so the comparison analysis between field and SfM values did not improve for this site. One segment from another reach (WL DS) did increase from category “1” to “2” after refraction correction, so an updated analysis was possible (given SfM depth had more than one level). However, the segment updated to a “2” depth category was categorized as a “1” in the field, so refraction correction did not result in an improved analysis, and comparison with field data showed no significant association between field and SfM depth categories (p = 0.63; Table 2.3.7).

Table 2.3.7. MBSS segment depth category contingency table, with SfM depth categories per segment derived from depth values corrected for refraction.

		SfM, Refrac. Corr.			
		1	2	3	Sum
Field	1	56	1	0	57
	2	13	0	0	13
	3	2	0	0	2
	Sum	71	1	0	72
Linear-by-linear p = 0.63					

2.3.3.b Discharge difference correction

Table 2.3.8 was modified from Table 1.3.3 to display just the field-aerial survey date discharge ratio per site. The remainder of the table represents the multipliers applied to width, depth and velocity for discharge correction, given the equations derived in section 2.2.3.b. Site WL had the largest discharge difference correction multipliers of the study sites for all three metrics (Table 2.3.8).

Table 2.3.8. Field and aerial survey date discharge ratios and correction factors derived from equations in section 2.2.3.b.

	144	253	CRAB	EBCH	PLUM, Rest	PLUM, US	ULP2	WL
Field / aerial Q (R)	0.337	2.011	0.532	0.335	0.669	0.302	1.579	2.762
Width correction factor ($R^{0.5}$)	0.580	1.418	0.729	0.579	0.818	0.550	1.257	1.662
Depth correction factor ($R^{0.4}$)	0.647	1.322	0.777	0.646	0.852	0.620	1.201	1.501
Velocity correction factor ($R^{0.1}$)	0.897	1.072	0.939	0.896	0.961	0.887	1.047	1.107

The discharge correction increased point based SfM depth estimates more than the refraction correction (Fig. 2.3.11 a-c), given the correction factor for depth at WL was approximately 1.5 (Table 2.3.8), while the refraction correction factor for depth was 1.34 (section 2.2.3.a). Again, the depth categories of the segments of focus in Fig. 2.3.10 did not change (e.g., mean discharge difference corrected depths per segment were <0.5m), though mean depths per segment did increase relative to uncorrected values (Table 2.3.6 a and c). But the same segment within the WL DS reach changed from a “1” to “2” depth category. Therefore, the results of the analysis comparing depth categories between field surveys and aerial surveys with data corrected for discharge differences yielded the same results as for refraction correction (Table 2.3.7).

2.3.3.c Combined refraction and discharge difference corrections

The combination of refraction and discharge difference corrections yielded the largest depth values within the pool of focus at site WL (Fig. 2.3.11 d). The deepest parts of the pool were between approximately 1.0-1.1m, which more closely approximated what was observed in the field in that location. Similarly, the mean depth corrected for both refraction and discharges differences within the river right segment was 0.57m, a “2” depth category, which matched the field assignment (Table 2.3.6 d). Table 2.3.9 displays the contingency table and analysis results comparing the combined corrections of SfM depths with field depths. Four segments were moved from the "1" to "2" depth category, reducing the matches with field segments categorized as a "1" for depth. However, 2 segments, including the river right segment of focus in Figure 2.3.11, were matched in the "2" depth category.

Table 2.3.9. MBSS segment depth category contingency table, with SfM depth corrected for refraction and discharge differences per site.

		SfM, Refrac. & Q Diff. Corr.			
		1	2	3	Sum
Field	1	53	4	0	57
	2	11	2	0	13
	3	2	0	0	2
	Sum	66	6	0	72
Linear-by-linear p = 0.61					

Velocities derived from refraction and discharge difference corrected values were also improved relative to uncorrected data. Table 2.3.10 displays velocity contingency tables for SfM velocities derived from corrected SfM velocities categorized based on mean and median velocity per segment. Both show improved association relative to uncorrected comparisons, neither of which were significant (Table 2.3.4), but only the median-summarized corrected velocity categories were significantly associated with field categories ($p < 0.05$). As with uncorrected values, summarizing velocities using medians instead of means per segment improved comparisons, likely because outliers in velocity within the segment were driving mean values that were less representative of the segment than median values.

Table 2.3.10. MBSS segment flow velocity category contingency table, with SfM velocity corrected for refraction and discharge differences per site.

		SfM, mn., Refrac. & Q Diff. Corr.		
		1	2	Sum
Field	1	39	16	55
	2	15	2	17
	Sum	54	18	72
Linear-by-linear p = 0.15				

		SfM, med., Refrac. & Q Diff. Corr.		
		1	2	Sum
Field	1	43	12	55
	2	17	0	17
	Sum	60	12	72
Linear-by-linear p = 0.04				

2.4 Discussion

Analyses revealed there was substantial agreement in the representation of field reaches by field and SfM habitat data (depth, width, sediment and velocity) when extracted at comparable intervals. Flexibility in SfM data extraction and correction allowed for improved comparisons at both the sampling unit and reach-aggregated levels. Even when sampling unit comparisons were subject to error, field and SfM data equivalently represented reaches at the reach-aggregated level important to many multimetric indices and models.

A key finding of this study conveyed the power and flexibility of SfM data. Comparisons between field and remotely sensed data were improved when adjusted for data extraction methods (e.g., sampling unit placement; null replacement; mean vs. median summarized segments), and corrections (e.g., refraction and discharge). SfM data was revealed to be locally heterogeneous, given the ability to obtain different results from slight adjustments to EMAP / SWAMP sampling unit placement. Reach-aggregated correlations and equivalencies varied with measurement or estimation methods in a similar manner to the sampling unit -based comparisons. Similarly, summaries of MBSS segments using median instead of mean flow velocities improved comparisons with field data, as median estimates were likely less subject to outliers in flow velocity. Ultimately, field and SfM surveys were comparable for most analyses because of the ability to flexibly adjust and correct the summaries of SfM data.

Other researchers have also demonstrated equivalence between field and remote measurement of physical channel features when sampled at similar intervals, particularly when remote data were flexibly summarized. Marcus et al. (2003) assessed the ability of high spatial resolution hyperspectral (HSRH) mapping to classify and map stream channel habitat types continuously across three reaches (2-5km). They found that field and remote methods represented trends in study systems in an equivalent manner, especially when they employed fuzzy boundaries to habitat types within the remote data to account for the subjectivities in manually mapping exact boundaries in the field. This similarly indicated that flexibility in remotely sensed data extraction may be necessary to account for subjectivities inherent in field estimates. Woodget et al. (2015, 2016) compared flow mapping between field and aerial surveys across a 50-m reach and noted similar findings in the ability of both methods to pick up overall trends of habitat units within the streams. Such findings suggest that remote assessment and field data provide equivalent representations of site conditions when sampled at analogously coarse intervals.

Part of the variation between paired SfM and field measurements was likely explained by the discharge difference between the measurements. Kauffman et al. (1999) define flow-dependent measures as particularly susceptible to low precision, given inevitable variability in flow over time. However, these types of measures (e.g., depth, width and velocity), are widely collected to assess stream physical habitat conditions. While this study and others perform assessments at baseflow to minimize flow variability, baseflow varies throughout the year (Fig 2.4.1; Mohamoud 2004). Figure 2.4.1 shows conceptual variation in the water balance, as represented by potential evapotranspiration (PET), precipitation (PREC), and streamflow, for watersheds that are comparable to study sites (e.g., Mid-Atlantic Piedmont). These data show how baseflow could have varied up to three-fold over the course of a year, which is near the level of variation observed at some of the study sites between aerial and field surveys (Table 1.3.3). The discharge at a site can be highly variable between late winter and early spring (when aerial surveys were completed) and summer (when most of the field surveys were completed). However, even sites with field surveys completed near the same time of the year as aerial surveys showed variable discharges (e.g., 253 and WL; Table 1.3.3 and Table 2.3.8).

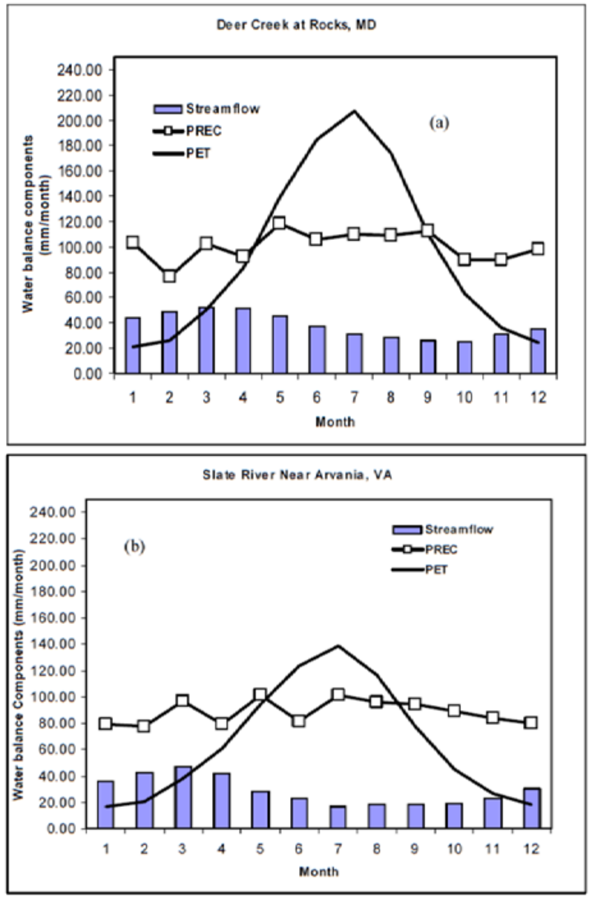


Figure 14. Monthly Water Balance Components of Two Piedmont Watersheds: (a) Deer Creek Watershed and (b) Slate Watershed.

Figure 2.4.1. Conceptual variation in the water balance, as represented by precipitation (PREC), potential evapotranspiration (PET) and streamflow, across a given year at two Piedmont watersheds (reprinted from Mohammad 2004).

Differences in discharge between surveys were due to logistical constraints, but future efforts would benefit from coordinating the timing of field and aerial surveys. Aerial surveys were completed opportunistically, coordinating with weather conditions and equipment use. EMAP / SWAMP field surveys were completed in the summer and prior to aerial surveys because the data was collected as a part of the partner study in PA (Kroll, *unpublished data*). MBSS field surveys were performed after aerial surveys with the thought that delineating habitat metrics from remotely sensed data prior to field delineation might reduce bias. However, this became an unrealistic goal given the extent of data processing necessary, so field surveys were completed as time allowed. In future assessments of this kind, it would be prudent to complete field surveys on the same day or under equivalent discharge conditions as aerial surveys.

Correction of depth and velocity values within MBSS segments proved helpful in reducing the influence of refraction and discharge differences at the level of the sampling unit. Refraction correction would likely be a fruitful endeavor to improve comparisons at all sites, given SfM depths were low and velocities were high relative to field depths in many comparisons, particularly at the sampling unit level. Refraction correction would increase depths, which would necessarily decrease cross-sectional flow velocity estimates (given methods described in section 1.2.4.f). Relative to the direction of bias in SfM depths and velocities relative to field values (negative and positive biases, respectively), such changes to SfM data have the potential to improve comparisons. Discharge correction would less clearly improve comparisons at all sites given the variability in the ratios between field and aerial survey discharges. Changes in SfM data due to discharge correction would not necessarily match the directional biases present between SfM and field data at all sites. However, the combination of refraction and discharge correction at MBSS sites improved velocity comparisons. Therefore, there may be some utility in discharge corrections, and further experimentation with data correction across sites may be helpful in further improving comparisons.

Field biases likely also contributed to variation between field and SfM values, particularly for the categories subjectively assigned during the MBSS field surveys. Field categorical estimates are often subjective and imprecise (e.g., Barbour et al. 1999; Kauffman et al. 1999). Observer bias in the field may have overestimated the influence of a pool within a segment, leading to a “2” or “3” depth category assignment (0.5-1.0m and >1.0m, respectively). Yet if all points within the segment were actually quantified, as in SfM data, the average may actually be <0.5m (a “1” depth category assignment). Categorization of flow across segments could have also been subject to observer bias and subjectivity. For instance, glides and pools within a given segment may have led to a “1” flow category ($\leq 0.3 ms^{-1}$), underestimating the influence of flow velocity within glides and riffles that, when the average or median velocity is summarized across the segment, does in fact lead to a “2” flow category ($>0.3 ms^{-1}$). Differing results in field-SfM comparison analyses when SfM data extraction was adjusted spatially revealed the potential biases present in field data when sampling units are subjectively placed and oriented. Such biases would go unnoticed without the continuous data provided by SfM.

Agreement at the reach-aggregated level suggests the equivalent utility of SfM and field data to researchers and practitioners, who typically describe reaches at this scale. Most researchers and practitioners are not as concerned with metrics on a sampling unit basis, but are interested in aggregate measures of the reach, typically for comparison with aggregate measures at other reaches (e.g., Barbour et al. 1999; Stranko et al. 2015; Doll et al. 2016). Such comparisons include multivariate habitat scoring techniques (e.g., Kauffman et al. 1999; Stranko et al. 2015) and multivariate analyses used to explain biotic condition (Kauffman et al. 1999; Doll et al. 2016). SfM data has a similar utility. Each individual pixel in a digital elevation model (DEM) or orthomosaic image generated by SfM data may not perfectly represent field conditions. However, there are a substantial number of pixel estimates continuously throughout the study area, which, on aggregate, should well represent the system as a whole and the variation within it. Such characteristics of SfM data were evident in the current study, as there was substantial agreement in reach-aggregated metrics even without agreement in sampling unit comparisons, and even in cases of three-fold differences in discharge between field and SfM surveys. Once again, the benefit of SfM data lies in the flexibility to scrutinize the data relative to extraction methods and necessary corrections. Therefore, it may in fact be more useful than field data toward researcher and practitioner goals when summarizing habitat.

If physical habitat metrics from SfM data can be seen as equivalent to those from field data when sampled similarly, SfM data can be used to not only ask questions typically posed using field data, but also to explore how answers might change when continuous data beyond the extent of conventional field reaches are available. Measuring habitat continuously across broad extents with high precision opens a new theoretical avenue, allowing for new questions about how to measure, summarize, and conceptualize physical habitat (e.g., Chapter 3).

Chapter 3 Viewing variation in restored channel habitat through the lens of SfM data

3.1 Introduction

Ecological restoration is "the process of assisting the recovery of an ecosystem that has been degraded, damaged, or destroyed," (Palmer et al. 2006). However, stream restoration can mean many things from place to place, with numerous and overlapping goals and strategies (Wohl et al. 2005). There remain many questions within the scientific community about how restoration practices affect aquatic communities (Palmer et al. 2006; Kroll et al. 2019). For instance, experts in ecological restoration have touted the importance of physical heterogeneity on multiple spatial scales to improving aquatic ecosystem outcomes (Larkin et al. 2006). Only some spatial scales (e.g., $10^{-1} – 10^{2}$m) are captured by conventional habitat assessment techniques, which may mean important broader-scale (e.g., $\geq 10^{3}$m) relationships between heterogeneity and ecosystem outcomes are not captured (e.g., Fausch et al. 2002). Greater specificity in physical habitat conditions over broader and more continuous extents provides perspectives of physical habitat variability that could enhance understanding of restoration outcomes in aquatic ecosystems.

