Hydrography Mapping Supporting Modeling and Targeted Conservation: Project Overview and Lessons Learned

a

Cooperative Agreement for Geospatial Support to the Chesapeake Bay Program Office

Prepared by
Chesapeake Conservancy
1212 West Street, Annapolis, MD 21401

Prepared for
U.S. Environmental Protection Agency
Region 3, Chesapeake Bay Program Office
1750 Forest Drive, Suite 130
Annapolis, MD 21401

September 30 2024
Revised: May 21 2025

Project Start Date: June 2018
Project End Date: June 2024

Funder: United States Environmental Protection Agency
Funding opportunity number: CB-96363002

Principal Investigators:
Dr. Matthew Baker
Professor of Geography and Environmental Systems
University of Maryland, Baltimore County
mbaker@umbc.edu
410-455-3759

David Saavedra
Senior Geospatial Technical Lead
Chesapeake Conservancy
dsaavedra@chesapeakeconservancy.org

Dr. Xuezhi Cang
Postdoctoral Associate
University of Maryland, Baltimore County
xcang@umbc.edu

Project Administrator:

Katie Walker
Geospatial Program Manager
Chesapeake Conservancy
kwalker@chesapeakeconservancy.org
443-345-5985

Technical Advisor:
Peter Claggett
Research Geographer
U.S. Geological Survey
Lower Mississippi-Gulf Water Science Center
Chesapeake Bay Program
pclagget@chesapeakebay.net

Project Officer:

Kaylyn Gootman
Life Scientist
U.S. Environmental Protection Agency
Chesapeake Bay Program Office
gootman.kaylyn@epa.gov
410-267-5705

Table of Contents

Introduction and Background

Stream channel mapping, particularly in headwaters, is critical for a suite of environmental applications including hydrologic and hydraulic modeling, soil conservation, wetland conservation, water quality analysis, and regulatory policy (Montgomery and Dietrich 1988; Hancock and Evans 2006; Baker et al. 2007; Julian et al. 2012). However, accurate mapping of channel heads is a long-standing challenge in geomorphology, largely due to the fact that channels form as a result of several different processes and because human activity can drastically alter both key processes as well as the geomorphic evidence used as criteria for delineation (Rodriguez-Iturbe and Rinaldo 2001). The complexity and variety of channel features often defies easy categorization, even by skilled interpreters, and thus complicates attempts at automated channel delineation. Accurately delineating stream channels from terrain data has been a long-standing challenge in the field of geomorphology.

A novel advance in the field of geomorphometry is the geomorphon approach to terrain classification (Jasiewicz and Stepinski 2013). Using line-of-sight concepts from computer vision, the method classifies a gridded terrain surface into its constituent ternary features. Geomorphon works by assessing the relative elevations along eight lines of sight surrounding a focal cell, out to a user-specified search radius. Based on the surrounding pattern of higher, lower, or equal elevations, the algorithm assigns a landform class to the focal cell.

Baker et al. (2018) explored the use of geomorphons for automated stream channel detection among other terrain interpretations including curvature, positive topographic openness, and negative topographic openness. The 2018 work served as a basis for the current project and involved a detailed literature review on the past and current state of automated channel mapping, a field-verification of channel heads in 14 study areas spanning 5 different physiographic provinces across the Chesapeake Bay watershed, an evaluation of the performance of different methods for channel detection, and established an automated workflow for producing connected stream maps from 1-meter (m) LiDAR-derived digital elevation models (DEMs) using geomorphons. Complementary work that helped inform and was informed by these initial findings has since been published elsewhere by Bezerra et al. (2020), Metes et al. (2022), and Metes et al. (2024).

Through this project, the methods developed in 2018 were refined, expanded, and applied to implement an automated technique for stream channel extraction and identification of other natural and man-made water conveyance features. These features were used to develop a connected network throughout the Chesapeake Bay watershed using 1-m LiDAR DEMs.

Data Description

Six spatial data layers were produced as outputs from this project: 1-m "channel-scale" geomorphon rasters, 10-m "valley-scale" geomorphon rasters, stream polygons, stream polylines, agricultural ditch polygons, and roadside ditch polygons. These data were produced for the full extent of the Chesapeake Bay watershed as delineated by the United States Geological Survey (USGS) Watershed Boundary Dataset and are packaged for distribution at the HUC8 level. Spatial data and corresponding metadata can be found at USGS ScienceBase (https://doi.org/10.5066/P1GRAPEX).

The 1-m channel-scale geomorphon and 10-m valley-scale geomorphon rasters characterize the terrain into ten constituent landforms according to line-of-sight surrounding focal cells, as described by Jasiewicz and Stepinski (2013). The 1-m channel-scale geomorphon rasters are produced using 1-m LiDAR DEMs and a narrow search radius, such that only local terrain is assessed during the line-of-sight analysis. The 10-m valley-scale geomorphon rasters are produced from a coarsened version of the same 1-m LiDAR DEMs, aggregated to 10-m resolution. The valley-scale geomorphons are generated using a very wide search radius so that broad landscape features including stream valleys and floodplains are perceived by the line-of-sight analysis. Both geomorphon layers are used in conjunction and serve as the basis for our delineation of hydrologic features.

The stream polygon layer is a two-dimensional representation of channel-like features that could be considered stream channels. The stream polygons are identified using an interpretation of channel-like depressions in the 1-m channel-scale geomorphons as they relate to the convergent terrain identified in the 10-m valley-scale geomorphons. Additional steps were taken in an attempt to remove non-fluvial features from the set of channel-like depressions and isolate the remaining fluvial features that could be considered part of a stream network. Channels visible in LiDAR DEMs may appear discontinuous where they are crossed by culverts or bridges, in areas where streams are routed underground via pipes, in areas of dense vegetation where LiDAR cannot penetrate effectively, or in areas where the channels are naturally less defined or flow underground (e.g. karst topography). In these instances, gaps are present in the stream polygon network corresponding to the locations where a visible stream channel could not be perceived from the DEM. In some cases, gaps may exist where mapped channel polygons were inadvertently removed from the initial set of 1-m geomorphon features during our automated filtering process.

The stream polyline layer connects the discontinuous stream polygons into a continuous one-dimensional linear network with attributes describing each reach in the network. Portions of each reach that span a gap between two or more channel-like features (i.e. the portion of the steam polyline that does not directly overlap a stream polygon) are considered "connectors" and their proportion is indicated in an attribute for each polyline. In most cases connectors represent stream length that is not mappable due to detection limits, anthropogenic modification, or natural disconnections, but across which it was desirable to link channel segments with a line. Other attributes describe the dimensions of

the stream reach, physical properties such as length and elevation, and its position within the larger stream network.

The agricultural and roadside ditch polygons are a subset of channel-like features identified from geomorphons that are most likely to be ditches based on interpretations of their physical properties such as shape, width, and depth, as well as their spatial context (e.g. near roads or agricultural fields). Like the stream polygons, agricultural and roadside ditch polygons are not necessarily contiguous and gaps in the network may be present where the ditches could not be detected in LiDAR. Agricultural and roadside ditches are not mutually exclusive of each other (e.g. ditches running parallel to roads and surrounded by agricultural fields may be present in both datasets), nor are they mutually exclusive of stream polygons; in some areas (e.g. Delmarva peninsula) the local drainage network is made up almost entirely of ditches and a given feature may be present in the stream channel, agricultural ditch, and roadside ditch datasets.

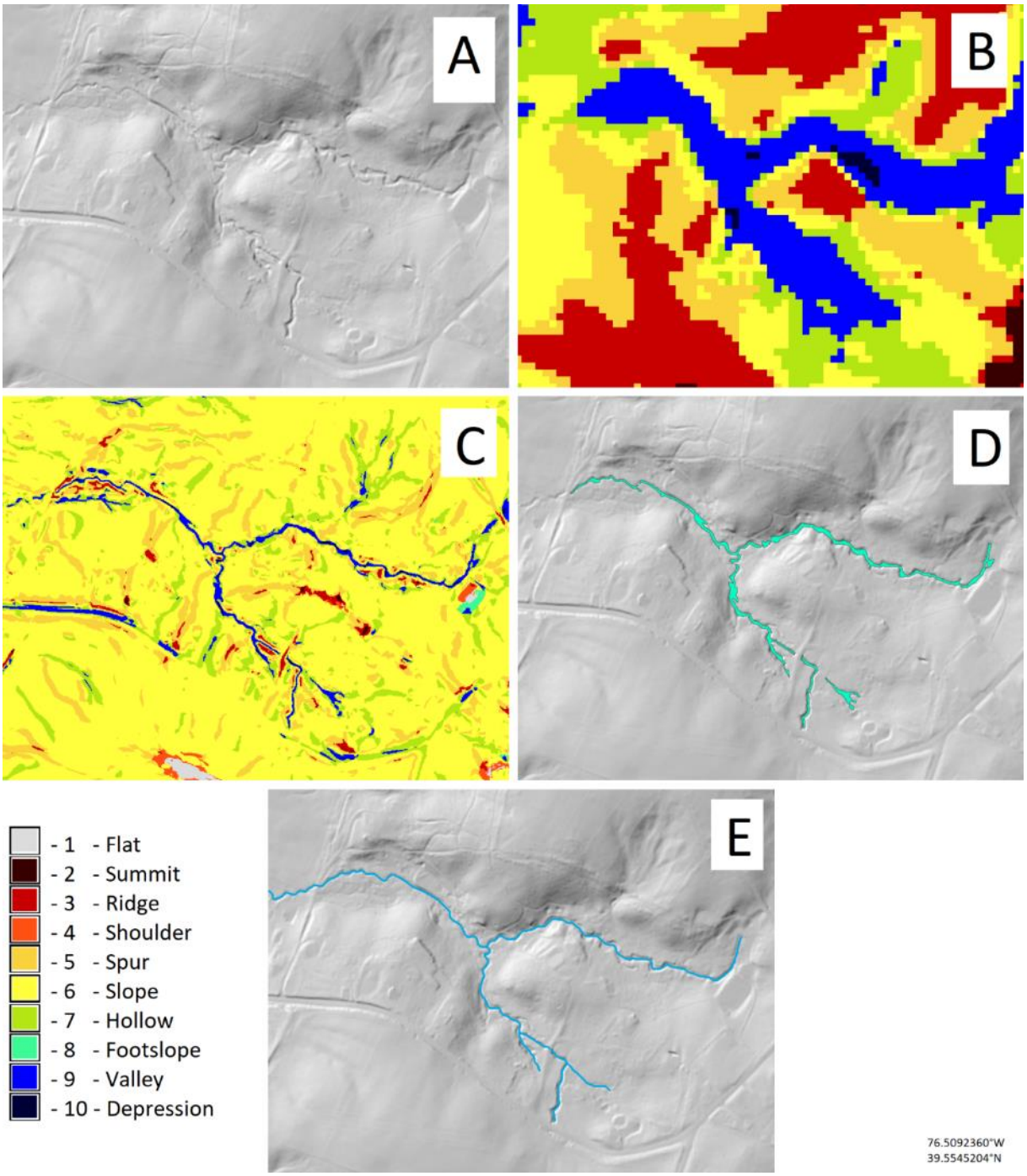


Figure 1: Summarization of stream channel mapping process. A) Shaded relief from LiDAR DEM showing stream channels in a human-modified landscape B) 10-m geomorphon classification C) 1-m geomorphon classification D) stream channel polygons extracted from geomorphons E) Connected stream polyline network

Methodology

Processing Elevation Data

Elevation data were crucial to the delineation of stream networks (Figure 1, Panel A). We obtained the majority of our DEMs from the USGS 3D Elevation Program's (3DEP) 1-m resolution DEM collection (USGS, 2024). In the few instances where 3DEP DEMs were not available, DEMs were obtained from state or local sources at the highest resolution available. Elevation sources were updated as new LiDAR was collected and new DEMs became available; for final data production, we relied on the most recent elevation data publicly available as of May 2024. An inventory of LiDAR elevation sources used for this project is provided in Appendix 1.