Stream restoration is used to address a variety of management goals, and many practices fit under its umbrella. Common goals of stream restoration include property protection, management of the effects of the local hydrologic regime, nutrient and/or sediment retention, and aquatic ecosystem improvement (Kroll, *unpublished data*; Bernhardt et al. 2005; Kroll et al. 2016). Restoration techniques include manipulations within and/or adjacent to stream channels. Off-stream practices commonly applied in the MidAtlantic region of the US include riparian buffer creation or enhancement, wetland creation or enhancement, stormwater best management practices (BMPs) and agricultural BMPs (Kroll, *unpublished data*; Bernhardt et al. 2005; Kroll et al. 2016). Instream practices commonly applied in the region include placing bank and/or channel structures, regrading channel beds and/or banks, and designing channels to a geomorphic type ("natural channel design"; Kroll, *unpublished data*; Bernhardt et al. 2005; Kroll et al. 2016). Many sites include a combination of multiple practices (e.g., Ch. 1.2; Kroll, *unpublished data*; Kroll et al. 2016). However, certain goals and practices inherently conflict. For instance, restoring "natural" features or processes may be antithetical to property protection. Restoring geomorphic and hydrologic processes may include creating channel units (pool-riffle sequences), encouraging overbank flow (floodplain connection via bank and/or channel grading), or increasing sinuosity. Yet pairing any of these practices with bank or channel structures disallows other natural geomorphic processes, such as channel migration and sediment movement. Similarly, creating idealized habitat conditions without improving degraded hydrologic and/or water quality conditions will not improve ecosystems (Karr and Chu 1999; Hilderbrand et al. 2005; Bernhardt and Palmer 2011; Anim et al. 2018).

Under the National River Restoration Synthesis Study (NRRSS), Bernhardt et al. (2005) identified common objectives of restoration projects. Figure 3.1.1 is reprinted from this study, which compared the top five objectives of stream and river restoration between U.S. regions where major restoration implementation had taken place. The green in this figure represents instream habitat improvement, and the box shows that the Chesapeake Bay region (CB), encompassing Virginia, Maryland and Pennsylvania, had the largest proportion of projects with instream habitat improvement as a major objective.

Fig. A1. Regional differences in the distribution of types of restoration efforts. To facilitate visual comparison only the top five intent categories for each node are shown in each stacked column. All other "non-dominant" intents are summed as part of the "non-dominant" category.

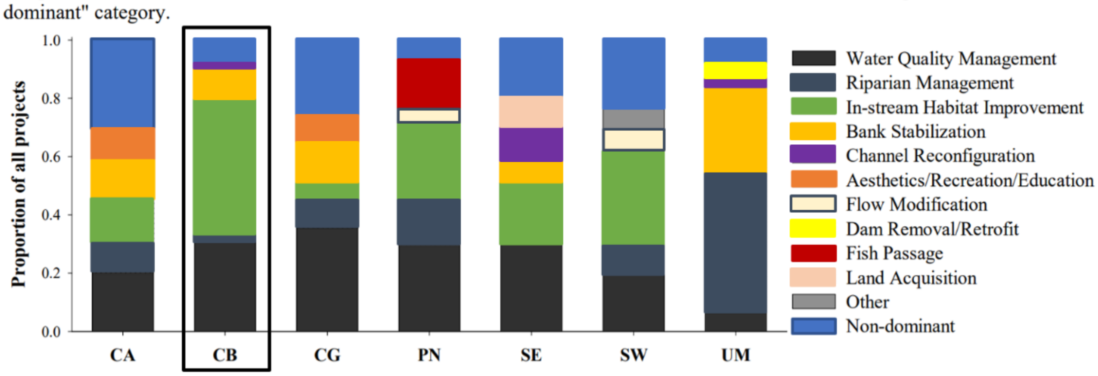


Figure 3.1.1. Common restoration goals in 7 U.S. regions with major restoration implementation. (Reprinted and modified from Bernhardt et al. 2005, Figure A1).

Criticisms of stream restoration goals and practices abound in the scientific literature (e.g., Hilderbrand et al. 2005; Lake et al. 2007; Simon et al. 2007; Palmer et al. 2010; Bernhardt and Palmer 2011). Hilderbrand et al. (2005) outline “myths” in restoration by questioning the assumptions of our goals and practices. Attempting to coerce a system into a single, unchanging state reveals a lack of appreciation for the dynamic nature of stream systems, which can ultimately lead to reduced resilience within the system (Holling 1973; Holling and Meffe 1996; Hilderbrand et al. 2005; Lake et al. 2007). The “natural channel design” method attempts to design streams relative to the way they “should” look based on a geomorphic classification system (Rosgen 1994, 1996; Simon et al. 2007; Lave 2009; Rosgen 2011). It has been an attractive, straight-forward and seemingly science-based solution to very complex problems (Lave 2009). Many researchers have cautioned against the use of channel classification schemes in restoration with the argument that they emphasize channel forms over channel processes (e.g., Poole et al. 1997; Ralph and Poole 2003; Wohl et al. 2005; Kondolf et al. 2006; Simon et al. 2007). Scientists have acknowledged the need to provide realistic alternatives to the complex issues facing streams rather than simply criticize existing efforts (e.g., Wheaton et al. 2004; Lave

2009; Rosgen 2011). In recent decades, scientists have worked to improve the collective knowledge about what kinds of restoration processes (e.g., goal development) and strategies (e.g., process-based and adaptive restoration) would improve outcomes (e.g., Palmer et al. 1997, 2005, 2006; Ralph and Poole 2003; Hilderbrand et al. 2005; Wohl et al. 2005; Kondolf et al. 2006; Larkin et al. 2006; Lake et al. 2007; Swan and Brown 2017).

Understanding restoration outcomes in aquatic ecosystems continues to be a difficult task (e.g., Palmer et al. 2005, 2006; Kroll et al. 2019). Ecological outcomes typically involve relating the magnitude of response in biotic conditions (or lack thereof) to changes in physical habitat. Comparing restored conditions to surrounding unrestored extents is helpful in assessing impacts of physical modification in restoration designs, as other catchment conditions (land use, hydrologic regime, water quality) are held approximately constant between extents (Kroll et al. 2019). However, linking changes in physical habitat from restoration and associated biotic responses has remained elusive (Larkin et al. 2006; Kroll and Oakland 2019; Kroll et al. 2019). If the objective of most restoration projects in the Chesapeake Bay region includes stream habitat improvement (Fig. 3.1.1), it is vital that scientists adequately monitor stream habitat to properly understand restoration outcomes.

Still, understanding restoration outcomes in aquatic ecosystems has been a difficult task within the scientific community (e.g., Palmer et al. 2005, 2006; Kroll et al. 2019). Understanding ecological outcomes typically involves relating the scale of response in biotic conditions (or lack thereof) to the scale of change in physical habitat conditions. Comparing restored conditions to surrounding conditions in unrestored extents is a helpful tool in assessing the impact of physical habitat modification in restoration designs, as other catchment conditions (land use, hydrologic regime, water quality) will be held approximately constant between extents (Kroll et al. 2019).

Yet there has been limited success finding direct links between changes in physical habitat from restoration and associated changes in biotic conditions (Larkin et al. 2006; Kroll and Oakland 2019; Kroll et al. 2019). If the objective of most restoration projects in the Chesapeake Bay region includes stream habitat improvement (Fig. 3.1.1), it is vital that scientists adequately monitor stream habitat to properly understand restoration outcomes.

One important question about how physical habitat restoration affects aquatic ecosystem outcomes involves the role of heterogeneity (Larkin et al. 2006). Physical heterogeneity has been emphasized within the ecological restoration community given its associations with well-founded ecological theories, such as niche availability, species coexistence, and biotic diversity (Larkin et al. 2006; Lake et al. 2007). For instance, foundational literature in ecology established the relationship between environmental heterogeneity and biotic diversity (Levene 1953; Vellend 2016) and community stability (Levin 1976; Larkin et al. 2006). Positive relationships between physical heterogeneity and stream ecosystems have been empirically related to refugia access (e.g., Palmer et al. 2000; Brown 2003; Larkin et al. 2006), near-bed hydraulic complexity important to spawning habitat (e.g., Cardenas et al. 2004; Wheaton et al. 2004; Mason et al. 2012), the provision of different resources for different life stages (e.g., Larkin et al. 2006; Yarnell 2008), ecosystem functioning (e.g., Cardinale et al. 2002; Palmer et al. 2005; Larkin et al. 2006), and physical and biotic resilience (e.g., Palmer et al. 2005; Larkin et al. 2006). Ecological benefits of physical heterogeneity have been observed at multiple spatial scales (e.g., Poff and Ward 1990; Power 1992; Palmer et al. 2000, 2010; Cardinale et al. 2002; Brown 2003; Wheaton et al. 2004; Yarnell 2008). However, there are still many questions about whether restoration practices that alter physical heterogeneity actually improve ecological conditions (Palmer et al. 2000; Wheaton et al. 2004; Larkin et al. 2006; Lake et al. 2007; Vellend 2016).

Restoration ecologists have questioned the spatial scales at which heterogeneity is important to biodiversity and ecosystem function (Palmer et al. 2000; Larkin et al. 2006), and whether site conditions necessarily support heterogeneous habitat (Larkin et al. 2006). Channel manipulation that recreates standard geomorphic forms from a design template would seem to introduce specific repetitive patterns into stream channels. The ability to quantify and scrutinize heterogeneity across multiple spatial scales has been limited by the extent, specificity and continuity of conventional field methods.

Measuring habitat continuously across broad extents with high precision opens a new theoretical avenue, allowing for new questions about the spatial variation in physical habitat across spatial scales. In a critical review of river restoration, Wohl et al. (2005) cite a lack of monitoring at scales that match management efforts as a major limitation to advancing knowledge about restoration effectiveness. Fausch et al. (2002) made a similar assertion that monitoring should be completed on scales that match critical biotic processes and management efforts. Use of continuous and more precise data over broad extents has allowed researchers to question perceptions of habitat developed through more conventional, field-based techniques. For instance, theoretical relationships in downstream hydraulic geometry were established from reach-scale field data (Leopold and Maddock Jr. 1953). These assumptions were put into question by researchers using continuous, remotely-sensed data to assess and predict downstream channel depth (Marcus and Fonstad 2008; Marcus et al. 2012) and width (Carbonneau et al. 2012). Their findings implied that downstream hydraulic geometry relationships did not strictly hold when the channel was viewed continuously over broad extents rather than viewed through the lens of conventional, discontinuous field stations. Such findings have major implications not only for geomorphic theory, but also for science-based river management (Marcus et al. 2012).

Similarly, understanding of restoration outcomes may be limited by conventional sampling of field reaches that is discontinuous and limited in extent. Manipulation of physical habitat during restoration, and associated ecological outcomes, may be perceived in novels ways when viewed over broader, more continuous extents.

Structure-from-motion (SfM) estimates of physical channel habitat provide an extensive and continuous view of stream channels. Mapped estimates were previously shown to be realistic (chapter 1) and representative (chapter 2) of restored streams in Maryland and Pennsylvania. Therefore, SfM data can be used to ask questions about restored habitat that might typically be asked with field estimates, but with greater continuity over broader extents. A seamless and extensive view of the stream channel allows physical habitat heterogeneity to be assessed at a variety of scales, providing new understanding of restored habitat. Accordingly, this chapter seeks to examine the consequences of restoration for physical habitat heterogeneity using continuous, high-resolution and extensive data. Given that restoration of habitat patterns often occurs as downstream sequences of channel units, the analysis will focus on longitudinal variation. Longitudinal sampling can show how habitat might vary within and across channel units, reaches and project extents. These analyses are predicted to reveal reduced variability in physical habitat features within restored channel extents when compared to unrestored channel extents on all scales. Ultimately, these analyses could help to test the assumption that environmental heterogeneity is an important component of ecological restoration, and if so, on what scales.

3.2 Methods

3.2.1 Data preparation

SfM data from restored and unrestored reaches were used to compare continuous characterizations of habitat. Derivation of habitat metrics used in these analyses follow the same process as described in Chapter 1. Restored and unrestored (upstream and/or downstream) extents were delineated using maps from project managers or practitioners and publicly available project reports (e.g., Figure 3.2.1; East Goshen Township n.d.; Clauser and Struble 2008; Bauer et al. 2009; Montgomery County DEP 2009, 2019b, 2019a; Birmingham and Koser 2015). SfM data from restored and unrestored reaches were used to compare continuous characterizations of habitat. Derivation of habitat metrics used in these analyses follow the same process as described in Chapter 1. Restored and unrestored (upstream and/or downstream) extents were delineated using maps from project managers or practitioners and publicly available project reports (e.g., Figure 3.2.1 (East Goshen Township n.d.; Clauser and Struble 2008; Bauer et al. 2009; Montgomery County DEP 2009, 2019b, 2019a; Birmingham and Koser 2015).

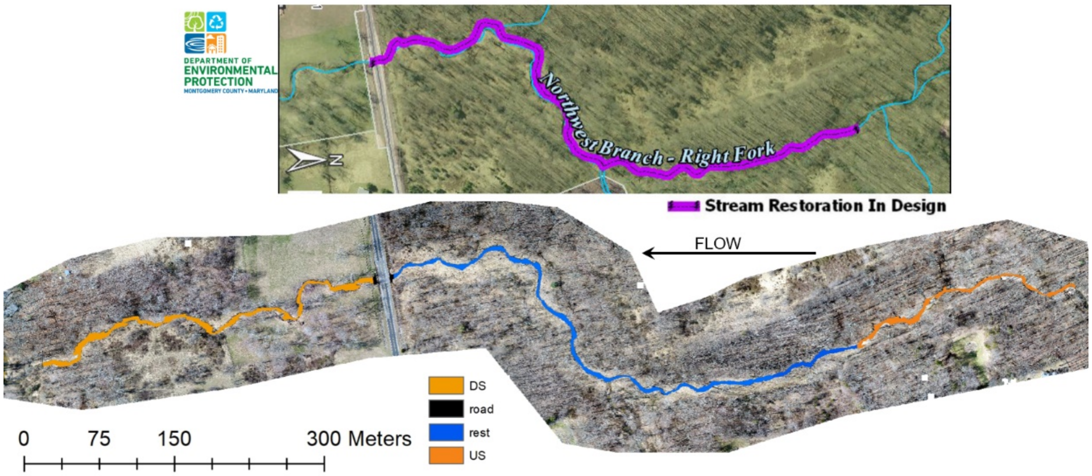


Figure 3.2.1. Example of a delineated restoration extent at site "WL." The upper figure is from the Montgomery County DEP report on the Woodlawn stream restoration (Mont. Co. DEP, 2019), which was used to delineate the restored and unrestored extents within SfM imagery, blue and orange, respectively, in the lower figure.