While the majority of DEMs were acquired at 1-m resolution, any DEMs obtained in a resolution coarser or finer than 1-m were resampled to 1-m resolution using bilinear interpolation to ensure uniformity and consistency in the final product. All DEMs were reprojected to the USA Contiguous Albers Equal Area Conic USGS projected coordinate system (WKID: 102039). Lastly, the cell alignment for all DEMs was snapped to a singular reference grid spanning the entirety of the Chesapeake Bay watershed, to ensure grid alignment among disparate HUC8 watersheds.

To remove noise and allow for cleaner and more contiguous feature identification, LiDAR elevation data are often denoised or smoothed using a feature-preserving filter (Passalacqua et al. 2010, Lashermes 2007). We pre-processed LiDAR DEMs using a mesh-denoising filter (Sun et al. 2007; Conrad et al. 2015) with a custom script in Python and GDAL prior to assessment of terrain properties. Mesh denoising of LiDAR DEMs removes many irregularities in a surface while preserving features and edges. Following mesh denoising, the 1-m DEMs were resampled to produce a 10-m DEM that aligned with 1-m pixels. Resampling further smoothed terrain and often, but not always, resolved connections between adjacent geomorphic features interrupted by anthropogenic modification. Other investigators have employed "context pixels" (i.e., segments of surrounding terrain in a DEM) and Convolutional Neural Network classifiers to attempt to improve the interpretation of interrupted valleys in terrain interpretations, yet were met with mixed results (Paredes-Tavares et al. 2023).

Identification of Channel-like Features

Geomorphons were calculated using the *r.geomorphon* extension of the free and open-source GRASS GIS 7.8.6 software. *r.geomorphon* has five required user-input parameters that can be customized to produce a landform classification map tailored to the user's needs. The five required parameters are: elevation, outer search radius, inner search radius, flatness threshold, and flatness distance. Several optional outputs provide additional information about the geometry and relative position of geomorphons in the surrounding terrain. To detect channel-like features using geomorphons, the outer search radius was set to a relatively small distance of 20 meters in order to constrain the line-of-sight analysis to the terrain in the local vicinity of the focal cell. The outer search radius defines the maximum

extent of terrain analyzed during the line-of-sight procedure. An inner search radius of 0 meters was specified so that no pixels were skipped and those immediately adjacent to the focal cell would be interpreted, and a 1-degree flatness threshold was used to detect features with sufficient relief. An inner search radius is used to ignore elevation values proximate to the focal cell, whereas the flatness threshold defines the angular difference between the elevation of a neighbor and the focal cell in order to be considered. By using this combination of outer and inner search radii with a detailed LiDAR DEM, channel banks were interpreted as valley walls, thus channel-like depressions were classified as valleys or pits (Figure 1, Panel C). In addition to the default classification of ten landforms, we produced rasters for all optional geomorphon outputs including: intensity (mean elevation difference between the focal cell and the visibility neighborhood), range (difference between the minimum and maximum elevations in the visibility neighborhood), variance (variance of the elevation of the visibility neighborhood), extend (area of the visibility neighborhood), azimuth (orientation of the least-square fit to the visibility neighborhood), elongation (ratio of the dimensions of the bounding box around the visibility neighborhood), width (shorter of the elongation dimensions), and the ternary code (a code indicating one of 498 unique geomorphon configurations).

Development of the Valley Network

We developed maps of river valleys following the suggestions of Jasiewicz and Metz (2011) and Hooshyar et al. (2016) who used delineation of river valleys as an initial geomorphic elimination of spurious depressions. To delineate valleys, the resampled 10-m DEM was provided as input to *r.geomorphon*. The outer search radius, which constrained the line-of-sight analysis, was set to 1,000 meters – a distance easily large enough to encompass headwaters and sufficient to capture most larger-order valleys. An inner search radius, also referred to as a "skip" radius, was also specified. This parameter further constrains the line-of-sight analysis such that any pixels between the focal cell and the skip radius are not considered in the calculation of geomorphons. We specified a 20-meter skip radius to prevent small incised channels and other local irregularities from interrupting the line-of-sight and allow better detection of broader features like headwater valleys. To improve detection of low relief valleys, we reduced the flatness threshold to 0.1 degree (Figure 1, Panel B).

To better distinguish features in the valley-extent geomorphon map associated with stream valleys, we focused on convergent features that were clearly connected to larger-order fluvial systems. The connected valley network thus consisted of all pit and valley classes from the valley-extent geomorphon map that were contiguous with each other or with fine resolution channel-like features. Valley contiguity was defined as all pit and valley cells connected through other pit, valley, or channel-like features at both valley and channel extents. In practice, major anthropogenic infrastructure could disrupt valley contiguity, even at a 10-m resolution. Therefore, we retained the largest percentile of valley features regardless of their contiguity. We used the resulting connected valley network as a geomorphic filter to eliminate channel-like depressions identified by geomorphons at the channel extent that were not contained by, or contiguous with, the valley network.

Isolating and Classifying the Channel Skeleton

Following the geomorphic filtering of channel-like features by the valley network, we assessed the shapes of the remaining features using the compact circle shape index. The compact circle index was calculated as the ratio of a feature's perimeter to the perimeter of a circle of the same area. Elongate features or those with more complex boundaries were preserved, whereas those with excessively compact, circular shapes were removed. In this manner, we aimed to eliminate small round ponds, detention features, and other spurious depressions in favor of longer linear features and meandering channels (Figure 1, Panel D).

In addition to its compact circle index, each feature was characterized by its area, perimeter, compact rectangle index, and fractal dimension. Within each feature, we summarized ancillary layers produced by *r.geomorphon* including width, extend, azimuth, elongation, variance, range, exposition, and intensity. Lateral width of the channel-like feature was computed for 5-m longitudinal zones as twice the maximum distance from each bank. Geomorphon intensity and range values were subsequently converted into bank heights using a custom protocol developed by M. Baker. Additional descriptors summarized in this manner included elevation, slope, and curvature values derived from the 10-m DEM as well as all similar ancillary geomorphon layers from the underlying 10-m valley-extent geomorphons. For each descriptor, we retained zonal mean, minimum, maximum, range, standard deviation, and variance, and the coefficient of variation was calculated for each feature using its mean and standard deviation. Subsequently, each channel-like feature was buffered by 10 meters and the buffers were used to summarize surrounding land cover proportions from 1-m resolution land cover (Chesapeake Bay Program, 2023). When completed, the feature descriptor table included 151 covariates (as described in Appendix 2) for each channel-like feature.

We fed the feature descriptor table into a Random Forest (RF) model trained using manual interpretation and classification of 89,136 features from the Lower Susquehanna HUC8 watershed. The Lower Susquehanna was notable for its diverse mosaic of land uses and physiography including forests, mountains, piedmont foothills, agricultural lands, karst features, urban centers, major dams, and the Susquehanna River itself. RF was employed to predict probability of membership in one of eight semantic classes: (1) Other; (2) Gully; (3) Roadside Ditch; (4) Agricultural Ditch; (5) Floodplain Depression; (6) Unknown feature; (7) Wetland; and (8) Stream Channel. Eight classes were selected (i.e., rather than a binary classification) to preserve flexibility in including or excluding different types of features in the resulting linework, and also to better understand sources of misclassification error for future implementations.

The RF model assigned eight semantic class probabilities for each feature. Out-of-bag (OOB) accuracy was 62.7%, indicating substantial confusion among classes, yet with a stream-class specific accuracy (i.e., 1 - Omission or *specificity*; Metes et al. 2022) of 88.9% and non-stream class specificity ranging from 50.9% to 2.5%. OOB *reliability* (i.e., 1 - Commission) was 69.4% for the stream class, and ranged from

100% to 46.5% for the other classes. A binary stream-nonstream version of the same data produced an OOB accuracy of 80.1% with a specificity of 75.7% and a reliability of 82% using the same data. Validation efforts withholding 30% of the observations suggested the models accounted for similar accuracy patterns for all eight feature classes including the stream class in both the training and validation subsets. Channel-like features were initially retained for the channel skeleton using probabilistic RF predictions of the stream class alone. Subsequent selections were based on additional interpretations using the decision tree presented in Figure 2 below. When predicted probabilities were definitive, there was little need for subsequent interpretation. However, in cases where the RF model failed to distinguish features clearly based on their descriptors (i.e., highest ranked probabilities were within a few percent or all probabilities were <0.30 for all classes), additional decision criteria were deemed necessary.

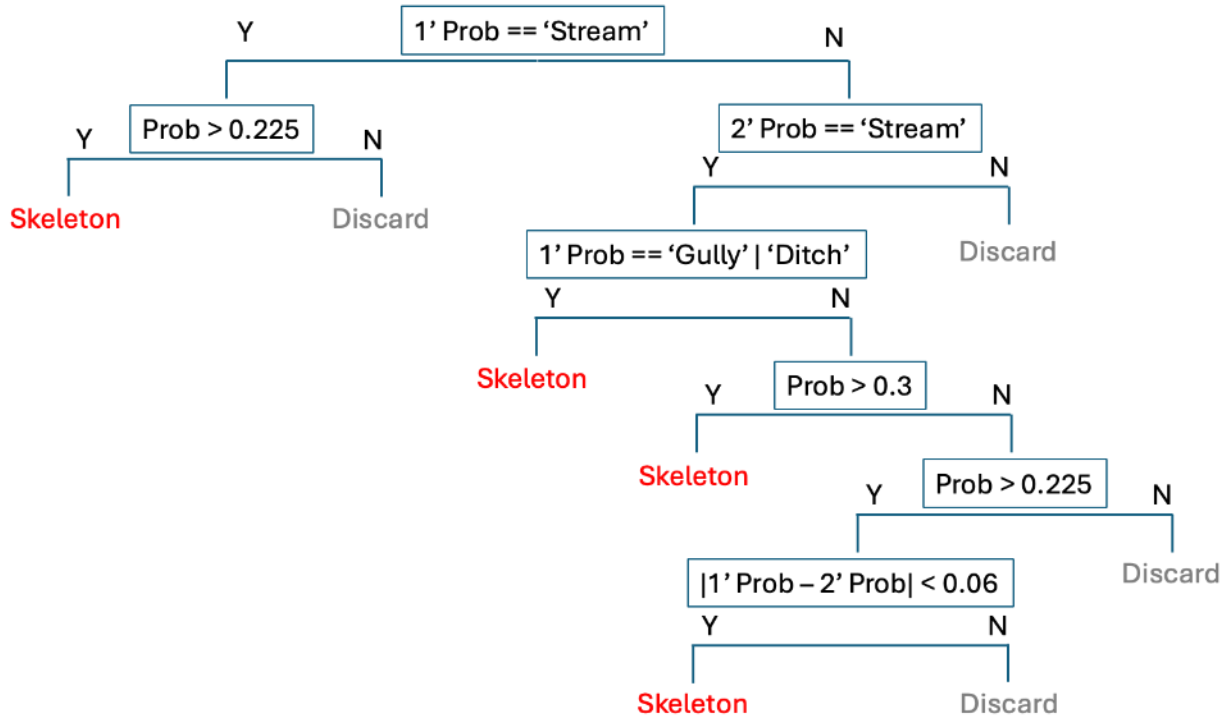


Figure 2: Decision criteria used to select channel-like geomorphon features based on a random forest model classification of channel types.

Because they were morphologically similar to streams and often functioned as part of the channel network, gullies and ditch classes were retained in the channel skeleton when streams (versus other similar classes) were the second highest class probability, or when the probability of streams was second to another class, but the secondary probability was greater than 0.3. Finally, when neither primary or secondary class probabilities were particularly definitive (and secondary probability was 'stream'), but both were similar in magnitude, we retained those features as likely channels. Given the rates of stream commission inherent in the RF model, the additional criteria were likely to result in additional commission, but would also better reflect an overall objective of mapping as many connected channels

as was feasible. Retained features were passed as the channel skeleton to subsequent processing for linear networking.

Linking the Channel Skeleton within a HUC12

Skeleton features were linked across each HUC12 watershed using the Priority Flood (PF) algorithm (Barnes et al. 2014). As opposed to conventional steepest-descent flow routing that follows the steepest slope downhill to connect pixels in a raster network, PF 'floods' DEMs inwards from their edges in an order determined by increasing elevation in a priority queue. Thus, PF works along a 'least cost path' from the lowest elevations upwards as opposed to downwards from the highest elevations. An advantage of the PF approach in LiDAR DEMs is that it is able to link all pixels in a flow direction surface without modifying the DEM, obviating any *a priori* need to 'fill sinks' or 'breach' other obstructions in the terrain. A disadvantage of the approach is that such connections are not always direct, can cross internal divides, and thus might incorrectly represent flow pathways over longer distances. We employed a weighted PF that lowered the elevation cost along direct routes between skeleton features to improve routing of connectors and halted flooding as soon as all skeleton features were connected within a network.

To connect the channel skeleton across terrain, we employed the weighted PF to create a flow direction layer, extending from the outlet(s) along the HUC12 boundary to all interior cells. Starting from headwater cells without upslope neighbors, the flow direction surface was followed towards the outlets, assigning paths as "streams" only when cells occurred within the channel skeleton. Within a skeleton feature, PF routing remained at the lowest portion of the DEM (i.e., the thalweg) as long as the PF entered the feature at the lowest point along its perimeter and a continuous channel was visible within the feature. Outside of skeleton features, PF routing connected skeleton features by linking them across terrain, open water, and roads through culverts, bridges, or other obstructions. Connections across terrain closely followed elevation gradients, yet flow paths over wetlands or open water would sometimes wander or follow one shoreline or the other due to hydro-flattening performed on the input DEMs. Modification of DEMs over water could lead to errors in flow routing along water margins, therefore we developed a custom script to automatically modify open water DEM surfaces so that they were "etched" with a reasonable approximation of a centerline flow pathway. This type of routing was particularly important when one end of an open water feature was bounded by a dam.

Open water regions of large tidal rivers, wetlands, and estuaries were typically represented in the DEMs as a flat surface without a geomorphological thalweg or any topographic signal. Additionally, the Chesapeake Bay watershed has numerous dams and reservoirs, obstructing or obfuscating flow routing. We employed 1-m land cover and geomorphons to identify open water areas. Cells classified as water and any adjacent cells that were classed as either barren or flat were designated as 'open water' for the purposes of etched flow routing. Subsequently, we identified dams by analyzing the lowest elevation within a 30-m buffer of a waterbody. If this buffer's minimum elevation was more than 2 meters lower

than the waterbody's lowest elevation, we considered this a likely indicator of a dam adjacent to the open water area.

After open water areas and dams were identified, we modified the elevation of the open water area by lowering the elevation of the potential dam region to route PF flow through the feature. For the open water area, we used distance from the water boundary to modify its elevation and identify its centerline. This modification ensured that PF routed flow along the centerline and that tributaries connected to the centerline along a direct pathway. For open water regions that contained dams, we adjusted the elevation of the buffer zone surrounding the open water. This adjustment involves lowering the elevation of the buffer area to match either the elevation of its own cells or the elevation of the nearest cell on the buffer's boundary, whichever was lower.

Although they did not obstruct PF connections, roads and other ridge-like features could cause PF routing to follow circuitous routes. In extreme cases when obstructions were extensive (e.g., multi-lane highways, cloverleafs, airports) or spanned the entire extent of the valley network, PF sometimes failed to connect all channel skeletons correctly. In large order rivers or reservoirs that crossed HUC boundaries, channel spanning bridges could cause flow from upstream areas to be directed upslope toward the HUC12 perimeter at a lower cost than crossing the span. To facilitate correct connections of channel skeletons using PF, we developed a two-stage process. In the first stage, a proximity-weighted DEM was used as a PF cost surface to produce an auxiliary flowline. This auxiliary flowpath illustrated potential relationships between skeleton features upstream and downstream of roads or ridges. The first stage thus resulted in directed acyclic stream networks connecting the stream channels and leading to downslope outlets on the HUC12 perimeter. However, crossings over roads were sometimes less than ideal, due to the presence of roadside ditch features in the skeleton or the absence of nearby channel-like features altogether.

The second stage of PF connection was concerned solely with refining road and impervious crossings from the proximity-weighted DEM. Here, our procedure considered the three closest skeleton features in sequence moving upstream and downstream of every crossing. Road or impervious crossings falling within 70 m of one another without intervening skeleton features were treated as a single entity for the purposes of sequencing. For upstream features, we established a line initiating from the lowest elevation point among upstream features falling within 120 m of the crossing to account for the fact that many proximate features included roadside ditches, so that the closest feature to the road or impervious surface was not always a channel pour point (i.e., the upstream entry of a bridge overpass, a culvert, or sewer drain). Candidate downstream features included three features in sequence below the crossing. Each feature was evaluated in turn as to whether half of its area fell within 20 m of the road or 10 m of impervious cover, where any single feature failing to meet this criterion in sequence would eliminate further features from consideration. The lowest 10th percentile of elevation in downstream candidate features was also required to be lower than the upstream initiation point to ensure at least the possibility of downslope flow, whereas the features were required to be within 120 m of the

auxiliary crossing to limit potential change in crossing location. For candidate features meeting all three criteria (proximity to the road/impervious surface, downslope flow potential, proximity to auxiliary crossing), the feature with the lowest elevation was selected as the downstream end of the crossing line. If a crossing did not have both an upstream and downstream feature, the auxiliary proximity-weighted PF solution was used. Once the crossing was constructed, it was burned into the original denoised DEM and PF connections were reinitiated, resulting in directed acyclic stream networks connecting skeleton features and leading them downslope to outlets on the HUC12 perimeter.

Validating Channel Networks

After generating stream networks at the HUC12 level, we merged flowlines to the HUC8 level. We assessed the validity of flow direction within the networks using topological analysis of flow lines as regional directed graphs (RDGs). Loops in the RDGs suggested errors of flow direction within the networks. These issues were rare but could stem from a failure to identify substantial obstacles to PF routing, such as a large dam or other obstructions within the terrain, and were subsequently corrected. Next, we evaluated the number of outlets associated with each HUC12. Most HUCs were expected to contain only a single outlet, but coastal HUCs could conceivably contain multiple independent drainages, all with their own outlets. And, because of the manner and the map scale at which they were delineated, HUCs at major confluences could have flowpaths that weaved in and out of adjacent, tributary HUCs before confluence with a downstream segment. Therefore, HUCs with multiple outlets or multiple networks were flagged for further consideration.

Across larger basin areas, we used a similar topological analysis of RDGs to connect sequential HUC8 flow lines. In this case, most of the connections were within larger, wider river channels. Here the challenge was often that HUC boundaries did not always respect the boundaries of large tributary channels and because of the flattened open water surface, it was often true that upstream flowlines would not cross HUC boundaries at locations within the channel near downstream HUC flowline initiation points. We developed a custom script to locate and connect lines within the channel, but in some cases a manual linkage was required. At the conclusion of this network validation it was possible to work across the connected flowlines to develop network attributes.

Attribution of Stream Reaches

Following completion of each HUC12 network, attributes were assigned to each reach, defined here as a polyline between two confluences or one confluence and either an end node or the HUC12 boundary. As a polyline attribute, the length of each reach was determined. During validation of the channel network, we computed several network metrics including Strahler stream order for each reach (strOrder), link number (i.e. Shreve magnitude; linkNum), and downstream-link number (dlinkNum) (i.e., the link number of the next downstream segment; Osborne and Wiley, 1992).

As a line segment, each reach was assigned elevation values from the 1-m DEM including distributional moments (elevMin, elevMax, elevRange, elevSd, elevAvg, elevMed, elevPct25, elevPct75) extracted from the overlay of the polyline with the DEM surface. Because each reach traced through a skeleton feature or linked two or more features between confluences, most lines were composed of both channel features and connectors. Thus, we calculated the proportion of each reach length formed by connectors as an attribute (cnctR). We also describe the proportion of connector length that crosses roads (rdCnctR), water (wtrCnctR), and other surfaces (othCnctR).

Within each skeleton feature, we measured channel width as twice the maximum distance from the centerline to the edge of the polygon (the distance from centerline to edge is presumed to be half the overall channel width, thus the value is doubled), summarized every five meters in the longitudinal direction. The distribution of widths was extracted as the reach polyline passed over each longitudinal zone, and its values summarized as distributional moments (wdMin, wdMax, wdRange, wdSd, wdAvg, wdMed, wdPct25, wdPct75). Similarly, bank heights were calculated as the difference between the elevation of a focal pixel and the average height of all higher features in the geomorphon visibility neighborhood. Each pixel within a skeleton feature had independent estimates of bank height, and these were summarized along each reach polyline as distributional moments (bkhtMin, bkhtMax, bkhtRange, bkhtSd, bkhtAvg, bkhtMed, bkhtPct25, bkhtPct75).

Mapping Agricultural and Roadside Ditches

Similar to our classification and selection of channel-like features for inclusion in the channel skeleton, we employed supervised random forest (RF) models to select channel-like features that were most likely to be agricultural and roadside ditches. However, in developing these models, we began by collecting additional, class-specific training data in different physiographies, augmenting each model with additional predictor variables beyond what was provided to the skeleton predictions. This process began with the manual classification of 13,747 agricultural ditches (Figure 3) across five HUC12 watersheds and 11,904 roadside ditches (Figure 4) across five different HUC12 watersheds to serve as training data.

The five watersheds selected for agricultural ditch training were chosen based on the presence of agricultural ditches in flat terrain. However, we were also careful to include distinct physiographic contexts in which agricultural ditches may be encountered (e.g. coastal plain, glaciated Appalachian plateau, and other poorly drained areas). These HUC12 watersheds were located within the Upper Susquehanna, Chenango, Lower Potomac, Pamunkey, and Pokomoke-Western Lower Delmarva HUC8 sub basins. Features selected during the manual interpretation of training data were linear, uniform in width, minimally meandering, and located within apparent agricultural areas as interpreted from aerial imagery. Certain types of features were specifically excluded from the training dataset including ditches in timber harvest areas, roadside ditches in agricultural areas (these were included in training data for the roadside ditch model), and large canals. In addition to the 151 covariates used in the stream-modeling RF described previously, the agricultural ditch-modeling RF incorporated a categorical variable

indicating whether the feature was located in 1) the Delmarva peninsula, 2) the glaciated portion of the Chesapeake watershed, or 3) any other location within the Chesapeake watershed. This variable accounted for the greater frequency of agricultural ditches (and, by extension, the increased likelihood that a feature may be an agricultural ditch) and the distinct contexts represented by the Delmarva and glaciated regions.

Figure 3: Channel polygons manually labeled as agricultural ditches for RF training (blue) and polygons not labeled as agricultural ditches (pink). Screenshot from Pocomoke watershed in Delmarva peninsula.

The five watersheds selected for roadside ditch training were chosen based on their inclusion of a variety of road types (e.g. large interstate highways, mid-sized state/county highways, and small residential streets) and representation of different physiographic contexts. These HUC12 watersheds were located within the Owego-Wappasening, Lower West Branch Susquehanna, South Fork Shenandoah, Middle Potomac-Anacostia-Occoquan, and Nanticoke HUC8 sub basins. Features selected during manual interpretation of roadside ditch training data were linear, uniform in width, and parallel or nearly parallel to roads, including ditches running parallel to roads in an otherwise agricultural setting. Features specifically excluded from the roadside ditch training dataset include ditches along railroads and shallow, paved street gutters. In addition to the 151 covariates used in the stream-modeling RF described above, the roadside ditch-modeling RF included five additional variables summarizing the distance of the feature to the nearest road, including mean distance, standard deviation of distance, minimum distance, maximum distance, and distance range. These variables were computed as zonal summaries of discrete polygons such that linear, uniform polygons running parallel

to roads (e.g. ditch polygons) had smaller standard deviations and ranges compared to meandering polygons of varying width (e.g. streams flowing alongside roads).

Figure 4: Channel polygons manually labeled as roadside ditches for RF training (blue) and polygons not labeled as roadside ditches (red). Screenshot from Lower West Branch Susquehanna watershed.

Using the manually-interpreted training datasets described above, separate RF models were developed to identify agricultural ditches and roadside ditches. The most influential variables in the agricultural ditch RF model were: number of pixels classified as “low vegetation” in the high-resolution land cover surrounding each feature and the average width of each feature. The most influential variables in the roadside ditch RF model were: number of pixels classified as “road” in the high-resolution land cover data surrounding each feature, average distance to nearest road, and fractal dimension – a quantification of the geometric complexity of the feature (or lack thereof).