The analyses of this chapter will focus on longitudinal variation in physical habitat features to assess longitudinal patterns of heterogeneity in the stream channels. Such patterns would be representative of how a habitat varies across important scales, such as the channel unit, reach and restoration project. To focus sampling longitudinally, cross-sectional means of depth, width and velocity were the data of focus (Fig. 3.2.2). Hereafter, cross-sectional means of the metrics are referred to just as depth, width or velocity. Extreme outliers of cross-section mean depth, width and velocity were identified per site as those that were less than the first quantile minus three times the interquartile range (IQR) or were greater than the third quartile plus three times the IQR. These identified values represented <1% of estimates per site and were removed prior to further analyses. Upstream and downstream extents were collectively considered the "unrestored" extent (orange in Fig. 3.2.2).

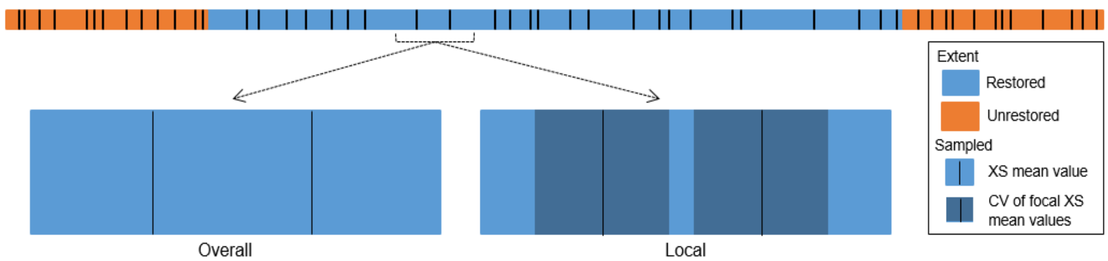


Figure 3.2.2. Depiction of the cross-section resampling regimes, with 25 cross-sections randomly sampled per extent (restored and unrestored). For the project-scale difference in variation, cross-section (XS) mean values of each metric were sampled (left), while for the local difference in variation, the coefficient of variation (CV) values of the focal XS mean values were sampled (right). In each case, the entire random sample was repeated 100 times.

3.2.2 Restoration project-scale variation

Habitat variables were sampled at the scale of the entire restoration project to assess whether there were differences in overall physical habitat magnitude and variation between restored and unrestored extents. Twenty-five randomly selected cross-sections (without replacement) were sampled from each of the restored and unrestored extents per site (Fig. 3.2.2). Unrestored channel extents were smaller than restored extents at each site, so the number of randomly selected cross-sections was chosen to represent only about 1/3 of the smallest unrestored extent. Each cross-section represents approximately 1m of the stream channel longitudinally, so 25 cross-sections were selected to ensure enough of the extents were sampled to approximately represent the variation, but not so much of the channel that entire extents would be represented by any samples.

To understand differences in magnitude of width, depth and velocity, sample distributions from restored and unrestored extents were displayed separately in a density ridge plot, and compared with a Kolmogorov-Smirnoff (K-S) test. Such a test indicates whether overall habitat variable magnitudes are different between extents, which represented how habitat metrics are often summarized and compared. Significance in the K-S test indicates that the rank ordered distributions are significantly different. Because any one randomly selected sample may fail to capture all the variation inherent to a reach, random sampling and successive analysis were repeated 100 times per site. The proportion of those 100 resamples with insignificant K-S test results ($p>0.05$) was displayed with the distributions. The collective sampled distributions were considered qualitatively different if this proportion was <0.3 (proportion significant >0.7). This criterion was chosen subjectively, as a qualitative judgment that the extents were fairly different if more than 70% of samples were significantly different.

Coefficients of variation (CVs) in habitat variables were also compared between extents to understand whether variability (heterogeneity) differed at the scale of the restoration project. For each of the 100 sample distributions, the CV was calculated per extent. The distribution of all CVs calculated per extent across all samples (100) were displayed as density plots and compared between restored and unrestored extents using a Student's t-test. A p-value less than the alpha (0.05) indicated the range of variability (CV values) between extents was significantly different, which would indicate heterogeneity was significantly different at the scale of the restoration project. The mean of the 100 CV values per distribution and the t-test p-value were also displayed on the plot.

3.2.3 Local variation

As discussed in chapter 2, smaller scales of stream habitat ($10^{-1} – 10^{2}$ m) are typically studied in the field. These scales are explicitly targeted to understand aquatic habitat in conjunction with aquatic biota sampled in the field (e.g., macroinvertebrates and fish). Therefore, local variation in habitat was assessed to understand the kind of heterogeneity that might be experienced by aquatic biota within a channel unit (or between adjacent channel units; e.g., 10^{1} m) or reach (10^{2}

m). These two scales were approximately representative of length scales of physical features in the study streams. For instance, 20 meters was the approximate average length of a pool or riffle unit (e.g., Fig. 1.3.5 and Fig. 1.3.6), so the local variation represented the variation within a pool or riffle, or the variation associated with the transition between the two channel units. The channel unit is also representative of the scale at which local habitat patches may be distributed and selected by individual organisms (e.g., contains microhabitats; Fausch et al. 2002). The 100-meter length scale represented that of the field reaches, which were designed to capture approximate scales of geomorphic processes that create habitat and at which humans typically assess habitat in the field (e.g., Ch. 2.1; Leopold and Wolman 1957; Karr et al. 1986; Kauffman et al. 1999; Fetscher et al. 2009).

To sample local variation on these length scales, the coefficient of variation of the distribution of values within 20m and 100m of a given cross section was calculated (Fig. 3.2.2). That meant cross-sections within 10m and 50m, respectively, of the focal cross-section were included in the calculation of the focal cross-section's "local CV" value. To remove edge effects and the influence of the boundary between restored and unrestored extents, null values were added to the ends of each extent (DS, restored and upstream). After these calculations, null values were removed from the ends of all extents, and unrestored reaches both upstream and downstream of the restoration were combined per site (orange in Fig. 3.2.2).

Consistent with the overall sampling above, 25 cross sections were randomly sampled to assess local variation from the restored and unrestored extents without replacement (Fig. 3.2.2). The resampled distributions of local CV values from each extent were extracted as vectors, the mean local CV values calculated per extent distribution. The two extent (restored and unrestored) distributions were compared using a Student's t-test. Again, this sampling regime and successive analysis was repeated 100 times. Each sampled distribution of local CV values was displayed as a density ridge plot. The global mean local CV across the sample distributions of mean local CV values was calculated per extent. The proportion of the 100 samples with insignificant t-test results ($p > 0.05$) was also calculated. The collective sampled distributions were considered significantly different if this proportion was <0.3 (proportion significant >0.7; assigned subjectively).

3.2.4 Validity of comparisons

A potential confounding factor in assessing differences in variation between extents was the difference in the view of the channel from the air between extents. The aerial view of the channel was typically less impeded within restored extents, as trees were often cleared around the channel during the restoration project (e.g., Fig. 1.3.10). Such systematic differences could potentially confound results, particularly of depth and velocity, if the vegetation surrounding the channel artificially created more variability in the unrestored extent of the digital elevation model (DEM; as opposed to more variability in channel elevation being a real feature). During the dominant sediment categorization, "exposed wood" was used to delineate portions of the channel that were obscured from view by overhanging trees or other vegetation (see Chapter 1.2). In most cases, a tree branch was blocking a small portion of the channel, but the edges of sediment classes within the channel could still be seen surrounding it, so the exposed wood category overlapped other sediment categories (e.g., the proportions of all categories, when exposed wood was included, summed to >1). As a coarse assessment of whether vegetation obscured more of the channel within unrestored extents, and therefore possibly confounded the variation within the DEM, the proportions of exposed wood per reach were compared between extents per site. The proportion of exposed wood within unrestored extents at Maryland sites were averaged between the upstream and downstream reaches. The pattern of difference per site in these proportions was compared to the pattern of results seen in the CV comparisons. Lastly, the proportion of exposed wood per reach was compared to the CV values of cross-section mean depth, width and velocity within the same reach via a Pearson correlation to assess whether channel view obstruction within the orthophoto and DEM may be significantly correlated with greater variation in physical channel features.

3.3 Results

The following displays results for physical habitat metric comparisons at two sites that represent contrasting results of the analyses. The same two sites, ULP2 and s144, are displayed throughout the section to compare alternative responses of different metrics within the same site and extent. However, these results are available for all site-metric combinations, which are summarized in Table 3.3.1.

3.3.1 Restoration project-scale variation

3.3.1.a Depth

None of the sites showed a significant difference between restored and unrestored distributions of randomly sampled depths at the project scale (Table 3.3.1; Fig. 3.3.7). However, for the same comparisons, 6 of 7 sites showed significantly different mean depth CV values between restored and unrestored extents (Table 3.3.1; Fig. 3.3.7). Of these, 5 sites showed higher mean depth CV values within unrestored extents, while 1 site showed higher mean depth CV values within restored extents (Table 3.3.1).

Figure 3.3.1 displays the comparison between project-scale depth variation within restored and unrestored extents at two sites. ULP2 displayed more spread in its distributions of depth within the unrestored extent than in its distributions of restored extent depths (Fig. 3.3.1, ULP2 a). Project-scale restored depth distributions from ULP2 also showed higher kurtosis, but the restored and unrestored resampled distributions were only significantly different for 9% of the resampled distribution comparisons (Fig. 3.3.1, ULP2 a). However, the unrestored extent did show a significantly ($p<0.05$) higher mean depth CV (Fig. 3.3.1, ULP2 b). Alternatively, S144 displayed more similarly shaped project-scale depth distributions between restored and unrestored extents, and the distributions were significantly different for fewer sub-sample comparisons (5%; Fig. 3.3.1, S144 a). The mean depth CV of the extent distributions at S144 was not significantly different between extents at the project scale (Fig. 3.3.1, S144 b).

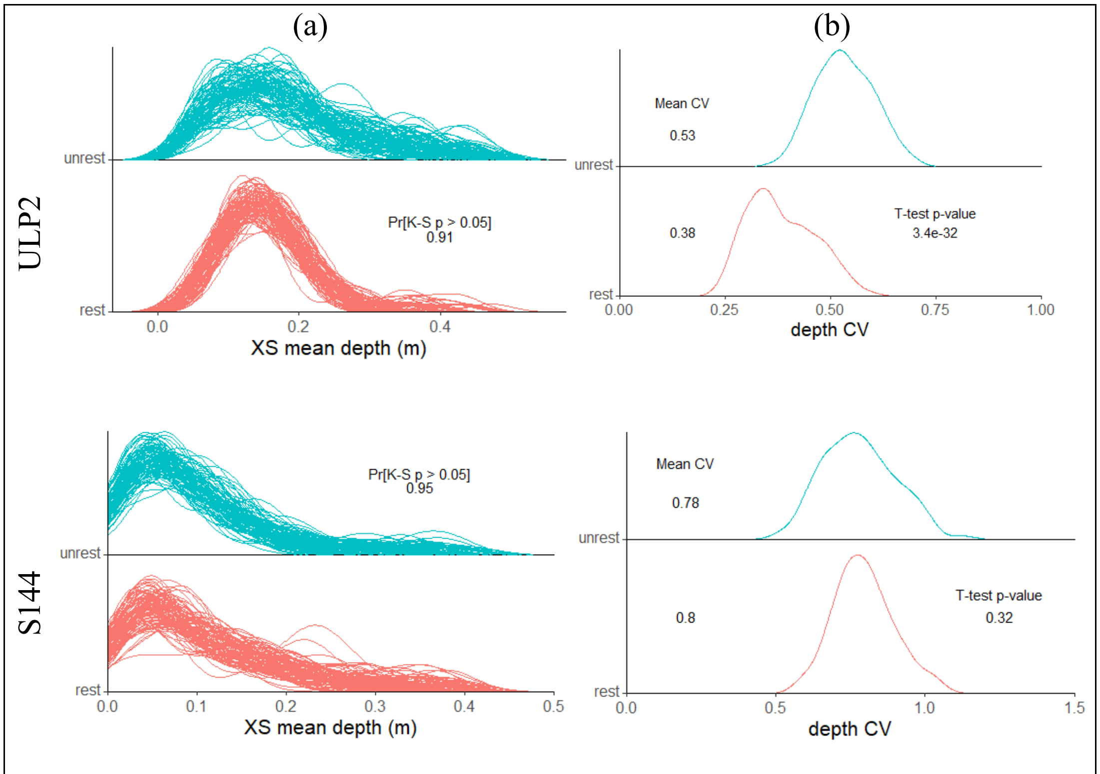


Figure 3.3.1. Randomly sampled cross-section depths per extent ("unrest" is unrestored; "rest" is restored). Rows represent the different sites displayed, ULP2 and S144. (a) Distributions of randomly sampled cross-section mean depths per extent with proportions of insignificant K-S tests. (b) The distributions of CVs from each resample with the mean CV per distribution and t-test p-value are displayed on the plot.

3.3.1.b Width

Project-scale variation was compared between extents for each sampled width distribution. Two sites showed a significant difference between restored and unrestored distributions of randomly sampled widths (Table 3.3.1; Fig. 3.3.7). However, for the same comparisons, 6 of 7 sites showed significantly different mean width CV values between extents at the scale of the restoration project (Table 3.3.1; Fig. 3.3.7). Of these, 4 sites had higher mean width CV values within unrestored extents, while two sites showed higher mean width CV values within restored extents (Table 3.3.1; Fig. 3.3.7).

Project-scale width distributions are displayed for ULP2 and S144 in Figure 3.3.2. Both sites displayed more spread in distributions of width within unrestored extents than within restored extents, and the restored extents showed higher kurtosis, but neither were significantly different for >70% of the resamples (Fig. 3.3.2.a). Alternatively, when looking to variation within width distributions at the scale of restoration projects, the unrestored extent of ULP2 did show a significantly ($p<0.05$) higher mean width CV (Fig. 3.3.2, ULP2 b). However, the mean width CV was not significantly different between extents at S144 (Fig. 3.3.2, S144 b). Restored and unrestored distributions of project-scale width CVs overlap in most of their range at S144 (Fig. 3.3.2, S144 b), while those ranges at ULP2 are almost completely distinct (Fig. 3.3.2, ULP2 b).