Although no formal accuracy assessment was conducted on the ditch predictions, RF model performance for agricultural and roadside ditch mapping was evaluated using OOB and cross-validated error estimates, as in the skeleton models described above but over a broader regional landscape. OOB accuracy for the agricultural ditch model was 95.1% with a specificity of 81.2% and a reliability of 94.4%. Cross-validation of the agricultural ditch model showed an overall accuracy of 94.9% with 95% confidence between 94.2% and 95.6%. OOB accuracy for the roadside ditch model was 95.8% with a specificity of 82.2% and a reliability of 87.9%. Cross validation of the roadside ditch model showed an

overall accuracy of 96.0% with 95% confidence intervals of 95.3% and 96.7%. We concluded that these models appeared to be generally effective when provided with additional predictors and regional variability, in sharp contrast to the success of the skeleton predictions, which showed class specificities of 42.7% and 25.2% for roadside and agricultural ditches, respectively. Despite their apparent successes, validation of these predictions is recommended as a best practice.

Accuracy Assessment Protocol

A thorough accuracy assessment was carried out to characterize the efficacy with which our automated method correctly mapped streams. It is perhaps worth noting up front that there have been very few examples of such assessments in the literature, simply because the relevant topographic data has not existed for very long and because few mapping efforts are so extensive as to defy reasonable field validation. Because of the substantial geographic extent of our data, we developed an accuracy assessment approach that would fairly characterize the automatically-generated stream maps using remotely-sensed data. Our accuracy assessment was carried out in seven HUC8 watersheds representing a diverse mix of physiographic and land use contexts: Gunpowder-Patapsco, Lower West Branch Susquehanna, Patuxent, Upper Rappahannock, Upper James, Choptank, and Chenango. External data in the form of field-collected stream measurements or other terrain-based stream mapping products from independent research efforts also exist in a number of these watersheds and may be used for subsequent evaluation of the stream dataset produced in this project.

Because a primary goal of this stream mapping effort was to accurately identify headwater channels, specifically, and depict their interconnectedness to larger waterways, the accuracy assessment protocol was developed to thoroughly evaluate the stream lines that represent an extension of existing stream maps, namely the National Hydrography Dataset (NHD). Therefore, we stratified the stream lines based on their degree of overlap with flowlines and water features mapped in the NHDPlus HR (USGS, 2024). NHD flowlines, NHD areas, and NHD waterbodies were combined and buffered by their published lateral accuracy of 40 feet. The fractional length of our stream lines intersecting this buffered version of NHD was tabulated and stream lines were put into four categories: those overlapping the buffered NHD for <25% of their length, 25-50% of their length, 50-75% of their length, and >90% of their length (Table 1). The majority of effort was focused on assessing stream lines overlapping the buffered NHD for <25% of their length, as these were thought to represent extensions of the existing hydrographic network and areas of the greatest uncertainty regarding accuracy. Many lines with 25%-50% NHD overlap were also extensions of the existing network (Figure 5), but as they typically represented ~10% or less of the network, extended headward up the same valley, or represented the same stream with different lateral placements, we deemed them less uncertain than the < 25% category. Stream lines with >25% NHD overlap were thus presumed to be more accurate and were evaluated using an alternate approach.

Table 1: Distribution of stream reaches overlapping buffered NHD for <25% of their length, 25-50% of their length, 50-75% of their length, and >90% of their length within the seven HUC8 watersheds selected for the accuracy assessment.

HUC8	Total reaches	<25% overlap (count)	25-50% overlap (count)	50-75% overlap (count)	>90% overlap (count)	<25% overlap (percent)	25-50% overlap (percent)	50-75% overlap (percent)	>90% overlap (percent)
Gunpowder	43,193	24,037	3,950	4,096	8,119	0.56	0.09	0.09	0.19
Lower West Susq.	45,492	21,715	4,667	4,954	10,508	0.48	0.10	0.11	0.23
Patuxent	68,753	47,728	5,373	4,530	8,546	0.69	0.08	0.07	0.12
Upper Rappa.	78,494	44,016	5,600	5,956	17,784	0.56	0.07	0.08	0.23
Upper James	135,167	77,755	9,047	9,984	30,045	0.58	0.07	0.07	0.22
Choptank	17,498	6,048	1,750	1,729	6,922	0.35	0.10	0.10	0.40
Chenango	35,152	15,229	4,098	4,492	8,242	0.43	0.12	0.13	0.23

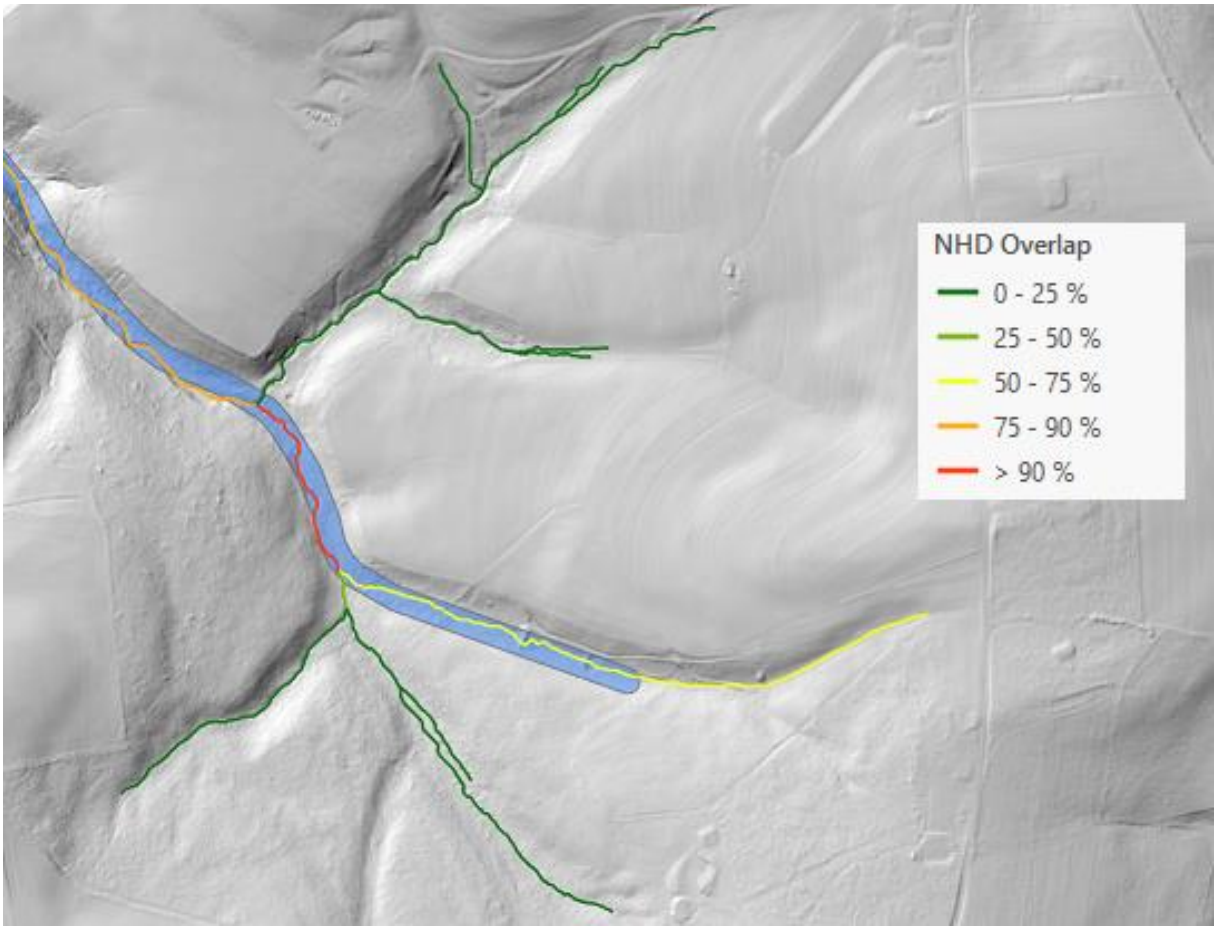


Figure 5: Example of stream stratification based on percentage overlap with buffered NHD (blue). Note the 75-90% class was ultimately not utilized in the accuracy assessment as these reaches represented a comparatively small portion of the network and were only marginally different from the >90% class.

To assess accuracy of streams overlapping the NHD for <25% of their length, a random subset of 1,000 reaches was selected from each of the seven HUC8 watersheds after a sample size of ~984 reaches was determined to be large enough to generate a sufficiently powerful and representative sample for a 95% confidence interval of approximately 2.5% and a user's accuracy of approximately 80% (Stehman and Foody, 2019). Collectively these 7,000 reaches were independently evaluated for accuracy by two reviewers and a third reviewer was consulted when the first two reviewers disagreed on the accuracy determination of a given reach, resulting in a majority rule for all 7,000 reaches. Stream reaches selected for accuracy assessment were evaluated against LiDAR hillshades from the same DEMs that were used to delineate them, as well as aerial imagery available in the ArcGIS Pro basemap gallery. Underlying channel skeleton polygons were also used to aid in the interpretation of accuracy. The following categories were used to describe the accuracy of the <25% NHD-overlapping streams:

- **Accurate** - Reaches were considered accurate when they originate within a channel visible in the LiDAR hillshade and/or aerial imagery and are traced correctly down the length of the channel. Reaches within non-stream conveyance features (e.g. ditches) were considered accurate if they were clearly and directly connected to the overall drainage network. Reaches that connect visible channel features but do not necessarily overlap a visible channel directly (e.g. connectors, culverts, reaches in karst terrain) were considered accurate if they connected the correct features in the correct direction.
- **Class commission** - Reaches were considered class commissions if they originated from a channel-like depression in the DEM that was not a stream or not connected to the overall drainage network (i.e., misclassed channel-like features). Examples of class commissions include detention features with no outlet, swales in agricultural fields, some floodplain depressions, roads and railroads with embankments on either side, and ditches without a clear connection to the drainage network.
- **Commission <25%** - This category describes reaches that originate from a channel skeleton polygon outside of or beyond a visible channel in the hillshade or imagery and persist outside of a channel for less than 25% of their length. A common example is a reach that extends further upslope than the visible stream head, but is otherwise accurate once the line is traced in the stream channel.
- **Commission 25-50%** - Similar to "Commission <25%", this category describes reaches that originate and persist outside of a channel for 25-50% of their length, with the remainder of the reach accurately traced in a channel.
- **Commission >50%** - Reaches that originate and persist outside of channels for greater than 50% of their length. This category also encompasses reaches that are entirely commissions, originating from spurious depressions that are unrelated to the drainage network.

- **Tracing error** - Reaches were considered tracing errors if portions of the line were drawn correctly within channels while other portions leave the channel erroneously or create inappropriate/unlikely connections between channel skeleton features.
- **Headwater omission** - Headwater omissions refer to reaches that *were mapped accurately*, however, the LiDAR hillshade showed visible evidence of unmapped channel(s) upslope of the headward end of the mapped reach. In order to be considered, the omitted channels had to be of similar size and relief as the correctly-mapped channel.
- **Tributary omission** - Similar to "Headwater omission", the "Tributary omission" label refers to reaches that *were mapped accurately* but had visible evidence of one or more unmapped tributaries flowing directly into the correctly-mapped reach that were of similar size and relief as the mapped channel.

Reaches overlapping the buffered NHD for greater than 25% of their length collectively made up a smaller percentage of the overall number of mapped reaches and there was a presumed direct positive correlation between degree of NHD overlap and accuracy. Reaches with increasing degree of NHD overlap were nonetheless examined to validate this presumption. From the same seven HUC8 watersheds that were previously assessed, 150 reaches were randomly selected from three categories: 50 reaches that overlap the buffered NHD for 25-50% of their length, 50 reaches that overlap the buffered NHD for 50-75% of their length, and 50 reaches that overlap the buffered NHD for >90% of their length. Reaches overlapping the NHD for 75-90% of their length were ultimately not assessed as these reaches represented a comparatively small portion of the network and were only marginally different from the >90% class. In total, 1,050 reaches were assessed in addition to the initial 7,000 reaches with <25% NHD overlap. These 1,050 additional reaches were categorized as "Accurate" or "Inaccurate" based on a single reviewer's interpretation of their agreement with visible stream channels in the LiDAR hillshade.

Accuracy Assessment Results

Accuracy results were compiled as a proportion of the 1,000 sampled reaches within each HUC8 watershed, as well as length-weighted proportions, which were arguably a more precise representation (Table 2). The first three columns indicate the watershed, the number of reaches sampled, and the total length of stream reaches sampled in the watershed.

The fourth column shows the percentage of reaches that were classified as "Accurate", "Headwater omission", or "Tributary omission" as a percentage of the total number reaches. As described previously, reaches categorized as headwater or tributary omissions were mapped accurately but included omissions in addition to the correctly-mapped channel; as such, they are counted toward the accurate total.

The fifth column shows a length-weighted accuracy metric. This metric attempts to more fairly represent the actual accuracy of the stream reaches by accounting for the accurate portion of reaches that reviewers determined were only partially inaccurate, namely those reaches classified as partial commissions. Length-weighted accuracy is calculated as the sum of the "Accurate" length, "Headwater omission" length, "Tributary omission" length, 75% of the "Commission <25%" length, and 50% of the "Commission 25-50%" length, divided by the total length of stream reaches in the sample.

The sixth column shows a modification of the length-weighted accuracy metric that demonstrates the impact of tracing errors on overall accuracy. This metric was computed the same as the length-weighted accuracy with the addition of "Tracing error" length in the summed numerator. Because reaches classified as "Tracing error" were correctly *detected* in the LiDAR but traced incorrectly by the routing algorithm, this metric separates the ability to correctly detect channels from the PF ability to trace them (i.e., column 6 - column 5).

The seventh and final column indicates the number of accurately-mapped reaches that included omitted headwaters or tributaries. Because we were unable to assess a true number of omissions (and thus its complement, producer's accuracy), collecting information about omission rates represented a challenge. As a partial solution, we tracked how often our sample included otherwise accurately-mapped reaches with significant portions of unmapped channel upslope of the initiation point (headwater omissions) or significant unmapped tributaries that flowed directly into the randomly sampled reach (tributary omissions).

Table 2: Accuracy assessment of reaches with <25% of their length within 40 feet of NHD water features, reported by count and length.

HUC8	Reach count	Total length (meters)	Accuracy (% by count)	Length-weighted accuracy (% by length)	Length-weighted accuracy, excluding tracing errors (% by length)	Omissions (count)
Gunpowder	1000	212714.80	0.64	0.66	0.77	41
Lower West Susq	1000	227932.98	0.68	0.65	0.82	51
Patuxent	1000	158566.04	0.76	0.75	0.83	10
Upper Rapp	1000	201450.24	0.67	0.70	0.81	34
Upper James	1000	184229.73	0.78	0.82	0.86	46
Choptank	1000	294935.38	0.75	0.68	0.89	155
Chenango	1000	235778.39	0.65	0.64	0.80	80
Average:		216515.37	0.70	0.70	0.82	60
Median:		212714.80	0.68	0.68	0.82	46

Collectively, our assessment indicated an average reliability (or user's accuracy) of approximately 70% to 82%, depending on whether tracing errors were considered commissions. There were notable differences among physiographic settings, with some Piedmont and Delmarva watersheds more

accurate than glaciated or Appalachian settings. However, in length-adjusted estimates, the Delmarva and Lower West Susquehanna maps were notably less accurate than any other setting when compared to count-based estimates, even though many of these were apparently tracing errors of real flow pathways. For other settings, length adjustments increased or did not have a substantial impact (i.e., > 2.5%) on accuracy estimates. Further, because detection errors were likely to be more prevalent in headwaters, whereas tracing errors may occur elsewhere in the network, it was worth examining the distribution of errors over stream orders.

Table 3 and Table 4, below, show the results of our principle accuracy assessment as a function of stream order. Table 3 shows the fraction of reaches that were classified as "Accurate" or one of the specific error types, as a proportion of the total number of reaches per order. Table 4 shows the proportion of accurate stream length, accounting for the accurate portion of reaches that were only partially inaccurate, per order. Table 4 also presents a length-weighted accuracy metric that excludes "Tracing errors" to distinguish the accuracy of channel detection from that of channel tracing.

Table 3: Accuracy assessment of reaches with <25% of their length within 40 feet of NHD water features, reported by count over Strahler stream orders.

Watershed	Stream order	Reach count	Accurate (% by count)	Commission <25% (% by count)	Commission 25-50% (% by count)	Commission >50% (% by count)	Class Commission (% by count)	Tracing Error (% by count)	Headwater Omission (% by count)	Tributary Omission (% by count)
Gunpowder	1	719	0.56	0.04	0.04	0.18	0.11	0.06	0.03	0.02
Gunpowder	2	229	0.81	0.00	0.00	0.04	0.05	0.09	0.00	0.01
Gunpowder	3	37	0.86	0.00	0.00	0.03	0.03	0.08	0.00	0.03
Gunpowder	4	10	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gunpowder	5	4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gunpowder	6	1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lower West Susq	1	727	0.63	0.01	0.02	0.18	0.07	0.08	0.05	0.02
Lower West Susq	2	192	0.75	0.01	0.01	0.07	0.03	0.14	0.00	0.01
Lower West Susq	3	53	0.87	0.00	0.00	0.00	0.02	0.11	0.00	0.00
Lower West Susq	4	20	0.95	0.00	0.00	0.00	0.05	0.00	0.00	0.00
Lower West Susq	5	7	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lower West Susq	6	1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Patuxent	1	600	0.64	0.04	0.07	0.18	0.03	0.04	0.01	0.01
Patuxent	2	252	0.92	0.00	0.00	0.02	0.01	0.05	0.00	0.00
Patuxent	3	103	0.92	0.00	0.00	0.01	0.01	0.06	0.00	0.01
Patuxent	4	32	0.94	0.00	0.00	0.00	0.00	0.06	0.00	0.00
Patuxent	5	13	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Upper Rappahannock	1	740	0.60	0.03	0.06	0.24	0.02	0.04	0.03	0.01
Upper Rappahannock	2	224	0.87	0.00	0.00	0.05	0.00	0.08	0.00	0.00
Upper Rappahannock	3	31	0.94	0.00	0.00	0.00	0.00	0.06	0.00	0.00
Upper Rappahannock	4	3	0.33	0.00	0.00	0.33	0.00	0.33	0.00	0.00
Upper Rappahannock	5	2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Upper James	1	706	0.73	0.03	0.02	0.17	0.03	0.02	0.03	0.03
Upper James	2	229	0.90	0.00	0.00	0.03	0.02	0.04	0.00	0.01
Upper James	3	46	0.96	0.00	0.00	0.00	0.00	0.04	0.00	0.00
Upper James	4	13	0.92	0.00	0.00	0.00	0.00	0.08	0.00	0.08
Upper James	5	4	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.25
Upper James	6	2	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.00
Choptank	1	756	0.72	0.01	0.02	0.12	0.02	0.11	0.15	0.04
Choptank	2	178	0.79	0.01	0.01	0.03	0.03	0.12	0.00	0.02
Choptank	3	42	0.90	0.00	0.00	0.00	0.00	0.10	0.00	0.05
Choptank	4	22	0.91	0.00	0.00	0.00	0.00	0.09	0.00	0.05
Choptank	5	2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Chenango	1	704	0.61	0.01	0.02	0.23	0.05	0.09	0.08	0.02
Chenango	2	195	0.72	0.00	0.00	0.12	0.02	0.14	0.00	0.02
Chenango	3	53	0.77	0.00	0.00	0.08	0.00	0.15	0.00	0.02
Chenango	4	33	0.88	0.00	0.00	0.00	0.03	0.09	0.00	0.00
Chenango	5	12	0.92	0.00	0.00	0.00	0.00	0.08	0.00	0.00
Chenango	6	2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Chenango	7	1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4: Accuracy assessment of reaches with <25% of their length within 40 feet of NHD water features, reported by length over Strahler stream orders.

Watershed	Stream order	Total length (meters)	Length-weighted accuracy (% by length)	Length-weighted accuracy, excluding tracing errors (% by length)
Gunpowder	1	153301.89	0.62	0.72
Gunpowder	2	50922.43	0.75	0.91
Gunpowder	3	6319.85	0.75	0.94
Gunpowder	4	1283.13	1.00	1.00
Gunpowder	5	631.59	1.00	1.00

Gunpowder	6	255.91	1.00	1.00
Lower West Susq	1	160002.47	0.65	0.78
Lower West Susq	2	50054.95	0.64	0.88
Lower West Susq	3	14019.07	0.73	0.99
Lower West Susq	4	2865.01	0.93	0.93
Lower West Susq	5	686.57	1.00	1.00
Lower West Susq	6	304.91	1.00	1.00
Patuxent	1	106252.00	0.69	0.78
Patuxent	2	33950.46	0.89	0.95
Patuxent	3	13478.27	0.84	0.90
Patuxent	4	3182.45	0.96	1.00
Patuxent	5	1702.87	1.00	1.00
Upper Rappahannock	1	154186.17	0.68	0.76
Upper Rappahannock	2	41473.92	0.77	0.95
Upper Rappahannock	3	4823.55	0.88	1.00
Upper Rappahannock	4	485.17	0.16	0.72
Upper Rappahannock	5	481.42	1.00	1.00
Upper James	1	134408.95	0.79	0.81
Upper James	2	38655.12	0.91	0.96
Upper James	3	8013.74	0.90	1.00
Upper James	4	1818.24	0.83	1.00
Upper James	5	1127.74	0.77	1.00
Upper James	6	205.95	0.39	1.00
Choptank	1	224272.72	0.66	0.87
Choptank	2	55504.88	0.68	0.93
Choptank	3	9734.21	0.91	1.00
Choptank	4	5272.88	0.87	1.00
Choptank	5	150.68	1.00	1.00
Chenango	1	164167.65	0.65	0.76
Chenango	2	49072.91	0.60	0.86
Chenango	3	12749.30	0.71	0.94

Chenango	4	6965.66	0.70	0.96
Chenango	5	2052.52	0.72	1.00
Chenango	6	591.51	1.00	1.00
Chenango	7	178.82	1.00	1.00

Patterns in Table 3 and Table 4 suggest accuracy rates were distinctly lower in first order channels and improved rapidly with increasing stream order across HUCs and regions. Improvements were largely a result of the lack of commission errors in larger-order (>2) channels, whereas tracing errors appeared marginally more common in second order and larger channels. These patterns differ substantially with physiographic and land use settings, with greater errors associated with intensive land use contexts, particularly those that include major transportation corridors (e.g., interstate highways, major railways, airports). Lower apparent accuracy was possible at higher orders in HUCs with fewer large channels in the sample. The overall implication was that low order channels showed the highest rates of commission, unsurprisingly. However, these results indicated lower commission rates than reported for field-surveyed Piedmont channel heads by Metes et al. (2022) using different techniques on similar lidar DEMs. While not a directly comparable analysis, their results provide important context for how challenging avoiding commission in terrain based delineation can be. Short of pruning the network, our finding suggested that any reduction of commission errors in headwaters should dramatically increase map accuracy. We thus assessed errors in reaches with more substantial overlap to verify this interpretation.

We conducted a secondary accuracy assessment on streams that overlapped with a buffered version of NHD streams for proportions of their length >25% (Table 5). The results of this assessment are a complement to those presented in Table 3 and Table 4 and confirm that accuracy rates improved as the degree of NHD overlap increased, much like accuracy rates improved with increasing stream order. This pattern was more or less expected as streamlines exhibiting large fractions of their length within 40 ft of an NHD feature were likely to represent increasing alignment with previously mapped hydrographic features. For these channels, errors primarily occurred as a result of commission in floodplains or tracing problems. Because many floodplain drainage channels were morphologically identical to streams, they posed a challenge for our geomorphic classifier. Many, but not all, of these commissions were eliminated by considering their internal relative gradient (i.e., generally low and similar across portions of their network upstream of the mainstem connection), with a large dlink number (because they are tributaries of rivers large enough to have channeled floodplains) and contributing drainage area (generally small as floodplain drainages often have few direct upslope contributions) as filtering criteria, but this did not resolve all of the issues. On the other hand, tracing errors often occurred when river corridors contained multiple braided channels or floodplain drainages where PF routing selected low-cost pathways that did not respect channel size. Because many large channels were hydro-flattened within the lidar DEMs, they provided limited topographic information for routing. These errors were initially less apparent until we had removed floodplain commissions.