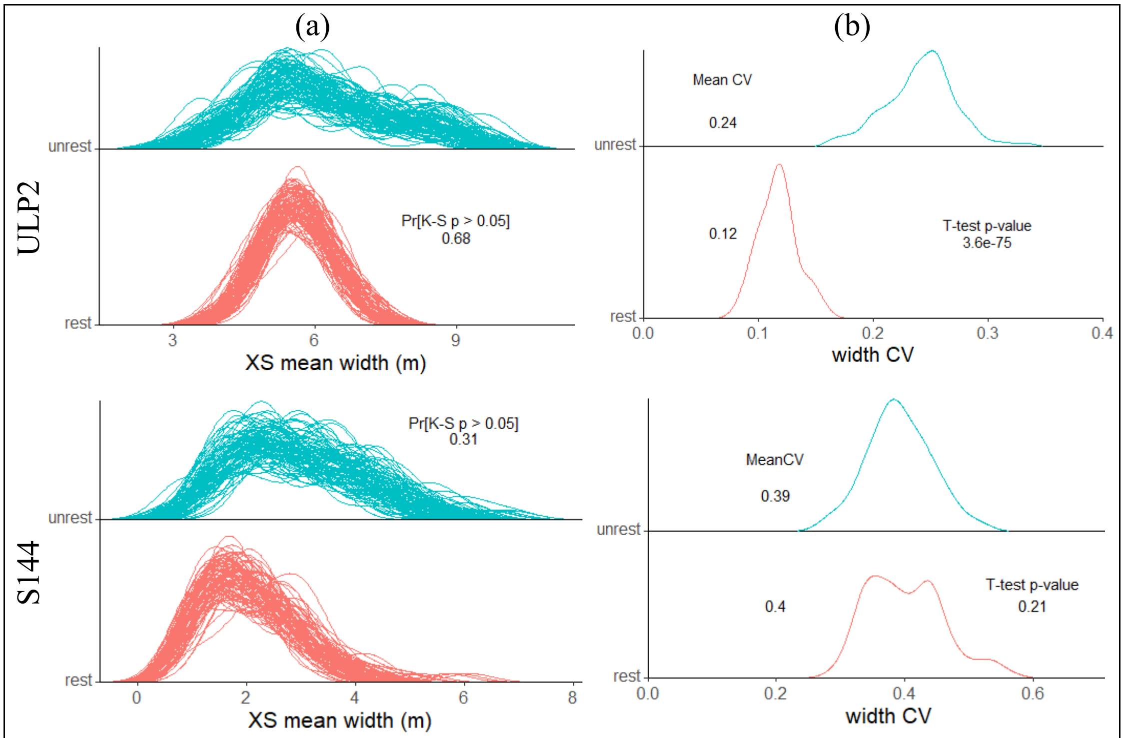


Figure 3.3.2. Randomly sampled cross-section width per extent. Rows represent the different sites displayed, ULP2 and S144. (a) Distributions of randomly sampled cross-section mean widths per extent with proportions of insignificant K-S tests. (b) The distributions of CVs from each resample with the mean CV per distribution and t-test p-value displayed on the plot.

3.3.1.c Velocity

Velocity distributions at the project scale were not significantly different between extents across study sites (Table 3.3.1; Fig. 3.3.7). However, for the same comparisons, showed significantly higher project-scale mean velocity CV values for the within unrestored extents (Table 3.3.1). This contrast between differences in distributions versus differences in variation was evident in Figure 3.3.3, which displays the distribution of restored and unrestored extent velocity samples at two study sites. ULP2 displayed more spread in its distributions of velocity within distributions of the unrestored extent than within distributions of the restored extent, but the restored and unrestored resampled distributions were only significantly different for 14% of the resample comparisons (Fig. 3.3.3, ULP2 a). However, the unrestored extent did show a significantly ($p < 0.05$) higher mean velocity CV at the project scale (Fig. 3.3.3, ULP2 b). Alternatively, S144 displayed more similarly shaped velocity distributions between restored and unrestored extents, and distributions were significantly different for only 11% of the resample comparisons (Fig. 3.3.3, S144 a). Also, the mean velocity CV of the unrestored extent distributions at S144 was not significantly different ($p > 0.05$) than that of the restored extent distributions (Fig. 3.3.3, S144 b).

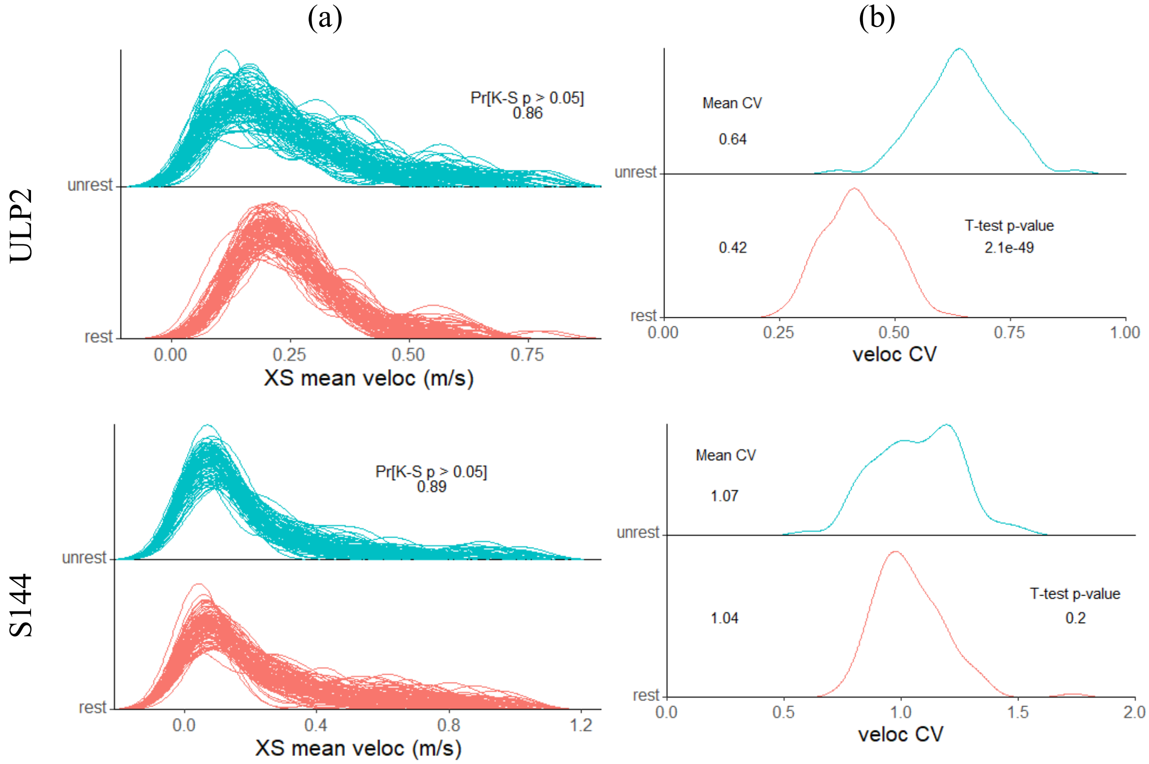


Figure 3.3.3. Randomly sampled cross-section velocity per extent. Rows represent the different sites displayed, ULP2 and S144. (a) Distributions of randomly sampled cross-section mean velocities per extent with proportions of insignificant K-S tests. (b) The distributions of CVs from each resample with the mean CV per distribution and t-test p-value displayed on the plot.

3.3.2 Local variation

3.3.2.a Depth

Local variation in depth was compared between extents at two spatial scales. Five of 7 sites showed significantly different mean 100m depth CV values between restored and unrestored extents (Table 3.3.2; Fig. 3.3.7). Of these, 4 sites showed significantly higher 100m variation in unrestored extents, while 1 site showed significantly higher 100m variation in its restored extent (Table 3.3.2). Alternatively, 3 of 7 sites showed significantly different mean 20m depth CV values between extents, all of which were higher within unrestored extents than within restored extents (Table 3.3.2; Fig. 3.3.7).

Column (a) of Figure 3.3.4 shows the distributions of 20m depth CV values, while column (b) displays the distributions of 100m depth CV values. ULP2 displayed more spread in its distributions of 20m depth CV values within its unrestored extent, and the mean 20m depth CV across samples was higher in the unrestored extent (Fig. 3.3.4, ULP2 a). The lowest 20m focal depth CV values were present within the restored extent of ULP2, with some distributions including 0 values (Fig. 3.3.4, ULP2 a). Alternatively, S144 displayed more similar distributions of 20m depth CV values between extents (Fig. 3.3.4, S144 a). The mean 20m depth CV of the distributions was not significantly different for 69% of the resamples, but the mean sample 20m depth CV value was slightly higher within restored extents (Fig. 3.3.4, S144 a). One-hundred - meter depth variation was again significantly higher in the unrestored extent at ULP2 (Fig. 3.3.4, ULP2 b). The shape of the unrestored and restored distributions of 100m depth CV values were similar to those of 20m focal CVs at ULP2, but with more restricted spread of 100m depth CV values than of 20m depth CV values, and slightly higher modes of the CV distributions (Fig. 3.3.4, ULP2). The same could be said of S144, with similar distribution shapes of extents between 20m focal CV values and 100m focal CV values, and slightly higher modes of the focal CV distributions (Fig. 3.3.4, S144). And similarly to the 20m depth CV distributions, the variation was higher in the restored extent, but the distribution of 100m depth CV values were significantly higher within restored extents for 93% of resample comparisons (Fig. 3.3.4, S144 b).

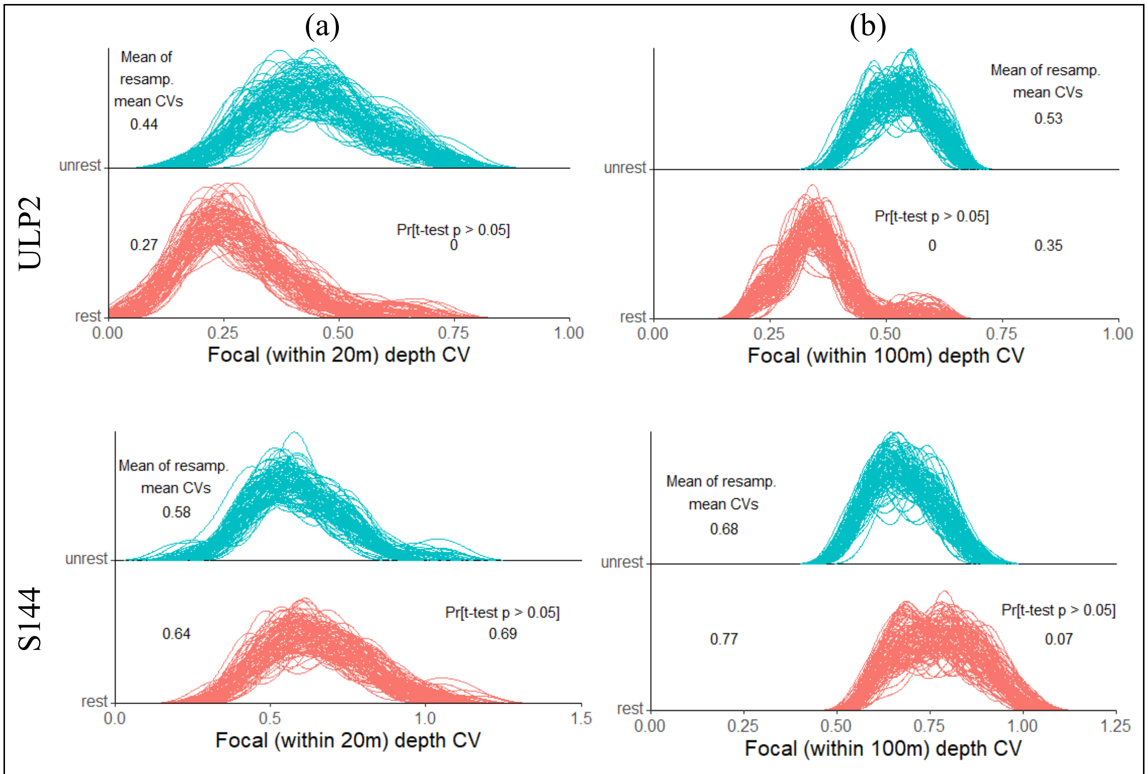


Figure 3.3.4. Distributions (100 trials) of 25 randomly sampled focal cross-section mean depth CVs per extent at ULP2 and S144. The distributions of randomly sampled CV values of cross-section mean depth within 20 meters (a) was compared to those CV values within 100m (b). The global mean across resamples of the distribution's mean focal CV per extent, and the proportion of trials with insignificant t-test results, are displayed on the plots.

3.3.2.b Width

Three of 7 sites showed significant differences between extents in mean 100m width CV values, all of which were higher in unrestored extents (Table 3.3.2; Fig. 3.3.7). Alternatively, only 1 of 7 sites showed significantly different 20m width CV values between extents, which again, was higher within the unrestored extent (Table 3.3.2; Fig. 3.3.7). Local width showed the lowest CV values across sites.

Again, column (a) of Figure 3.3.5 displays distributions of 20m CV values, and column (b) 100m CV values, at two study sites. ULP2 displayed more spread in its distributions of 20m width CV within the unrestored extent, and 93% of resampled unrestored distributions showed significantly higher mean 20m width CV values than those of restored distributions (Fig. 3.3.5, ULP2 a). The lowest 20m width CV values were present within the restored extent of ULP2, with some distributions including 0 values (Fig. 3.3.5, ULP2 a). However, the unrestored extents of both ULP2 and S144 also included 0 values of 20m width CV (Fig. 3.3.5, ULP2 a and S144 a, respectively). In contrast to ULP2, S144 displayed similar distributions of 20m width CV values between extents (Fig. 3.3.5, S144 a). Distributions of 20m width CV values were not significantly different for 83% of samples, but mean 20m width CV was slightly higher within the restored extent (Fig. 3.3.5, S144 a). The mean width CV within 100m of randomly sampled cross sections was again significantly higher in the unrestored extent at ULP2 (Fig. 3.3.5, ULP2 b). The unrestored distributions of 100m width CV values at ULP2 was multimodal, with the major mode at a little over 0.2, but additional minor modes at a little over 0.1 and approximately 0.3, indicating there were three distinct levels of 100m focal variation in width within the unrestored extent at ULP2 (Fig. 3.3.5, ULP2 b). The 100m width CV distributions at S144 were not significantly different for 44% of the resample comparisons, though the mean 100m width CV was slightly higher within the restored extent (Fig. 3.3.5, S144 b). The distributions of 100m width CV for the unrestored extent of S144 had higher spread and included lower CV values than the restored extent (Fig. 3.3.5, S144 b).

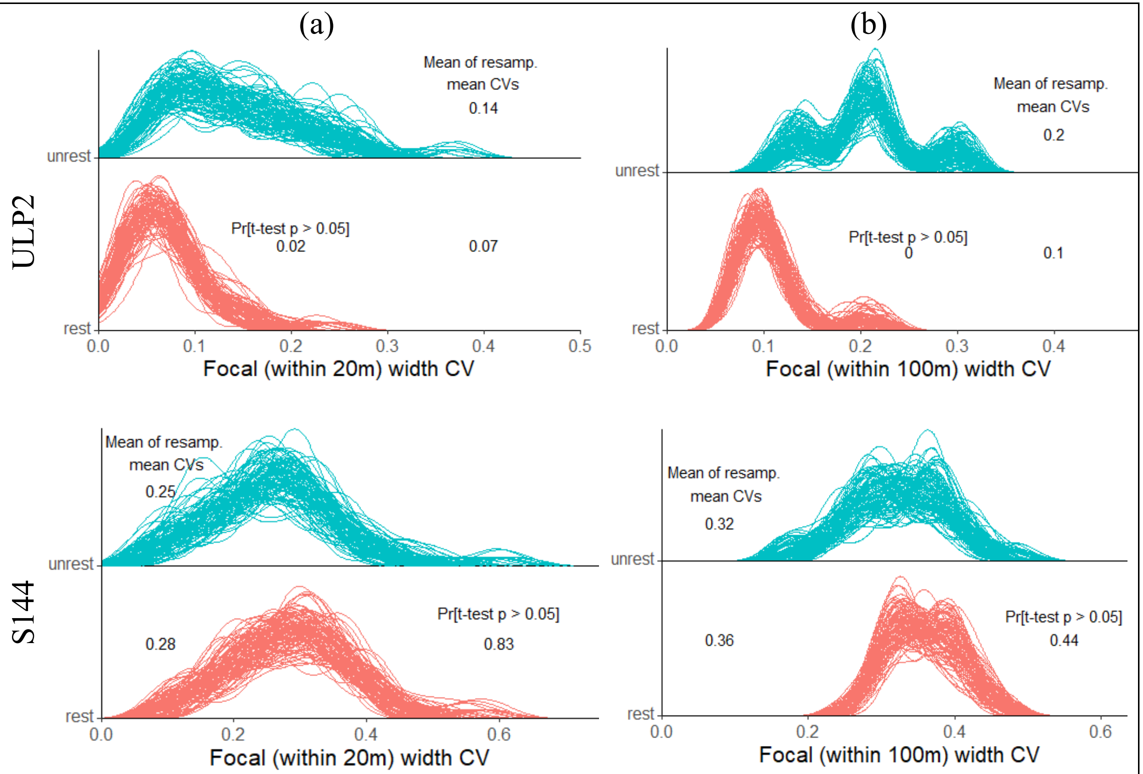


Figure 3.3.5. Distributions (100 trials) of 25 randomly sampled focal cross-section mean width CVs per extent at ULP2 and S144. The distributions of randomly sampled CV values of cross-section mean width within 20 meters (a) was compared to those CV values within 100m (b). The global mean across resamples of the distribution's mean focal CV per extent, and the proportion of trials with insignificant t-test results, are displayed on the plots.