For channels with lower NHD overlap, errors included the two types mentioned above, as well as the types of commissions depicted in the <25% assessment above. This occurred because channels with 25-50% overlap tended to represent a combination of aligned channels and segments that proceeded upslope into headwater valleys without any corresponding NHD features. For such likely commissions, further refinement of the skeleton should reduce their frequency.

Table 5: Accuracy assessment of reaches with >25% of their length within 40 feet of NHD water features, reported by count and length.

HUC8	25-50% NHD overlap			50-75% NHD overlap			>90% NHD overlap		
	Total reaches	Accurate (% by count)	Accurate (% by length)	Total reaches	Accurate (% by count)	Accurate (% by length)	Total reaches	Accurate (% by count)	Accurate (% by length)
Gunpowder	50	0.74	0.78	50	0.90	0.93	50	0.82	0.92
Lower West Susq	50	0.66	0.82	50	0.88	0.97	50	0.88	0.96
Patuxent	50	0.72	0.83	50	0.90	0.89	50	0.90	0.98
Upper Rapp	50	0.68	0.75	50	0.82	0.81	50	0.98	1.00
Upper James	50	0.64	0.73	50	0.84	0.92	50	0.86	0.94
Choptank	50	0.74	0.67	50	0.82	0.81	50	0.90	0.90
Chenango	50	0.78	0.81	50	0.76	0.83	50	0.80	0.90
Average:		0.71	0.77		0.85	0.88		0.88	0.94
Median:		0.72	0.78		0.84	0.89		0.88	0.94

Overall, our assessment indicated a reasonably accurate map that could nonetheless benefit from additional improvements at several key points in the automated workflow. In larger order systems, it was apparent that further work on routing within and around larger order channels as well as reducing the inclusion of floodplain drainage channels would likely substantially reduce errors and increase general agreement with the NHD in excess of 90%. On the other hand, precise agreement with existing maps was neither expected nor desired in most cases. In lower order systems, feature detection results were adequate for an automated algorithm, but our assessment suggested that some small refinements of the feature classification and selection process could produce large reductions in commissions. Further, regional variation in assessment accuracy revealed the physiographic basis for many tracing errors and provided some indication of how they might be addressed.

Upper Rappahannock Field Validation

As a part of a separate grant funded by National Fish and Wildlife Foundation, non-profit group Friends of the Rappahannock (https://riverfriends.org/) conducted field validation of a subset of mapped reaches in the Upper Rappahannock watershed in collaboration with our group. A stratified random sample of 200 stream reaches was selected from the Upper Rappahannock watershed to guide their efforts. The distribution of stream orders in the sample matched that of the watershed map as a whole,

i.e. 50% of streams in the Upper Rappahannock watershed were first order streams, 23% were second order, and 12% were third order, thus 50%, 23%, and 12% of streams in the sample were first, second, and third order, respectively. The sample size was intentionally large enough to allow ample options in the event the field crew encountered access limitations at any of the 50 planned site visits. Ultimately, the field crew visited 21 of the randomly selected sites and validated the presence or absence of a stream channel, indicated whether there was flowing water, standing water, or no water present in the channel, and collected measurements of channel dimensions at three transects along each channel. The results of their field data collection effort are presented in Table 6.

Table 6: Verification data on channel presence, flow presence, and channel dimensions collected in the field by Friends of the Rappahannock.

Reach ID	Stream order	Field-measured bank height (meters)	Field-measured width (meters)	Stream channel present	Flow observation
4023	1	0.53	12.11	yes	flowing water
4845	1	0.57	4.8	yes	standing water
20877	1	1.1	6.67	no	no water
22001	1	0.22	1.46	yes	no water
23990	1	0.53	4.89	yes	flowing water
37786	1	0.16	1.23	yes	no water
39015	1	1.1	1.27	yes	no water
46998	1	0.44	3.54	no	no water
59932	1	0.39	3.6	yes	flowing water
62698	1	0.23	1.1	yes	no water
2325	2	0.38	1.37	yes	standing water
24672	2	0.52	1.01	no	no water
24745	2	0.3	3.36	yes	flowing water
46772	2	0.3	1.23	yes	standing water
174	3	0.41	2.75	yes	flowing water
54632	3	0.45	2.19	yes	no water
34178	4	0.54	3.29	yes	flowing water
46216	4	0.69	4.12	yes	flowing water
56849	5	1.04	9.52	yes	flowing water
80010	5	0.68	10.89	yes	flowing water
2434	7	1.21	9.61	yes	flowing water

Evidence of an ephemeral or perennial stream channel was visible in 18 of the 21 sites visited by Friends of the Rappahannock for a user's field accuracy of 85.7%. Two of the remaining sites featured linear, convergent features that could be described as swales or other nascent drainages yet lacked evidence of a defined stream channel (Reach IDs 20877 and 46998). The last remaining site lacked evidence of a stream channel and featured only subtly convergent topography, described by the field crew as more

akin to a game trail than a stream channel (Reach ID 24627). Photos of the field sites taken by the field crew are included in Appendix 3.

Lessons Learned

During this mapping effort, we encountered many challenges both expected and unexpected. Foremost, the Chesapeake Bay watershed is a huge land area with variable physiography and a long history of many types of anthropogenic terrain modifications that can impact interpretations of terrain-based flow analysis. No automated technique would be immune from error because all would rely on simplification of real-world complexity. Our original automation was really developed with headwater channels in mind and not specifically designed for larger order floodplain or estuarine systems. In producing output, we realized that the nature of our routing, especially in hydro-flattened channels, often made connections across HUC boundaries unexpectedly challenging. We did not initiate this effort with a specific approach for routing across reservoirs or open water bodies but developed one to include these areas in the resulting network. Similarly, use of HUC12 boundaries for parallel processing greatly accelerated automated runtimes, but introduced many complex and unique routing scenarios, especially in tidal systems. We briefly considered defaulting to existing hydrography where it existed in larger order systems, but ultimately decided to attempt a comprehensive approach to enable future improvement.

Although we are confident that our work remains innovative and includes many technical advances and insights, in hindsight we are very aware of efficiencies we have yet to exploit and further opportunities for error reduction. Most opportunities for improvement fall into one of four general categories: (1) improved removal of erroneous channel-like depressions from the channel skeleton to reduce commissions; (2) better retention of features in the channel skeleton to improve continuity and accuracy of flow routing; (3) improvements to tracing routines and road crossing protocols in complex anthropogenic terrains (e.g. major highways, storm sewers, culverts, or ditches); and (4) improved stream attribution especially including channel dimensions. In general, these improvements should include efforts to bring our data products closer to being in line with the 3DHP cartographic standards.

Accuracy assessment revealed that some of the most prevalent sources of error, especially in first-order streams, were those of commission (Table 3). These commissions appeared primarily a result of the inclusion of non-stream features in the channel skeleton (unchannelized valleys, detention features, floodplain channels, incised roadways and trails, other natural and artificial depressions), which were in turn connected to the stream network by routing and tracing algorithms. Improved removal of non-stream features from the potential channel skeleton should reduce the number of commissions in our dataset and could be accomplished in several different ways. A relatively minor change that may produce a noticeable impact on the channel skeleton would be to replace or augment the training data developed in the Lower Susquehanna HUC with features selected from a broader range of geographic contexts.

Although training data used in the initial skeleton selection model contained a variety of land use and physiography, they did not necessarily reflect the breadth of variation in channel features found across the Chesapeake Bay watershed. Moreover, the predictive skill of the RF classifier was likely influenced by the distribution of classes in the training data. For example, the Lower Susquehanna HUC was dominated by 'stream' and 'gully' (>55%) features so the classifier may have appeared more accurate than it really was simply by defaulting to those predominant classes. Ditch classes, corresponding to less than 15% of the total training features, had to be more definitive in order for the classifier to identify them. Further, initial training data were characterized by single individuals within a 5-6 person interpretive team, each solely responsible for classifying separate portions of the HUC8 watershed. This approach made efficient use of time, but may have led to interpretive inconsistencies that would function as noise in resulting predictions. By contrast, the accuracy assessment of the final data was randomly selected over a number of diverse and unique physiographic areas within the Chesapeake watershed and was carried out by three trained interpreters using majority rule to make the final determination for every reach. Even so, we were surprised by how often experienced interpreters could disagree, and such semantic inconsistencies may detract from predictive skill. Thus, one outcome from our assessment processes should be a cleaner set of training data, which of course might be augmented with additional attributes we have since discovered can impact model performance. Either or both may produce more effective feature selection for the channel skeleton.

Other approaches for improved removal of erroneous channel-like features could be accomplished outside of RF modeling by adjusting the way we implement various geomorphic (i.e. valley network filtering) and morphometric filters (i.e. shape index filtering). For example, assessment revealed many headwater omissions in Appalachian HUCs that resulted from the use of an overly limited delineation of the valley network in mountains, so conceptual and practical extensions of the valley network concept might address these omissions. In low-relief landscapes with limited valley extents and frequent ditching, minor discontinuities in topographic ditch signatures could result in the omission of large sections of the drainage. Reconsideration of the notion of a 'valley' in such landscapes may also be necessary.

Shape-based filtering may also remove channel features that nonetheless belong in the network and thus, refining their application may provide another opportunity for improvement. We have learned that although it is true that many non-stream depressions lack an elongate shape, segments of real stream channels are also sometimes depicted by geomorphons as truncated lengths that have oblong or circular shapes, particularly in landscapes with many channel crossings. While removal of non-stream features from the channel skeleton is desirable to reduce commissions in the dataset, if the algorithms implemented for removal of non-stream features inadvertently remove real stream features from the channel skeleton, it can introduce problems in flow routing and tracing procedures. Inadequate continuity in the channel skeleton can alter the apparent cost surface for the PF algorithm, leading to difficulty connecting a given channel skeleton feature to the correct upslope skeleton feature by the most direct topographic route. Therefore, channel segments can represent essential "breadcrumbs"

that enable more accurate routing by preventing tracing errors, which Table 3 suggested was the most common type of error encountered in second-order and larger streams.

Although retention of stream features and more precise removal of non-stream features in the channel skeleton should improve overall flow routing and tracing, complex anthropogenic infrastructure such as multi-lane highway interchanges, bridges, and large dams involving massive rerouting of flow and novel landforms can still pose a challenge for stream tracing and represents an opportunity for improvement in our approach. This signal was apparent in the accuracy of the Gunpowder/Patapsco HUC, where many commissions and tracing errors resulted from terrain modification, redirected drainage, and other infrastructure. Here we have two foci. For delineations specific to the Chesapeake Bay watershed, many routing problems arising from major infrastructure can be readily addressed manually and subsequently incorporated into future iterations of hyper-resolution hydrography once they are identified. However, our goal is to develop approaches that will reduce the need for such manual interventions in the future to accommodate additional infrastructure and other landscapes outside the watershed. Although we observed some greater precision of road crossings as a result of our two-phased approach in highly modified landscapes, the current version remains vulnerable to unusual contexts and may yet produce unintended errors. Before introducing additional algorithmic complexity, it will be important to better understand interplay between the two crossing approaches and the broader accuracy of the skeleton.

Though not an explicit requirement or goal at the outset of this project, continuous estimates detailing channel dimensions and other attributions were an important output of this work. Initial validation data (such as those collected by Friends of the Rappahannock and other independent research efforts by graduate students at UMBC) suggests promise for our approach with a number of important caveats. Initial analysis indicates that automated channel width estimates generated for this dataset are likely to overestimate widths around tributary confluences because they do not necessarily distinguish one contiguous channel from another. Overestimates can also occur in areas where the line-of-sight constraints reach valley walls. Because of how they are delineated, geomorphons developed at the channel extent from a 1-m DEM are unlikely to reliably detect shallow channels less than 2-2.5 m wide. Instead, they are more likely to detect the local valley context as a channel-like feature. If that valley context is a v-shaped depression in an incised gorge or large gully, delineation of the channel-like feature may proceed up the valley wall. If the valley context is a flat-bottomed, u-shaped depression, the delineated feature may represent the valley floor. The extent of these features and their resulting width estimates, computed as twice the maximum distance from each edge for 5-m longitudinal zones along the presumed channel, are critical for interpreting estimates of bank height. The automated bank height estimates derived from line-of-sight constraints employed in this project are more likely to overestimate than underestimate bank heights where the channel's topographic signal is limited and overshadowed by proximate valley walls or infrastructure (e.g., embankments under bridges). However, bank height and width underestimates are likely where large channel features are wider than the channel-extent geomorphon search radius and thus confound local line-of-sight measurements.