3.3.2.c Velocity

Four of 7 sites showed significant differences between extents in mean 100m velocity CV values, all of which were higher within the unrestored extent (Table 3.3.2; Fig. 3.3.7). Alternatively, 2 of 7 sites displayed significantly higher 20m velocity CV within the unrestored extents (Table 3.3.1; Fig. 3.3.7). Velocity showed the highest CV values per site across scales of comparison.

Distributions of velocity CV values within 20m (Fig. 3.3.6 a) and 100m (Fig. 3.3.6 b) of randomly sampled cross sections are again presented for two sites. ULP2 displayed more spread in 20m velocity CV distributions within the unrestored extent than the restored extent, and all of the resampled unrestored distributions showed significantly higher mean 20m velocity CV (Fig. 3.3.6, ULP2 a). The lowest 20m velocity CV values were present within the restored extent of ULP2, with some distributions including 0 values (Fig. 3.3.6, ULP2 a). However, the unrestored extent of S144 also included 0 values of 20m velocity CV (Fig. 3.3.6, S144 a). In contrast to ULP2, S144 displayed more similar distributions of 20m velocity CV values between extents (Fig. 3.3.6, S144 a). The mean 20m velocity CV values were not significantly different between extents for 74% of samples, but the mean 20m velocity CV was slightly higher within the restored extent (Fig. 3.3.6, S144 a).

For the velocity CV within 100m of randomly sampled cross sections, the mean of all resampled distributions was again significantly higher in the unrestored than in the restored extent at ULP2 (Fig. 3.3.6, ULP2 b). Both the restored and unrestored distributions of 100m velocity CV values at ULP2 were bimodal. The unrestored extent had two major modes of 100m velocity CV values, at approximately 0.5 and 0.75 (Fig. 3.3.6, ULP2 b). The restored extent had a major mode of 100m velocity CV at a little over 0.25, and a minor mode at approximately 0.5 (Fig. 3.3.6, ULP2 b). The mean 100m velocity CV at S144 was not significantly different for 89% of the resample comparisons, but this mean was slightly higher within the restored extent, and displayed another bimodal distribution with a major mode at approximately 1 and a minor mode at approximately 1.25 (Fig. 3.3.6, S144 b). The distributions of 100m velocity CV values for the unrestored extent of S144 had higher spread and included higher CV values than the restored extent (Fig. 3.3.6, S144 b). Overall, both the 20m and 100m focal velocity CV distributions at S144 had higher CV values than for those distributions of ULP2 (Fig. 3.3.6).

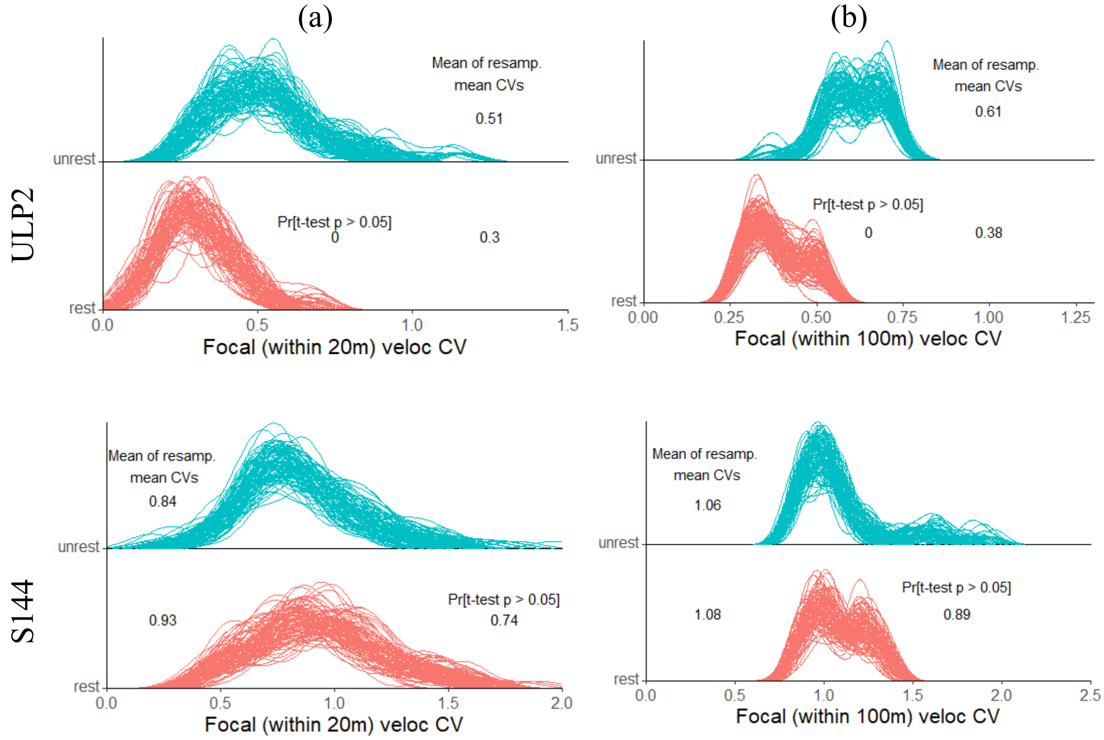


Figure 3.3.6. Distributions (100 trials) of 25 randomly sampled focal cross-section mean velocity focal CVs per extent at ULP2 and S144. The distributions of randomly sampled CV values of cross-section mean velocity within 20 meters (a) was compared to those CV values within 100m (b). The global mean across resamples of the distribution’s mean focal CV per extent, and the proportion of trials with insignificant t-test results, are displayed on the plots.

3.3.3 Comparisons across sites and scales of variation

At the scale of the restoration project, only 2 of 21 comparisons showed significantly different distributions between restored and unrestored extents (K-S; Table 3.3.1). However, considering differences in variation between extents at this scale, 15 of 21 comparisons were significantly different. Of those 15 comparisons, 12 showed significantly greater variation within unrestored extents, while only 3 showed significantly greater variation within restored extents (Table 3.3.1).

Table 3.3.1. Project-scale comparisons across sites. Rank ordered distributions were compared (top), with proportions of insignificant K-S tests bolded where <0.3 of sample t-tests were insignificant. Mean CV values across distributions were compared (bottom), with CV values bolded where <0.3 of sample t-tests were insignificant.

				s144	s253	CRAB	EBCH	PLUM	ULP2	WL
Project scale	Proportion of insignificant (p > 0.05) K-S tests	Depth		0.95	0.82	0.71	0.96	0.95	0.91	0.96
		Width		0.31	0.75	0.47	**0.03**	**0**	0.68	0.96
		Veloc		0.89	0.59	0.95	0.77	0.52	0.86	0.95
	Mean CV	Depth	Unrest	0.78	**0.58**	**0.75**	**0.59**	**0.71**	**0.53**	**0.77**
			Rest	0.8	**0.7**	**0.6**	**0.52**	**0.58**	**0.38**	**0.68**
		Width	Unrest	0.39	**0.31**	**0.29**	**0.26**	**0.36**	**0.24**	**0.4**
			Rest	0.4	**0.35**	**0.32**	**0.22**	**0.26**	**0.12**	**0.27**
		Veloc	Unrest	1.07	0.84	0.79	**0.74**	**1.03**	**0.64**	0.89
			Rest	1.04	0.81	0.75	**0.67**	**0.82**	**0.42**	0.91

Fewer comparisons were significantly different when assessing local variation. When assessing variation within 100m (approximate reach scale in study streams), 11 of 21 comparisons displayed significantly different global mean CV values between extents. Of these, 10 instances indicated greater 100m variation within the unrestored extent, and only 1 instance indicated greater 100m variation within the restored extent. At the approximate channel unit scale in study streams (20m), local variation was significantly different for 6 of 21 comparisons, all of which indicated greater global CV values within unrestored extents (Table 3.3.2).

Table 3.3.2. Local scale CV value comparisons across sites. Bold values indicate substantial differences (<0.3 of t-tests were insignificant).

				s144	s253	CRAB	EBCH	PLUM	ULP2	WL
Local scales	100m mean CV	Depth	Unrest	**0.68**	0.57	**0.73**	**0.55**	**0.69**	**0.53**	0.74
			Rest	**0.77**	0.65	**0.57**	**0.42**	**0.56**	**0.35**	0.66
		Width	Unrest	0.32	0.28	0.27	0.19	**0.32**	**0.2**	**0.32**
			Rest	0.36	0.3	0.25	0.45	**0.24**	**0.1**	**0.26**
		Veloc	Unrest	1.06	0.84	**0.8**	**0.72**	**0.95**	**0.61**	0.85
			Rest	1.08	0.93	**0.61**	**0.57**	**0.8**	**0.38**	0.88
	20m mean CV	Depth	Unrest	0.58	0.51	0.59	**0.47**	**0.6**	**0.44**	0.62
			Rest	0.64	0.51	0.49	**0.35**	**0.48**	**0.27**	0.54
		Width	Unrest	0.25	0.22	0.2	0.13	0.22	**0.14**	0.23
			Rest	0.28	0.21	0.19	0.12	0.2	**0.07**	0.21
		Veloc	Unrest	0.84	0.66	**0.69**	0.52	0.79	**0.51**	0.73
			Rest	0.93	0.65	**0.55**	0.43	0.69	**0.3**	0.74

The proportion of site extent comparisons with significant differences across scales is displayed in Figure 3.3.7. Color in the figure represents significance and directionality of the significance. It was evident that most of the comparisons showed no significant difference, but a large portion of comparisons showed lower variation within the restored extent than within the unrestored extent.

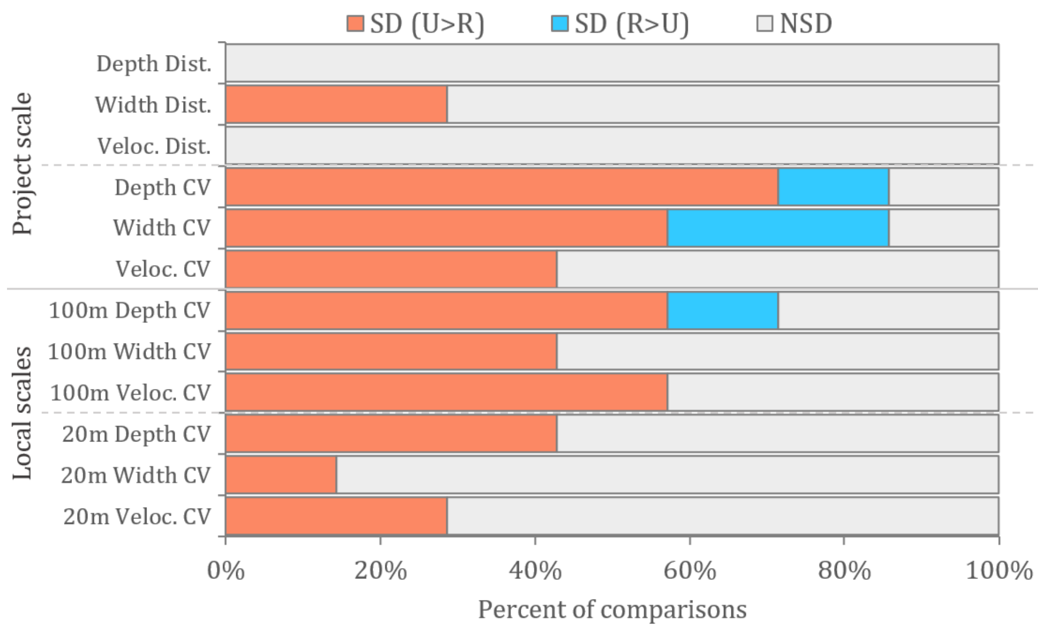


Figure 3.3.7. Substantial differences (<0.3 insignificant comparisons across 100 samples) between restored and unrestored extents across comparisons and sites. The first three bars represent extent magnitude distributions ("Dist."), whereas the remainder represent differences in CV values among extents.

3.3.4 Validity of comparisons

There was not a clear pattern of greater proportions of exposed wood within unrestored extents (Fig. 3.3.8). While some sites did show much greater proportions of exposed wood within the unrestored reach(es) (e.g., s144), others showed much greater proportions within the restored reach (e.g., CRAB; Fig. 3.3.8). Site EBCH showed the greater overall proportions of exposed wood within both the restored and unrestored reaches (Fig. 3.3.8), but did not display the greatest variation in metrics (Table 3.3.1).

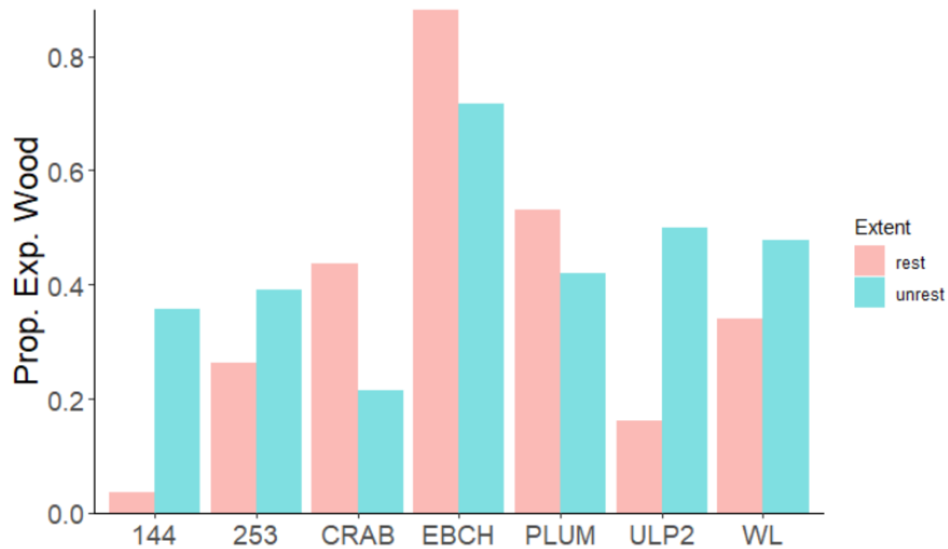


Figure 3.3.8. Comparison of proportion exposed wood within the restored and unrestored extents.

The pattern of difference between proportions of exposed wood per reach across sites did not match the pattern of difference between CV values per extent across sites. For instance, both ULP2 and s144 showed much greater proportions of exposed wood within unrestored extents (Fig. 3.3.8), but ULP2 displayed the greatest difference in CV values between extents, while s144 showed the least difference (Table 3.3.1 – Table 3.3.2). Similarly, the trend of CV differences between reaches (for depth, width and velocity) was not correlated with the trend of proportion exposed wood between reaches (correlation p-values were 0.24 for depth, 0.61 for width and 0.76 for velocity).

3.4 Discussion

Analyses indicated substantial differences in physical channel habitat heterogeneity between restored and unrestored reaches among study sites. Substantial differences in variation were evident across all spatial scales analyzed, but differences became less pronounced at more local scales. Such differences reflected a recurring pattern of higher variation within unrestored reaches relative to restored extents.

Project-scale habitat metrics indicated that even when magnitudes were comparable between restored and unrestored extents, their degree of variation could be very different. Rank order distributions between restored and unrestored extents at the project scale were not significantly different in most cases. However, when the distributions of CV values from all samples were compared, most were significantly different. Such findings implied that comparing just habitat magnitudes would not have revealed the whole story of restored habitat characteristics within study streams.

Local scale comparisons sought to understand whether there was a structure in the patterns of variation within the project scale. Variation at the project scale was most reflected in the local structure of habitat metrics across approximately 1 – 2 meander wavelengths (100m), only some of which was reflected at the scale of adjacent channel units (20m) within those wavelengths. Variation across meander wavelengths is thought to represent the scale of hydrologic and geomorphic processes that create habitat, the extent over which stream habitat is studied in the field, and the variation in habitat that might be experienced by mobile biotic communities moving between channel units (e.g., Leopold and Wolman 1957; Karr et al. 1986; Kauffman et al. 1999; Fetscher et al. 2009). Many of the same differences observed at the project scale were reflected at this reach scale, with many comparisons indicating greater habitat heterogeneity within unrestored extents. Variation within and among adjacent channel units (20m) indicated habitat that might be experienced by biotic communities that dwell predominantly within one channel unit (e.g., within a pool or riffle), or that migrate between channel units for different life stages (e.g., Larkin et al. 2006; Yarnell 2008) or for refugia access (e.g., Palmer et al. 2000; Brown 2003; Larkin et al. 2006). At this channel unit scale, most comparisons indicated no difference in habitat variation. Comparisons with substantial differences in channel unit variation were consistent with patterns observed at broader scales. Analyses of variation over smaller spatial scales helped to reveal that the variation at the project scale was locally structured, and that there were differences in habitat heterogeneity at scales important to stream biota.

Variation comparisons were not influenced by differences in vegetative cover between extents. The proportion of "exposed wood" within field reaches represented a catch-all for any non-wetted vegetation (typically woody) that covered the view of the channel in the orthoimagery. There were no patterns evident in proportions of "exposed wood" between extents' field reaches (e.g., it was not always higher within either extent; Fig. 3.3.7). There was also no correlation between the pattern of CV values between extents and the pattern of exposed wood between extents (e.g., the relative proportions were not associated with relative variation; Fig. 3.3.7 and Table 3.3.1). Such results suggested that differences in CV values were not related to differences in channel visibility between restored and unrestored extents.

Drainage land cover characteristics were related to variation comparison outcomes more than watershed size. A portion of Table 1.2.1 is reprinted below to display differences in sites' physical characteristics (Table 3.4.1). The drainages of EBCH, PLUM and ULP2, each of which had many instances of reduced variation in habitat across scales, are very different sites in terms of size (3.25km^2, 16.26km^2 and 26.03 km^2, respectively), and Strahler stream orders (first, second and fourth, respectively; Table 3.4.1). However, these sites did stand out in terms of developed and impervious land cover. Sites CRAB, EBCH, PLUM and ULP2 had the highest developed and impervious land cover (Table 3.4.1).

Table 3.4.1. Select site upstream drainage area characteristics.

	144	253	CRAB	EBCH	PLUM	ULP2	WL
Drainage Area (km^2)[a]	3.02	8.82	3.12	16.26	3.25	26.03	6.69
Strahler stream order[a]	2	3	1	2	1	4	4
% US Developed[b]	27.4	34	52.1	73.4	74	48.5	36.1
% US Impervious[b]	3.78	3.86	11.2	17.7	13.1	7.61	5.91

(a) (USEPA and USGS 2005; Wieczorek and LaMotte 2010; MD iMAP 2011,2013)
(b) (USGS 2016b, 2016a)

Drainage land cover characteristics were also associated with different restoration objectives at these sites (Tables 3.4.1 – 3.4.2), which corresponded with the patterns of variation between extents across scales. Impervious land, in particular, is infamously associated with altered hydrologic regimes and associated bank erosion and channel incision (Leopold 1968; Dunne and Leopold 1978; Walsh et al. 2005). Table 3.4.2 outlines the restoration objectives associated with study sites. Restoration projects at CRAB, EBCH, PLUM, ULP2 and WL were initiated with hydrologic regime management as a main objective in each case (Table 3.4.2). These restoration objectives imply practices that purposefully reduce some of the physical heterogeneity because zones of high velocity associated with variable hydrology likely lead to the issues with bank erosion and channel bed incision (e.g., Thompson 1986; Thompson et al. 1999). High variability in depth and width leads to high variability in velocity, as constrictions and pool-riffle transitions create zones of very high velocity and very low velocity (Thompson 1986; Thompson et al. 1999). Therefore, hydrologic regime management may have inherently reduced variability in physical habitat.

Table 3.4.2. Site restoration objectives.

	144[a]	253[b]	CRAB[c]	EBCH[d]	PLUM[e]	ULP2[f]	WL[g]
General ecological benefit	X	X	X	X	X	X	X
Explicitly, habitat creation	X	X			X		X
Nutrient reduction						X	
General sediment reduction	X	X	X	X	X	X	
Explicitly, bank erosion	X	X		X	X	X	X
Explicitly, bed incisision			X	X		X	
Hydrologic regime control			X	X	X	X	X

(a) (Montgomery County DEP 2009); (b) (Montgomery County DEP 2012); (c) (Bauer et al. 2009); (d) (East Goshen Township n.d.); (e) (Clauser and Struble 2008); (f) (Birmingham and Koser 2015); (g) (Montgomery County DEP 2019b).

Alternatively, neither sites 144 or 253 were targeted for hydrologic regime management, and both were explicitly targeted to habitat restoration (Table 3.4.2). These sites also had relatively less developed and impervious land cover in their respective drainage areas (Table 3.4.1). Further, these were the two sites that displayed either no significant difference or higher variation within their restored extents, and no comparisons with significantly higher variation in unrestored extents (Table 3.3.1). The combination of explicit habitat restoration objectives and relatively less stress associated with developed and impervious land cover may have led to more habitat heterogeneity at these two sites (sites 144 and 253).

Overall patterns across scales of analysis revealed that restoration either decreased or had no effect on habitat heterogeneity. Reductions in heterogeneity within restored extents may have major implications for ecological outcomes at these restoration sites. Heterogeneity has been emphasized as an important feature of ecosystem resilience (e.g., Holling 1973; Palmer et al. 2005; Larkin et al. 2006; Lake et al. 2007) and functioning (e.g., Cardinale et al. 2002; Palmer et al. 2005; Larkin et al. 2006), as well as opportunities for biotic survival (Brown 2003; Larkin et al. 2006). Therefore, it has been suggested as an important component of ecological restoration (Larkin et al. 2006). However, as was evident at some of the study sites (particularly ULP2), some restoration practices have been shown to reduce physical heterogeneity (Holling and Meffe

1996; Hilderbrand et al. 2005; Wohl et al. 2005). And similarly to other study sites (e.g., s144, s253 and CRAB), some studies have shown mixed results of restoration effects on heterogeneity (Larkin et al. 2006; Mason et al. 2012; Rubin et al. 2017). If restoration projects are targeted to prevent or reduce flood disturbance, channel movement, and other natural hydrologic and geomorphic processes, they may also prevent or reduce natural variation important to ecosystem structure and function (e.g., Larkin et al. 2006; Poole 2010). Local scales of habitat heterogeneity indicate access to refugia (Brown 2003) and to a variety of habitats necessary for different stages of aquatic biota life cycles (e.g., Wheaton et al. 2004; Yarnell 2008). Therefore, reduction in local habitat heterogeneity within restoration projects could have major implications on aquatic ecosystems.

One of the great benefits of this kind of continuous, high-resolution and extensive data is in the diversity of questions that can be addressed. Previous studies have shown that this kind of data can be used to map fish habitat suitability continuously over broad extents (Wheaton et al. 2004; Carbonneau et al. 2012; Tamminga et al. 2015). This data and the analyses presented also provide a powerful new tool to assess ecological outcomes of restoration. There are numerous possibilities just in comparing this continuous physical habitat data with biotic community outcomes. Of course, there is a strong desire to link biotic outcomes to specific restoration practices to bolster arguments of ecological benefit (e.g., Wohl et al. 2005). Some have questioned the application of heterogeneity theory to stream and river restoration, citing a lack of substantial evidence of its association with improved ecosystems (Wheaton et al. 2004; Larkin et al. 2006; Palmer et al. 2010). Future work could use analyses of variation across multiple scales, as presented in this study, to assess biotic community outcomes of restoration. Such an analysis could help to tease out the degree and scale(s) of association between environmental heterogeneity and biotic diversity in restored streams. Variation analyses in the present study provides an objective means to test that relationship on multiple spatial scales. Particularly pairing this spatially explicit physical habitat data with spatially explicit biotic samples (e.g., on the microhabitat scale, 10^{-1} - 10^{0}m) may help to “improve understanding of how heterogeneity influences the biota and functions targeted in restoration in specific ecological systems.” (*Foundations of Restoration Ecology*, pg. 285).

Conclusion

Measurement of physical stream habitat typically involves a tradeoff between precision, extent and spatial explicitness. Precise and spatially explicit measurements, such as those made while surveying cross-section profiles, are collected over limited extents that must be interpolated to represent broader extents. Alternatively, estimates of habitat features, such as the extent of a riffle over a given reach, are more continuous over broader extents, but their characterizations are often imprecise and subjective. Tradeoffs between measurement precision, spatial extent and spatial explicitness can lead to error and thereby reduce associations between proxy estimates and the concept of "physical habitat." Similarly, error can propagate through reduced associations between physical habitat and other components of stream integrity that could limit ability to detect relationship between physical and ecological processes, and therefore effective freshwater management. Physical habitat improvement is a primary objective of many restoration projects, particularly in the study region. It is therefore necessary to adequately measure how physical habitat has changed due to restoration to fully understand ecological outcomes from restoration.

Aerial surveys employing Structure-from-Motion (SfM) photogrammetry, a technique that uses overlapping photos to produce 3D models and large orthophotos, provide a means to obtain physical habitat measures that are more continuous, precise, objective, spatially explicit and repeatable than conventional field survey techniques. Small, readily-available unmanned aerial vehicles (UAVs) with high-resolution cameras make these methods efficient and accessible. Methods were established to extract habitat metrics commonly used in regional monitoring programs from SfM data products (digital elevation models (DEMs) and orthophotos). The wetted channel extent across entire survey areas, as well as dominant sediment classes within reaches, were manually delineated. This allowed for continuous estimates of pixelated depth and width. Channel cross-sections were approximated as distinct integer distances to the downstream end of study areas and used as zones over which to summarize mean depth and width. Resulting cross-sectional areas were used in conjunction with discharge measurements taken during aerial surveys to delineate continuous estimates of cross-section mean flow velocity. Drainage area correction and outlier removal improved the ranges of velocity magnitude per site. Delineated SfM estimates of physical habitat were realistic in pattern, magnitude and extent relative to field observations.

Physical channel habitat within sampling reaches was equivalently represented by field and aerial surveys. There was significant correlation, association and equivalence in the representation of study stream reaches when field and aerial surveys were similarly sampled. Comparisons that were not significantly associated at the scale of spatially paired sampling units became significantly associated or equivalent at the scale of the spatially paired reach. Different methods of data extraction from aerial surveys displayed varying strength in relating field and SfM -based datasets, in some cases resulting in a near 1:1 relationship. Improvement in comparisons from slight changes to placement and orientation of sampling transects at EMAP / SWAMP sites indicated sensitivity of these measurements to transect placement. This suggested the potential for unrecognized bias in field sampling when field technicians decide where to locate cross sections. Transect placement sensitivity also conveyed that SfM was able to pick up heterogeneity at local scales, given different results from slight differences in placement. Depth and flow velocity corrections within MBSS “facies” segments indicated effects of refraction and discharge differences on comparisons between field and SfM data. The great benefit of SfM data is in the ability to revisit the data, to see exactly how and where data were sampled, and to scrutinize where and why the data may need to be corrected.

Because SfM estimates were realistic and representative of study streams, aerial survey data were used to ask questions about restoration outcomes that might typically be asked with field data, but with the advantage of continuity over the entire surveyed extent. Variation in metrics at multiple scales was compared between restored and unrestored extents. Analyses yielded many examples of lower variation within restored extents as compared with unrestored extents, and few examples of higher variation. Reduced habitat heterogeneity, particularly on the local scales of habitat important to aquatic biota, may have major implications for restored ecosystems. Differences in habitat variation were likely a result of the differences in catchment-level conditions, which appeared to drive restoration objectives. Analyses of heterogeneity across scales – including the scale of the restoration project, reach and channel unit – represented a powerful new method for assessing and understanding restoration outcomes. Future work should consider how differences in habitat heterogeneity relate to biotic outcomes, and on what spatial scales, perhaps addressing the role of heterogeneity in restored aquatic ecosystems.

Bibliography

Agisoft. 2019. Orthophoto and dem (with gcps). PhotoScan Tutorial. Accessed September 16. https://www.agisoft.com/pdf/PS_1.3%20-Tutorial%20(BL)%20-%20Orthophoto,%20DEM%20(GCPs).pdf.

Anim, D.O., T.D. Fletcher, G.J. Vietz, G.B. Pasternack, and M.J. Burns. 2018. Restoring in-stream habitat in urban catchments: modify flow or the channel? Ecohydrology 12 (1). doi:10.1002/eco.2050.

Baker, M.E., D.E. Weller, and T.E. Jordan. 2006a. Improved methods for quantifying potential nutrient interception by riparian buffers. Landscape Ecology, no. 21: 1327–45. doi:10.1007/s10980-006-0020-0.

Baker, M.E., M.J. Wiley, M. Carlson Mazur, and P.W. Seelbach. 2006b. A gis model of subsurface water potential for aquatic resource inventory, assessment, and environmental management. Environmental Management 32 (6): 706–19.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, second edition. US Environmental Protection Agency; Office of Water.