Improvements to all of this attribution presents an opportunity to make these data more useful to researchers, planners, and project implementers alike.

Conclusion

In this report, we present an automated approach to stream mapping using high-resolution LiDAR elevation data and describe its implementation across the Chesapeake Bay watershed. Although we describe non-trivial amounts of error in this, our initial attempt at hydrographic production, it is novel as the first automated mapping effort of its kind over such a broad extent and, as such, represents an important advance in hydrography. Our automated stream mapping approach utilizes computer-vision interpretations of terrain to detect channel-like depressions, employs a suite of geomorphic and morphometric filters to isolate depressions for further consideration as stream channels, applies machine learning to produce a probabilistic classification of depressions as streams and other conveyance features, links stream features within HUC12 watersheds and across HUC8 watersheds, and attributes linked stream reaches with information on network position and physical channel properties estimated from DEM data. This effort represents a fundamental shift in the production of hydrography data because it can be produced and reproduced within a fraction of time required for conventional digital mapping (i.e., the entire Chesapeake basin in two weeks or less on UMBC's high-performance computing facility). Our next challenge will be to use this capacity to iteratively improve the accuracy and value-added contributions of future data products. A thorough accuracy assessment revealed an overall user's accuracy of approximately 70-82%, with accuracy rates improving rapidly as stream order increased. We then applied a similar approach to map agricultural and roadside ditches across the basin. Several areas for future research and development presented themselves during the course of this work, primarily related to better removal of non-stream depressions and retention of real stream depressions during filtering, improved tracing processes in complex anthropogenic landscapes, and improved stream attribution including channel dimensions.

Acknowledgements: We thank the UMBC students who helped develop the initial RF training dataset on the Lower Susquehanna: Breanna Byrd, Michelle Chan, Ethan Crookshank, Ioana Draghici, and Nicati Robidoux; UMBC High-Performance Computing Facility staff and graduate students who assisted us in parallelizing our code, and Friends of the Rappahannock staff who field-validated our data.

References

Baker, M. E., Weller, D. E., & Jordan, T. E. (2007). Effects of stream map resolution on measures of riparian buffer distribution and nutrient retention potential. Landscape Ecology, 22(7), 973-992.

Baker, M. E., Saavedra, D., & Norton, M. (2018). *Scope #10: Methodology for developing high-resolution stream and waterbody datasets for the Chesapeake Bay watershed*. Final project report to the Chesapeake Bay Trust. Accessed online at: https://cbtrust.org/wp-content/uploads/2018_Scope10_FinalReport_wAppendix.pdf

Barnes, R., Lehman, C., & Mulla, D. (2014). Priority-flood: An optimal depression-filling and watershed-labeling algorithm for digital elevation models. *Computers & Geosciences*, *62*, 117-127.

Bezerra, M., Baker, M. Palmer, M., & Filoso, S. (2020). Gully formation in headwater catchments under sugarcane agriculture in Brazil. Journal of Environmental Management 110271.
Chesapeake Bay Program, 2023, Chesapeake Bay Land Use and Land Cover Database 2022 Edition: Land Cover: U.S. Geological Survey data release. Developed by the University of Vermont Spatial Analysis Lab, Chesapeake Conservancy, and U.S. Geological Survey, https://doi.org/10.5066/P981GV1L.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. Geoscientific model development, 8(7), 1991-2007.

Hancock, G. R., & Evans, K. G. (2006). Channel head location and characteristics using digital elevation models. Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group, 31(7), 809-824.

Hooshyar, M., Wang, D., Kim, S., Medeiros, S. C., & Hagen, S. C. (2016). Valley and channel networks extraction based on local topographic curvature and k-means clustering of contours. *Water Resources Research*, *52*(10), 8081-8102.

Jasiewicz, J., & Metz, M. (2011). A new GRASS GIS toolkit for Hortonian analysis of drainage networks. *Computers & Geosciences*, *37*(8), 1162-1173.

Jasiewicz, J., & Stepinski, T. F. (2013). Geomorphons—a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, *182*, 147-156.

Julian, J. P., Elmore, A. J., & Guinn, S. M. (2012). Channel head locations in forested watersheds across the mid-Atlantic United States: A physiographic analysis. Geomorphology, 177, 194-203. Kiang, J.E., Flynn, K.M., Zhai, T., Hummel, P., Granato, G. 2018. "SWToolbox: A surface-water tool-box for statistical analysis of streamflow time series" U.S. Geological Survey Techniques and Methods, book 4, chap. A–11, 33 p.

Lashermes, B., Foufoula-Georgiou, E., & Dietrich, W. E. (2007). Channel network extraction from high resolution topography using wavelets. *Geophysical Research Letters*, *34*(23).

Metes, M.J., Jones, D.K., Baker, M.E., Miller, A.J., Hogan, D.M., Loperfido, J.V. and Hopkins, K.G. (2022), Ephemeral Stream Network Extraction from Lidar-Derived Elevation and Topographic Attributes in Urban and Forested Landscapes. J Am Water Resour Assoc, 58: 547-565. https://doi.org/10.1111/1752-1688.13012

Metes, M., Miller, A., Baker, M., Hopkins, K., & Jones, D. (2024). A lidar based method to remotely maps gullying and incision in Maryland Piedmont headwater streams. Geomorphology 109205.

Montgomery, D. R., & Dietrich, W. E. (1988). Where do channels begin? Nature, 336(6196), 232.

Osborne, L.L. and Wiley, M.J., 1992. Influence of tributary spatial position on the structure of warmwater fish communities. Canadian Journal of Fisheries and Aquatic Sciences, 49(4), pp.671-681.

Paredes-Tavares, J., Lopez-Farias, R., Valdez, S.I. and Lamphar, H.S., 2023, September. Valley Classification using Convolutional Neural Network and a Geomorphons Map. In 2023 Mexican International Conference on Computer Science (ENC) (pp. 1-6). IEEE.

Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G., & Dietrich, W. E. (2010). A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths. Journal of Geophysical Research: Earth Surface, 115(F1).

Rodriguez-Iturbe, I., & Rinaldo, A. (2001). Fractal river basins: chance and self-organization. Cambridge University Press.

Stehman, S. V., & Foody, G. M. (2019). Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, *231*, 111199.

Sun, X., Rosin, P. L., Martin, R., & Langbein, F. (2007). Fast and effective feature-preserving mesh denoising. *IEEE transactions on visualization and computer graphics*, *13*(5), 925-938.

U.S. Geological Survey, 2024, 3D Elevation Program 1-Meter Resolution Digital Elevation Model, accessed May 2024 at URL https://www.usgs.gov/the-national-map-data-delivery

U.S. Geological Survey, 2024, NHDPlus High Resolution, accessed May 2024 at URL https://www.usgs.gov/national-hydrography/nhdplus-high-resolution

Abbreviations

3DEP - 3D Elevation Program
DEM - Digital Elevation Model
GDAL - Geospatial Data Abstraction Library
HUC - Hydrologic Unit Code (i.e. sub-basin or sub-watershed)
LiDAR - Light Detection and Ranging
UMBC - University of Maryland, Baltimore County
NHD - National Hydrography Dataset
OOB - Out-of-bag
PF - Priority-flood
RDG - Regional directed graph
RF - Random forest
USGS - United States Geological Survey
WKID - Well-known ID

Suggested Citation

Baker, M. E., Saavedra, D., Cang, X., and Walker, K. (2024). *Hydrography Mapping Supporting Modeling and Targeted Conservation: Project Overview and Lessons Learned*. Final project report to the U.S. EPA. Accessed online at: https://doi.org/10.5066/P1GRAPEX

Appendix 1 - LiDAR inventory used for channel mapping

The following table details the LiDAR sources that were used in this mapping effort. Most LiDAR collections were part of the 3DEP's National Elevation Dataset though some LiDAR collections were not part of 3DEP and were contracted through state or other local agencies. The column "QL" denotes the USGS Quality Level of the LiDAR collection. USGS QL specifications can be found here: https://www.usgs.gov/3d-elevation-program/topographic-data-quality-levels-qls. The column "Collection Year" denotes the year that the LiDAR acquisition was completed, as stated in the contractor-provided metadata. Note that the year acquisition was completed is often different from the nominal year in the project name or the year that the LiDAR collection was made publicly available.

LiDAR Collection Name	QL	Collection Year
MD_Western_2021_D21	2	2021
VA_NorthernShenandoah_2020_D20	2	2021

MD_Southeast_2019_D20	2	2021
MNCPPC Planimetric LiDAR 2020	1	2021
NY_FEMAR2_Central_2018_D19	2	2020
MD_VA_NorthChesapeakeBay_KGeorge_2020_D20	2	2020
VA_UpperMiddleNeck_2018_D18	2	2020
PA_WesternPA_2019_D20	2	2020
VA_SouthamptonHenricoWMBG_2019_B19	2	2020
WV_FEMAR3_Southcentral_B1_2018	2	2020
Anne Arundel County LiDAR 2020	1	2020
PA_Northcentral_2019_B19	2	2019
PA_3_County_South_Central_2018_D18	2	2018
VA_FEMA_NRCS_SouthCentral_2017_D17	2	2018
VA_West_Chesapeake_Bay_Watershed_Lidar_2017_B17	2	2018
VA_Fairfax_County_2018_D18	1	2018
PA_LuzerneCounty_2018_A18	1	2017
PA_South_Central_2017_D17	2	2017
PA_South_Central_2017_D17	2	2017
VA_FEMA_R3_SW_Lidar_2016_D16	2	2017
VA_FEMA_R3_SW_Lidar_2016_D16	2	2017
NY_FEMA_R2_LiDAR_2016_D17	2	2017
2017 Calvert County Lidar	1	2017
PA_Dauphin_County_LiDAR_2016_B16	2	2016
VA_ChesapeakeBay_2015	2	2016
NY_Southwest_2_County_2016_A16	2	2016
WV_FEMA_R3_East_Lidar_2016_D16	2	2016
VA_ChesapeakeBay_2015	2	2016
Garrett_County_MD_QL2_LiDAR	2	2015
Central_NY_Madison_Otsego_Counties_Lidar	2	2015

Eastern_Shore_Virginia_QL2_LiDAR_BAA	2	2015
Sandy_Supplemental_MD_PA_QL2_LiDAR	2	2015
NRCS_Shenandoah_VA_LiDAR	2	2015
Virginia_SANDY_QL2_LiDAR	2	2014
PA_Pennsylvania_SANDY_LiDAR_15	2	2014
VA_Central_Virginia_Seismic_LiDAR_2013_14	2	2014
Sandy_Supplemental_NCR_VA_MD_DC_QL2_LiDAR	2	2014
Sandy_Supplemental_DE_and_MD_QL2_LiDAR	2	2014
USGS_Schoharie2014	2	2014
FEMA_SenecaWatershed2012	3	2014
2015 Baltimore County Lidar	2	2014
Norfolk_VA_LiDAR	2	2013
USDA_Livingston2011	2	2012
VA_FEMA_MiddleCounties_2011	2	2012
VA_LouisaCo_2012	2	2012
VA_LoudounCo_2012	2	2012
VA-WV-MD_FEMA_REGION3_UTM18_2012	2	2012
VA_FEMA_NorthCounties_2011	2	2011
County_Tompkins2008	3	2008

Appendix 2 - Covariates used in random forest modeling

The following table describes the covariates provided to the random forest models for classifying channel-like features into stream, ditch, and other categories. Each covariate was computed as a zonal statistic using the channel-like polygon as a zone unless otherwise noted.