Bauer, R., P. Goodman, and C. Marshall. 2009. Lower crabby creek restoration. Valley Creek Restoration Partnership. https://www.delawareriverkeeper.org/sites/default/files/resources/Reports/Lower_Crabby_Creek_Restoration-Good_and_Not_As_Good.pdf.

Bergeron, N., and P. Carbonneau. 2012. 9. geosalar: innovative remote sensing methods for spatially continuous mapping of fluvial habitat at riverscape scale. In Fluvial Remote Sensing for Science and Management, 193–213. Advancing River Restoration and Management 1. Wiley-Blackwell.

Bernhardt, E.S., M. Palmer, J.D. Allan, G. Alexander, K. Barnas, S. Brooks, J. Carr, S. Clayton, C. Dahm, J. Follstad-Shah, and others. 2005. Synthesizing u. s. river restoration efforts. Science 308 (5722): 636–637.

Bernhardt, E.S., and M.A. Palmer. 2011. River restoration: the fuzzy logic of repairing reaches to reverse catchment scale degradation. Ecological Applications 21 (6): 1926–31. doi:10.1890/10-1574.1.

Birmingham, A., and J. Koser. 2015. Upper little patuxent stream restoration: a vision for success, a design-build project. presented at the 2015 Villanova Urban Stormwater Partnership Stormwater Management Symposium. https://www1.villanova.edu/content/dam/villanova/engineering/vcase/sym-presentations/2015/Presentationpdfs/4C2-ULP%20Presentation_VUSP%202015.pdf.

Bisson, P.A., J.L. Nielson, R.A. Palmason, and L.E. Grove. 1982. A system of naming habitat types in small streams, with examples of habitat utilization by salmonids during low stream flow. In American Fisheries Society Symposium, edited by Armantrout, N.A., 62–73. Portland, OR: American Fisheries Society.

Brown, B.L. 2003. Spatial heterogeneity reduces temporal variability in stream insect communities. Ecology Letters 6 (4): 316–25. doi:10.1046/j.1461-0248.2003.00431.x.

Carbonneau, P., M.A. Fonstad, W.A. Marcus, and S.J. Dugdale. 2012. Making riverscapes real. Geomorphology 137 (2012): 74–86.

Carbonneau, P., and H. Piegay. 2012. Advancing River Restoration and Management Ser. : Fluvial Remote Sensing for Science and Management (1). Somerset, GB: Wiley-Blackwell. http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10593155.

Carbonneau, P.E., N. Bergeron, and S.N. Lane. 2005. Automated grain size measurements from airborne remote sensing for long profile measurements of fluvial grain sizes. Water Resources Research 41 (11). doi:10.1029/2005WR003994.

Carbonneau, P.E., and J.T. Dietrich. 2017. Cost-effective non-metric photogrammetry from consumer-grade suas: implications for direct georeferencing of structure from motion photogrammetry. Earth Surface Processes and Landforms 42 (3): 473–86. doi:10.1002/esp.4012.
Carbonneau, P.E., S.N. Lane, and N.E. Bergeron. 2004. Catchment-scale mapping of surface grain size in gravel bed rivers using airborne digital imagery. Water Resources Research 40 (7). doi:10.1029/2003WR002759.
Cardenas, M.B., J.L. Wilson, and V.A. Zlotnik. 2004. Impact of heterogeneity, bed forms, and stream curvature on subchannel hyporheic exchange. Water Resources Research 40 (8). doi:10.1029/2004WR003008.
Cardinale, B.J., M. Palmer, C.M. Swan, S. Brooks, and N.L. Poff. 2002. The influence of substrate heterogeniety on biofilm metabolism in a stream ecosystem. Ecology 83 (2): 412–22.
Carrivick, J.L., and M.W. Smith. 2019. Fluvial and aquatic applications of structure from motion photogrammetry and unmanned aerial vehicle/drone technology. Wiley Interdisciplinary Reviews: Water 6 (1): e1328. doi:10.1002/wat2.1328.
Clauser, A., and R. Struble. 2008. Restoration plan for plum run watershed chester county, pennsylvania. Rettew Associates, Inc.; Brandywine Red Clay Alliance. http://www.brandywineredclay.org/documents/Plum%20Run%20Restoration%20Plan.pdf.
Dietrich, J. 2015. Advanced geographic research: agisoft photoscan crash course (updated for version 1.1.6). Advanced Geographic Research. June 24. http://adv-geo-research.blogspot.com/2015/06/photoscan-crash-course-v1-1.html.
Dietrich, J.T. 2016. Riverscape mapping with helicopter-based structure-from-motion photogrammetry. Geomorphology 252 (2016): 144–57.
Dietrich, J.T. 2017. Bathymetric structure-from-motion: extracting shallow stream bathymetry from multi-view stereo photogrammetry. Earth Surface Processes and Landforms 42 (2): 355–64. doi:10.1002/esp.4060.
Doll, B., G. Jennings, J. Spooner, D. Penrose, J. Usset, J. Blackwell, and M. Fernandez. 2016. Can rapid assessments predict the biotic condition of restored streams? Water 8 (4): 143. doi:10.3390/w8040143.
Dunne, T., and L.B. Leopold. 1978. Water in Environmental Planning. New York: W.H. Freeman and Company.
East Goshen Township. n.d. Project narrative for chester creek stream restoration.
ESRI. 2019. ArcGIS Desktop (version 10.6). Redlands, CA: Environmental Systems Research Institute.
Fausch, K.D., C.E. Torgersen, C.V. Baxter, and H.W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. BioScience 52 (6): 483–98. doi:10.1641/0006-3568(2002)052[0483:LTRBTG]2.0.CO;2.
Fetscher, A.E., L. Busse, and P.R. Ode. 2009. Standard operating procedures for collecting stream algae samples and associated physical habitat and chemical data for ambient bioassessments in california. California State Water Resources Control Board Surface Water Ambient Monitoring Program (SWAMP) Bioassessment SOP 2. http://www.researchgate.net/profile/Lilian_Busse/publication/271078928_Standard_Operating_Procedures_for_Collecting_Stream_Algae_Samples_and_Associated_Physical_Habitat_and_Chemical_Data_for_Ambient_Bioassessments_in_California/links/54bdc2a50cf218da9391be8e.pdf.
Fonstad, M.A., J.T. Dietrich, B.C. Courville, J.L. Jensen, and P.E. Carbonneau. 2013. Topographic structure from motion: a new development in photogrammetric measurement. Earth Surface Processes and Landforms 38 (4): 421–30. doi:10.1002/esp.3366.
Frissell, C.A., W.J. Liss, C.E. Warren, and M.D. Hurley. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. Environmental Management 10 (2): 199–214. doi:10.1007/BF01867358.

Gleason, C.J., L.C. Smith, and J. Lee. 2014. Retrieval of river discharge solely from satellite imagery and at-many-stations hydraulic geometry: sensitivity to river form and optimization parameters. Water Resources Research 50 (12): 9604–19. doi:10.1002/2014WR016109.

Hack, J.T. 1957. Studies of longitudinal stream profiles in virginia and maryland. USGS Numbered Series 294-B. Professional Paper. http://pubs.er.usgs.gov/publication/pp294B.

Hall, R.K., R.L. Watkins, D.T. Heggem, K.B. Jones, P.R. Kaufmann, S.B. Moore, and S.J. Gregory. 2009. Quantifying structural physical habitat attributes using lidar and hyperspectral imagery. Environmental Monitoring and Assessment 159 (1): 63. doi:10.1007/s10661-008-0613-y.

Hannaford, M.J., M.T. Barbour, and V.H. Resh. 1997. Training reduces observer variability in visual-based assessments of stream habitat. Journal of the North American Benthological Society 16 (4): 853–60. doi:10.2307/1468176.

Hawkins, C.P., J.L. Kerschner, P.A. Bisson, M.D. Bryant, L.M. Decker, S.V. Gregory, D.A. McCullough, C.K. Overton, G.H. Reeves, R.J. Steedman, and M.K. Young. 1993. A hierarchical approach to classifying stream habitat features. Fisheries, Environment, 18 (6): 3–12.

Hawkins, C.P., R.H. Norris, J.N. Hogue, and J.W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10 (5): 1456–77.

Hilderbrand, R. unpublished data. Ecological outcomes at maryland stream restoration sites.

Hilderbrand, R.H., A.C. Watts, and A.M. Randle. 2005. The myths of restoration ecology. Ecology and Society 10 (1). doi:10.5751/ES-01277-100119.

Holling, C.S. 1973. Resilience and stability of ecological systems. Annual Review of Ecology and Systematics 4 (1): 1–23. doi:10.1146/annurev.es.04.110173.000245.

Holling, C.S., and G.K. Meffe. 1996. Command and control and the pathology of natural resource management. Conservation Biology 10 (2): 328–37.

Hothorn, T., H. Winell, K. Hornik, M.A. van de Wiel, and A. Zeileis. 2019. Coin: Conditional Inference Procedures in a Permutation Test Framework (version 1.3-1). http://coin.r-forge.r-project.org/.

Karr, J.R., and E.W. Chu. 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press. https://islandpress.org/book/restoring-life-in-running-waters.

Karr, J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant, and I.J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. Special Publication 5. Champaign, IL: Illinois Natural History Survey. https://semspub.epa.gov/work/01/554353.pdf.

Karr, J.R., and C.O. Yoder. 2004. Biological assessment and criteria improve total maximum daily load decision making. Journal of Environmental Engineering 130 (6): 594–604. doi:10.1061/(ASCE)0733-9372(2004)130:6(594).

Kauffman, P.R., P. Levine, E.G. Robinson, C. Seelinger, and D.V. Peck. 1999. Quantifying physical habitat in wadeable streams. EPA/620/R-99/003. U.S. Environmental Protection Agency. https://archive.epa.gov/emap/archive-emap/web/pdf/phyhab.pdf.

Kominoski, J.S., T.J. Hoellein, C.J. Leroy, C.M. Pringle, and C.M. Swan. 2010. Beyond species richness-expanding biodiversity-ecosystem functioning theory in detritus-based streams. River Research and Applications 26 (2010): 67–75.

Kondolf, G.M., A.J. Boulton, S. O'Daniel, G.C. Poole, F.J. Rahel, E.H. Stanley, E. Wohl, Å. Bång, J. Carlstrom, C. Cristoni, H. Huber, S. Koljonen, P. Louhi, and G. Nakamura. 2006. Process-based ecological river restoration : visualizing three-dimensional connectivity and dynamic vectors to recover lost linkages. Ecology & Society 11 (2): 1–16.

Kroll, S. unpublished data. Ecological outcomes at pennsylvania stream restoration sites.

Kroll, S.A., R.J. Horwitz, D.H. Keller, B.W. Sweeney, J.K. Jackson, and L.B. Perez. 2019. Large-scale protection and restoration programs aimed at protecting stream ecosystem integrity: the role of science-based goal-setting, monitoring, and data management. Freshwater Science 38 (1): 000–000.

Kroll, S.A., H. Oakland, C. Collier, and S. Klein. 2016. Year 1 progress report to the pa dep: past restorations – success or failure? PA DEP. The Academy of Natural Sciences of Drexel University.

Kroll, S.A., and H.C. Oakland. 2019. A review of studies documenting the effects of agricultural best management practices on physiochemical and biological measures of stream ecosystem integrity. Natural Areas Journal 39 (1): 58–77.
Lake, P.S., N.R. Bond, and P. Reich. 2007. Linking ecological theory with stream restoration. Freshwater Biology 52.
Lamouroux, N., S. Dolédec, and S. Gayraud. 2004. Biological traits of stream macroinvertebrate communities: effects of microhabitat, reach, and basin filters. Journal of the North American Benthological Society 23 (3): 449–466.
Larkin, D.J., G.L. Bruland, and J.B. Zedler. 2006. Heterogeneity theory and ecological restoration. In Foundations of Restoration Ecology, edited by Palmer, M.A., J.B. Zedler, and D.A. Falk, 2nd ed. Washington, D.C., USA: Island Press. https://islandpress.org/books/foundations-restoration-ecology-0.
Lave, R. 2009. The controversy over natural channel design: substantive explanations and potential avenues for resolution1. JAWRA Journal of the American Water Resources Association 45 (6): 1519–32. doi:10.1111/j.1752-1688.2009.00385.x.
Legleiter, C.J., and M.A. Fonstad. 2012. 3. an introduction to the physical basis for deriving river information by optical remote sensing. In Fluvial Remote Sensing for Science and Management, 43–69. Advancing River Restoration and Management 1. Wiley-Blackwell.
Leopold, L.B. 1968. Hydrology for urban land planning - a guidebook on the hydrologic effects of urban land use. Geological Survey Circular 554. U. S. Geological Survey. https://pubs.usgs.gov/circ/1968/0554/report.pdf.
Leopold, L.B., and T. Maddock Jr. 1953. The hydraulic geometry of stream channels and some physiographic implications. USGS Numbered Series 252. Professional Paper. Washington, D.C.: U.S. Government Printing Office. http://pubs.er.usgs.gov/publication/pp252.
Leopold, L.B., and M.G. Wolman. 1957. River channel patterns: braided, meandering, and straight. USGS Numbered Series 282-B. Professional Paper. Washington, D.C.: U.S. Government Printing Office. http://pubs.er.usgs.gov/publication/pp282B.
Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. The American Naturalist 87 (836): 331–33.
Levin, S.A. 1976. Population dynamic models in heterogeneous environments. Annual Review of Ecology and Systematics 7: 287–310.
Mallison, H. 2015. Photogrammetry tutorial 11: how to handle a project in agisoft photoscan. Photogrammetry Tutorial. https://dinosaurpalaeo.wordpress.com/2015/10/11/photogrammetry-tutorial-11-how-to-handle-a-project-in-agisoft-photoscan/.
Marcus, W.A., and M.A. Fonstad. 2008. Optical remote mapping of rivers at sub-meter resolutions and watershed extents. Earth Surface Processes and Landforms 33 (1): 4–24. doi:10.1002/esp.1637.
Marcus, W.A., M.A. Fonstad, and C.J. Legleiter. 2012. Management applications of optical remote sensing in the active river channel. In Fluvial Remote Sensing for Science and Management. Vol. 1. Advancing River Restoration and Management. United Kingdom: Wiley-Blackwell.
Marcus, W.A., C.J. Legleiter, R.J. Aspinall, J.W. Boardman, and R.L. Crabtree. 2003. High spatial resolution hyperspectral mapping of in-stream habitats, depths, and woody debris in mountain streams. Geomorphology 55 (2003): 363–80.
Marteau, B., D. Vericat, C. Gibbins, R.J. Batalla, and D.R. Green. 2017. Application of structure-from-motion photogrammetry to river restoration. Earth Surface Processes and Landforms 42 (3): 503–15. doi:10.1002/esp.4086.
Mason, S.J.K., B.L. McGlynn, and G.C. Poole. 2012. Hydrologic response to channel reconfiguration on silver bow creek, montana. Journal of Hydrology 438–439 (May): 125–36. doi:10.1016/j.jhydrol.2012.03.010.
MD iMAP. 2011. Maryland lidar howard county - dem meters. Elevation. Maryland iMAP. https://data.imap.maryland.gov/datasets/a93ed945fd4b4b30a802bba36a2a7987.

MD iMAP. 2013. Maryland lidar montgomery county - dem meters. Elevation. Maryland iMAP. https://data.imap.maryland.gov/datasets/bd7ac606a4d3400db91c923801b8a1d7.

Meador, M.R., and R.M. Goldstein. 2003. Assessing water quality at large geographic scales: relations among land use, water physicochemistry, riparian condition, and fish community structure. Environmental Management 31 (4): 0504–0517.

Mohamoud, Y. 2004. Comparison of hydrologic responses at different watershed scales. EPA/600/R-04/103. Athens, GA: U.S. Environmental Protection Agency Ecosystems Research Division. https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=525020.

Montgomery County DEP. 2009. Batchellors run east report. 3.8. Montgomery County Maryland Department of Environmental Protection. https://www.montgomerycountymd.gov/DEP/Resources/Files/downloads/restoration/monitoring/Batchellors%20Run%20East%20report_2009.pdf.

Montgomery County DEP. 2012. Watershed restoration fact sheet: batchellors run i & ii stream restoration. Montgomery County Maryland Department of Environmental Protection. https://www.montgomerycountymd.gov/DEP/Resources/Files/downloads/restoration/batchellors-run-east/Project-fact-sheet.pdf.

Montgomery County DEP. 2019a. Batchellors run i & ii stream restoration project. Montgomery County Maryland Department of Environmental Protection. https://www.montgomerycountymd.gov/water/restoration/batchellors-run-east.html.

Montgomery County DEP. 2019b. Woodlawn stream restoration. Montgomery County Maryland Department of Environmental Protection. https://www.montgomerycountymd.gov/water/restoration/woodlawn.html.

National hydrography dataset (nhd). 2017. Reston, VA: U.S. Geological Survey. https://nhd.usgs.gov/.

Palmer, M.A., R.F. Ambrose, and N.L. Poff. 1997. Ecological theory and community restoration ecology. Restoration Ecology 5 (4): 291–300. doi:10.1046/j.1526-100X.1997.00543.x.

Palmer, M.A., E.S. Bernhardt, J.D. Allan, P.S. Lake, G. Alexander, S. Brooks, J. Carr, S. Clayton, C.N. Dahm, J. Follstad Shah, and others. 2005. Standards for ecologically successful river restoration. Journal of Applied Ecology 42 (2): 208–217.

Palmer, M.A., H.L. Menninger, and E. Bernhardt. 2010. River restoration, habitat heterogeneity and biodiversity: a failure of theory or practice? Freshwater Biology 55 (s1): 205–22. doi:10.1111/j.1365-2427.2009.02372.x.

Palmer, M.A., C.M. Swan, K. Nelson, P. Silver, and R. Alvestad. 2000. Streambed landscapes: evidence that stream invertebrates respond to the type and spatial arrangement of patches. Landscape Ecology 15 (6): 563–76. doi:10.1023/A:1008194130695.

Palmer, M.A., J.B. Zedler, and D.A. Falk, eds. 2006. Foundations of Restoration Ecology. 2nd ed. Washington, D.C., USA: Island Press. https://islandpress.org/books/foundations-restoration-ecology-0.

Patrick, R. 1949. A proposed biological measure of stream conditions, based on a survey of the conestoga basin, lancaster county, pennsylvania. Proceedings of the Academy of Natural Sciences of Philadelphia 101: 277–341.

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. Rapid bioassessment protocols for use in streams and rivers- benthic macroinvertebrates and fish. Protocol EPA/440/4-89/001. Washington D.C.: U.S. Environmental Protection Agency. nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=9100LGCA.TXT.

Poff, N.L., and J.V. Ward. 1990. Physical habitat template of lotic systems: recovery in the context of historical pattern of spatiotemporal heterogeneity. Environmental Management 14 (5): 629. doi:10.1007/BF02394714.

Poole, G.C. 2010. Stream hydrogeomorphology as a physical science basis for advances in stream ecology. Journal of the North American Benthological Society 29 (1): 12–25. doi:10.1899/08-070.1.

Poole, G.C., C.A. Frissell, and S.C. Ralph. 1997. In-stream habitat unit classification: inadequacies for monitoring and some consequences for management1. JAWRA Journal of the American Water Resources Association 33 (4): 879–96. doi:10.1111/j.1752-1688.1997.tb04112.x.
Power, M.E. 1992. Habitat heterogeneity and the functional signifiance of fish in river food webs. Ecology 73 (5): 1675–88. doi:10.2307/1940019.
R Core Team. 2018. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Ralph, S.C., and G.C. Poole. 2003. 9. putting monitoring first: designing accountable ecosystem restoration and management plans. In Restoration of Puget Sound Rivers, 226–47. http://wvvw.krisweb.com/biblio/gen_uofw_ralphetal_ip.pdf.
Robinson, A. 2016. Equivalence: Provides Tests and Graphics for Assessing Tests of Equivalence. (version R package version 0.7.2). https://CRAN.R-project.org/package=equivalence.
Robinson, A.P., and R.E. Froese. 2004. Model validation using equivalence tests. Ecological Modelling 176 (3): 349–58. doi:10.1016/j.ecolmodel.2004.01.013.
Rosgen, D. 1996. Applied River Morphology. 2nd ed. Wildland Hydrology Consultants. https://wildlandhydrology.com/books/?id=32&course=Applied+River+Morphology.
Rosgen, D.L. 1994. A classification of natural rivers. CATENA 22 (3): 169–99. doi:10.1016/0341-8162(94)90001-9.
Rosgen, D.L. 2011. Natural channel design: fundamental concepts, assumptions, and methods. In Stream Restoration in Dynamic Fluvial Systems, edited by Simon, A., S.J. Bennett, and J.M. Castro, 69–93. American Geophysical Union. doi:10.1029/2010GM000990.
Rubin, Z., G.M. Kondolf, and B. Rios-Touma. 2017. Evaluating stream restoration projects: what do we learn from monitoring? Water 9 (174). doi:10.3390/w9030174.
Seelbach, P.W., M.J. Wiley, M.E. Baker, and K.E. Wehrly. 2006. Initial classification of river valley segments across michigan's lower peninsula. American Fisheries Society Symposium 48: 25–48.
Seelbach, P.W., M.J. Wiley, J.C. Kotanchik, and M.E. Baker. 1997. A landscape-based ecological classification system for river valley segments in lower michigan (mi-vsec version 1.0). 2036. Fisheries Division Research Report. State of Michigan Dept of Natural Resources. https://www.researchgate.net/publication/267702539_A_Landscape-Based_Ecological_Classification_System_for_River_Valley_Segments_in_Lower_Michigan_MI-VSEC_version_10.
Simon, A., M. Doyle, M. Kondolf, F. Shields, B. Rhoads, and M. MCPHILLIPS. 2007. Critical evaluation of how the rosgen classification and associated natural channel design methods fail to integrate and quantify fluvial processes and channel response. Journal of the American Water Resources Association 43 (5): 1117–1131.
Simonson, T. 1993. Correspondence and relative precision of stream habitat features estimated at two spatial scales. JOURNAL OF FRESHWATER ECOLOGY 8 (4): 363–73.
SonTek/YSI. 2007. FlowTracker technical manual. https://www.uvm.edu/bwrl/lab_docs/manuals/Flow_Tracker_Manual.pdf.
Stranko, S., D. Boward, J. Kilian, A. Becker, M. Ashton, M. Southerland, B. Franks, W. Harbold, and J. Cessna. 2015. Maryland biological stream survey round four sampling manual. Maryland Department of Natural Resources. https://dnr.maryland.gov/streams/Publications/R4Manual.pdf.
Swan, C.M. 2011. Consumer presence and resource diversity independently induce stability of ecosystem function in a piedmont stream. Ecosphere 2 (12).
Swan, C.M., and B.L. Brown. 2017. Metacommunity theory meets restoration: isolation may mediate how ecological communities respond to stream restoration. Ecological Applications, May, 1–11.
Tamminga, A., C. Hugenholtz, B. Eaton, and M. Lapointe. 2015. Hyperspatial remote sensing of channel reach morphology and hydraulic fish habitat using an unmanned aerial vehicle (uav): a first assessment in the context of river research and management. River Research and Applications 31 (3): 379–91. doi:10.1002/rra.2743.

The national elevation dataset (ned). 2016. Reston, VA: United States Geological Survey, The National Map. https://nationalmap.gov/elevation.html.
Thompson, A. 1986. Secondary flows and the pool-riffle unit: a case study of the processes of meander development. Earth Surface Processes and Landforms 11: 631–41.
Thompson, D.M., E.E. Wohl, and R.D. Jarrett. 1999. Velocity reversals and sediment sorting in pools and riffles controlled by channel constrictions. Geomorphology 27 (3–4): 13. doi:10.1016/S0169-555X(98)00082-8.
Thomson, J.R., M.P. Taylor, K.A. Fryirs, and G.J. Brierley. 2001. A geomorphological framework for river characterization and habitat assessment. Aquatic Conservation: Marine and Freshwater Ecosystems 11 (5): 373–89. doi:10.1002/aqc.467.
Turnipseed, D.P., and V.B. Sauer. 2010. Discharge measurements at gaging stations. USGS Numbered Series 3-A8. Techniques and Methods. U.S. Geological Survey. http://pubs.er.usgs.gov/publication/tm3A8.
US EPA. 2004. Wadeable Streams Assessment : Quality Assurance Project Plan. Washington, DC: U.S. Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds.
USEPA, and USGS. 2005. National hydrography dataset plus – nhdplus – strahler order. Tabular digital data. US Environmental Protection Agency; US Geological Survey. http://www.horizon-systems.com/NHDPlus/StrahlerList.php.
USGS. 2016a. The StreamStats Program for Maryland. U.S. Geological Survey. http://water.usgs.gov/osw/streamstats/maryland.html.
USGS. 2016b. The StreamStats Program for Pennsylvania. U.S. Geological Survey. http://water.usgs.gov/osw/streamstats/pennsylvania.html.
Vellend, M. 2016. The Theory of Ecological Communities. Monographs in Population Biology 57. Princeton and Oxford: Princeton University Press.
Ventura, D., M. Bruno, G. Jona Lasinio, A. Belluscio, and G. Ardizzone. 2016. A low-cost drone based application for identifying and mapping of coastal fish nursery grounds. Estuarine, Coastal and Shelf Science 171 (March): 85–98. doi:10.1016/j.ecss.2016.01.030.
Walsh, C.J., T.D. Fletcher, and A.R. Ladson. 2005. Stream restoration in urban catchments through redesigning stormwater systems: looking to the catchment to save the stream. Journal of the North American Benthological Society 24 (3): 690–705. doi:10.1899/04-020.1.
Wang, L., T.D. Simonson, and J. Lyons. 1996. Accuracy and precision of selected stream habitat estimates. North American Journal of Fisheries Management 16 (2): 340–47.
Wellek, S. 2003. Testing Statistical Hypotheses of Equivalence. Chapman and Hall/CRC.
Wellek, S. 2010. Testing Statistical Hypotheses of Equivalence and Noninferiority. 2nd ed. Boca Raton: Chapman and Hall/CRC. http://proxy-bc.researchport.umd.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid&db=nlebk&AN=342376&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover.
Wheaton, J.M., J. Brasington, S.E. Darby, J. Merz, G.B. Pasternack, D. Sear, and D. Vericat. 2010. Linking geomorphic changes to salmonid habitat at a scale relevant to fish. River Research and Applications 26 (4): 469–86. doi:10.1002/rra.1305.
Wheaton, J.M., G.B. Pasternack, and J.E. Merz. 2004. Spawning habitat rehabilitation -i. conceptual approach and methods. International Journal of River Basin Management 2 (1): 3–20. doi:10.1080/15715124.2004.9635218.
Wheaton, Joseph M., Pasternack, Gregory B., and Merz, Joseph E. 2004. Use of habitat heterogeneity in salmonid spawning habitat rehabilitation design. In Aquatic Habitats: Analysis & Restoration, 791–96. Madrid.
Whitlock, M., and D. Schluter. 2015. The Analysis of Biological Data. 2nd ed. Macmillan. https://whitlockschluter.zoology.ubc.ca/.

Wieczorek, M.E., and A.E. LaMotte. 2010. Attributes for nhdplus catchments (version 1.1): basin characteristics, 2002. Tabular digital data. Reston, Virginia: U.S. Geological Survey. https://water.usgs.gov/lookup/getspatial?nhd_bchar.

Wohl, E., P.L. Angermeier, B. Bledsoe, G.M. Kondolf, L. MacDonnell, D.M. Merritt, M.A. Palmer, N.L. Poff, and D. Tarboton. 2005. River restoration. Water Resources Research 41 (10): W10301. doi:10.1029/2005WR003985.

Wolman, M.G. 1954. A method of sampling coarse river-bed material. Eos, Transactions American Geophysical Union 35 (6): 951–56. doi:10.1029/TR035i006p00951.

Woodget, A.S. 2015. Quantifying physical river habitat parameters using hyperspatial resolution uas imagery and sfm-photogrammetry. University of Worcester. http://eprints.worc.ac.uk/3830/1/PhD_AmyWoodget_.pdf.

Woodget, A.S., and R. Austrums. 2017. Subaerial gravel size measurement using topographic data derived from a uav-sfm approach. Earth Surface Processes and Landforms 42 (9): 1434–43. doi:10.1002/esp.4139.

Woodget, A.S., R. Austrums, I.P. Maddock, and E. Habit. 2017. Drones and digital photogrammetry: from classifications to continuums for monitoring river habitat and hydromorphology. Wiley Interdisciplinary Reviews: Water 4 (4): e1222. doi:10.1002/wat2.1222.

Woodget, A.S., P.E. Carbonneau, F. Visser, and I.P. Maddock. 2015. Quantifying submerged fluvial topography using hyperspatial resolution uas imagery and structure from motion photogrammetry. Earth Surface Processes and Landforms 40 (2015): 47–64.

Woodget, A.S., C. Fyffe, and P.E. Carbonneau. 2018. From manned to unmanned aircraft: adapting airborne particle size mapping methodologies to the characteristics of suas and sfm. Earth Surface Processes and Landforms 43 (4): 857–70. doi:10.1002/esp.4285.

Woodget, A.S., F. Visser, I.P. Maddock, and P.E. Carbonneau. 2016. The accuracy and reliability of traditional surface flow type mapping: is it time for a new method of characterizing physical river habitat? River Research and Applications 32 (9): 1902–14. doi:10.1002/rra.3047.

Yarnell, S. 2008. QUANTIFYING physical habitat heterogeneity in an ecologically meaningful manner: a case study of the habitat preferences of the foothill yellow-legged frog. In Landscape Ecology Research Trends, 24. Nova Science Publishers.

Yuen, K.K. 1974. The two-sample trimmed t for unequal population variances. Biometrika 61 (1): 165–70. doi:10.1093/biomet/61.1.165.

Yuen, K.K., and W.J. Dixon. 1973. The approximate behaviour and performance of the two-sample trimmed t. Biometrika 60 (2): 369–74. doi:10.2307/2334550.