Covariate name	Description
azim_1m_minimum	Minimum azimuth value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the

	geomorphon.
azim_1m_maximum	Maximum azimuth value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azim_1m_range	Range of azimuth values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azim_1m_average	Average azimuth value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azim_1m_stddev	Standard deviation of azimuth value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azim_1m_number	Number of 1-m geomorphon azimuth pixels in channel-like feature.
elong_1m_minimum	Minimum elongation value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
elong_1m_maximum	Maximum elongation value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
elong_1m_range	Range of elongation values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
elong_1m_average	Average elongation value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
elong_1m_stddev	Standard deviation of elongation value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.

elong_1m_number	Number of 1-m geomorphon elongation pixels in channel-like feature.
ext_1m_minimum	Minimum extend value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
ext_1m_maximum	Maximum extend value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
ext_1m_range	Range of extend values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
ext_1m_average	Average extend value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
ext_1m_stddev	Standard deviation of extend values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
ext_1m_number	Number of 1-m geomorphon extend pixels in channel-like feature.
inten_1m_minimum	Minimum intensity value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
inten_1m_maximum	Maximum intensity value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
inten_1m_range	Range of intensity values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
inten_1m_average	Average intensity value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
inten_1m_stddev	Standard deviation of intensity values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
inten_1m_number	Number of 1-m geomorphon intensity pixels in channel-like feature.
range_1m_minimum	Minimum range value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.

range_1m_maximum	Maximum range value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.
range_1m_range	Range of range values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.
range_1m_average	Average range value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.
range_1m_stddev	Standard deviation of range values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.
range_1m_number	Number of 1-m geomorphon range pixels in channel-like feature.
var_1m_minimum	Minimum variance value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_1m_maximum	Maximum variance value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_1m_range	Range of variance values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_1m_average	Average variance value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_1m_stddev	Standard deviation of variance values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_1m_number	Number of 1-m geomorphon variance pixels in channel-like feature.
width_1m_minimum	Minimum width value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.

width_1m_maximum	Maximum width value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
width_1m_range	Range of width values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
width_1m_average	Average width value in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
width_1m_stddev	Standard deviation of width values in the channel-scale geomorphons, calculated from 1-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
width_1m_number	Number of 1-m geomorphon width pixels in channel-like feature.
azimuth_10m_minimum	Minimum azimuth value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azimuth_10m_maximum	Maximum azimuth value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azimuth_10m_range	Range of azimuth values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azimuth_10m_average	Average azimuth value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azimuth_10m_stddev	Standard deviation of azimuth values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon azimuth is defined as the orientation of the long side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
azimuth_10m_number	Number of 10-m geomorphon azimuth pixels in channel-like feature.
elong_10m_minimum	Minimum elongation value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.

elong_10m_maximum	Maximum elongation value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
elong_10m_range	Range of elongation values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
elong_10m_average	Average elongation value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
elong_10m_stddev	Standard deviation of elongation values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon elongation is defined as the proportion between sides of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
elong_10m_number	Number of 10-m geomorphon elongation pixels in channel-like feature.
expo_10m_minimum	Minimum exposition value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon exposition is defined as the distance between the central cell and the furthest vertex of the 8 vertices that comprise the geomorphon.
expo_10m_maximum	Maximum exposition value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon exposition is defined as the distance between the central cell and the furthest vertex of the 8 vertices that comprise the geomorphon.
expo_10m_range	Range of exposition values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon exposition is defined as the distance between the central cell and the furthest vertex of the 8 vertices that comprise the geomorphon.
expo_10m_average	Average exposition value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon exposition is defined as the distance between the central cell and the furthest vertex of the 8 vertices that comprise the geomorphon.
expo_10m_stddev	Standard deviation of exposition values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon exposition is defined as the distance between the central cell and the furthest vertex of the 8 vertices that comprise the geomorphon.
expo_10m_number	Number of 10-m geomorphon exposition pixels in channel-like feature.
extend_10m_minimum	Minimum extend value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
extend_10m_maximum	Maximum extend value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8

	vertices that comprise the geomorphon.
extend_10m_range	Range of extend values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
extend_10m_average	Average extend value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
extend_10m_stddev	Standard deviation of extend values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon extend is defined as the area of the polygon created by the 8 vertices that comprise the geomorphon.
extend_10m_number	Number of 10-m geomorphon extend pixels in channel-like feature.
intensity_10m_minimum	Minimum intensity value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
intensity_10m_maximum	Maximum intensity value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
intensity_10m_range	Range of intensity values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
intensity_10m_average	Average intensity value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
intensity_10m_stddev	Standard deviation of intensity values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon intensity is defined as the mean difference in elevation between the central cell and the 8 vertices that comprise the geomorphon.
intensity_10m_number	Number of 10-m geomorphon extend pixels in channel-like feature.
range_10m_minimum	Minimum range value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.
range_10m_maximum	Maximum range value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.
range_10m_range	Range of range values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.

range_10m_average	Average range value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.
range_10m_stddev	Standard deviation of range values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon range is defined as the difference between the maximum and minimum elevations of the 8 vertices that comprise the geomorphon.
range_10m_number	Number of 10-m geomorphon range pixels in channel-like feature.
var_10m_minimum	Minimum variance value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_10m_maximum	Maximum variance value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_10m_range	Range of variance values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_10m_average	Average variance value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_10m_stddev	Standard deviation of variance values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon variance is defined as the average difference between the individual elevations and the mean elevation of the 8 vertices that comprise the geomorphon.
var_10m_number	Number of 10-m geomorphon variance pixels in channel-like feature.
width_10m_minimum	Minimum width value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
width_10m_maximum	Maximum width value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
width_10m_range	Range of width values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.

width_10m_average	Average width value in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
width_10m_stddev	Standard deviation of width values in the valley-scale geomorphons, calculated from 10-m DEM. Geomorphon width is defined as the length of the short side of the least-square bounding box surrounding the 8 vertices that comprise the geomorphon.
width_10m_number	Number of 10-m geomorphon width pixels in channel-like feature.
area2*	Area of channel-like feature, in square meters.
perim*	Perimeter of channel-like feature, in meters.
cmp_sq*	Compactness of channel-like feature as the ratio of the feature's perimeter to the perimeter of a circle of the same area.
cmp_cir*	Compactness of channel-like feature as the ratio of the feature's perimeter to the perimeter of a square of the same area.
fractdim*	Fractal dimension of channel-like feature. Describes the complexity of the perimeter as two times the ratio of the logarithm of perimeter to logarithm of area.
water**	Percent of 10-m buffer surrounding the channel-like feature classified as water, as computed from the 2018 high-resolution land cover dataset.
emergw**	Percent of 10-m buffer surrounding the channel-like feature classified as wetland, as computed from the 2018 high-resolution land cover dataset.
canopy**	Percent of 10-m buffer surrounding the channel-like feature classified as tree canopy, as computed from the 2018 high-resolution land cover dataset.
shrub**	Percent of 10-mbuffer surrounding the channel-like feature classified as shrub, as computed from the 2018 high-resolution land cover dataset.
grass**	Percent of 10-m buffer surrounding the channel-like feature classified as low vegetation, as computed from the 2018 high-resolution land cover dataset.
barren**	Percent of 10-m buffer surrounding the channel-like feature classified as barren, as computed from the 2018 high-resolution land cover dataset.
build**	Percent of 10-m buffer surrounding the channel-like feature classified as structure, as computed from the 2018 high-resolution land cover dataset.
imperv**	Percent of 10-m buffer surrounding the channel-like feature classified as impervious surface, as computed from the 2018 high-resolution land cover dataset.
road**	Percent of 10-m buffer surrounding the channel-like feature classified as road, as computed from the 2018 high-resolution land cover dataset.

can_bld**	Percent of 10-m buffer surrounding the channel-like feature classified as tree canopy over structure, as computed from the 2018 high-resolution land cover dataset.
can_imp**	Percent of 10-m buffer surrounding the channel-like feature classified as tree canopy over impervious surface, as computed from the 2018 high-resolution land cover dataset.
can_rds**	Percent of 10-m buffer surrounding the channel-like feature classified as tree canopy over road, as computed from the 2018 high-resolution land cover dataset.
all_rds**	Percent of 10-m buffer surrounding the channel-like feature classified as road or tree canopy over road, as computed from the 2018 high-resolution land cover dataset.
all_bldg**	Percent of 10-m buffer surrounding the channel-like feature classified as structure or tree canopy over structure, as computed from the 2018 high-resolution land cover dataset.
all_dev**	Percent of 10-m buffer surrounding the channel-like feature classified as structure, impervious surface, road, tree canopy over structure, tree canopy over impervious surface, or tree canopy over road, as computed from the 2018 high-resolution land cover dataset.
bnkHt_1m_minimum	Minimum bank height value estimated using channel-scale geomorphons, calculated from 1-m DEM.
bnkHt_1m_maximum	Maximum bank height value estimated using channel-scale geomorphons, calculated from 1-m DEM.
bnkHt_1m_range	Range of bank height values estimated using channel-scale geomorphons, calculated from 1-m DEM.
bnkHt_1m_average	Average bank height value estimated using channel-scale geomorphons, calculated from 1-m DEM.
bnkHt_1m_stddev	Standard deviation of bank height values estimated using channel-scale geomorphons, calculated from 1-m DEM.
bnkHt_1m_var	Variance of bank height values estimated using channel-scale geomorphons, calculated from 1-m DEM.
bnkHt_1m_cv	Coefficient of variation of bank height values estimated using channel-scale geomorphons, calculated from 1-m DEM.
bnkHt_1m_number	Number of 1-m bank height pixels in channel-like feature.
elev_10m_minimum	Minimum elevation value in 10-m DEM.
elev_10m_maximum	Maximum elevation value in 10-m DEM.
elev_10m_range	Range of elevation values in 10-m DEM.

elev_10m_average	Average elevation value in 10-m DEM.
elev_10m_stddev	Standard deviation of elevation values in 10-m DEM.
elev_10m_var	Variance of elevation values in 10-m DEM.
elev_10m_cv	Coefficient of variation of elevation values in 10-m DEM.
elev_10m_number	Number of 10-m DEM pixels in channel-like feature.
slope_10m_minimum	Minimum slope calculated from 10-m DEM.
slope_10m_maximum	Maximum slope calculated from 10-m DEM.
slope_10m_range	Range of slope values calculated from 10-m DEM.
slope_10m_average	Average slope calculated from 10-m DEM.
slope_10m_stddev	Standard deviation of slope values calculated from 10-m DEM.
slope_10m_var	Variance of slope values calculated from 10-m DEM.
slope_10m_cv	Coefficient of variation of slope values calculated from 10-m DEM.
slope_10m_number	Number of slope pixels in channel-like feature, calculated from 10-m DEM.
pcurve_10m_minimum	Minimum profile curvature value calculated from 10-m DEM. Profile curvature is calculated parallel to the direction of maximum slope and affects the acceleration or deceleration of flow across the surface.
pcurve_10m_maximum	Maximum profile curvature value calculated from 10-m DEM. Profile curvature is calculated parallel to the direction of maximum slope and affects the acceleration or deceleration of flow across the surface.
pcurve_10m_range	Range of profile curvature values calculated from 10-m DEM. Profile curvature is calculated parallel to the direction of maximum slope and affects the acceleration or deceleration of flow across the surface.
pcurve_10m_average	Average profile curvature value calculated from 10-m DEM. Profile curvature is calculated parallel to the direction of maximum slope and affects the acceleration or deceleration of flow across the surface.
pcurve_10m_stddev	Standard deviation of profile curvature values calculated from 10-m DEM. Profile curvature is calculated parallel to the direction of maximum slope and affects the acceleration or deceleration of flow across the surface.
pcurve_10m_var	Variance of profile curvature values calculated from 10-m DEM. Profile curvature is calculated parallel to the direction of maximum slope and affects the acceleration or deceleration of flow across the surface.
pcurve_10m_cv	Coefficient of variation of profile curvature values calculated from 10-m DEM. Profile curvature is calculated parallel to the direction of maximum slope and affects the acceleration or deceleration of flow across the surface.
pcurve_10m_number	Number of profile curvature pixels in channel-like feature, calculated from 10-m

	DEM.
tcurve_10m_minimum	Minimum tangential curvature value calculated from 10-m DEM. Tangential curvature is calculated perpendicular to the direction of maximum slope and affects the convergence or divergence of flow across the surface.
tcurve_10m_maximum	Maximum tangential curvature value calculated from 10-m DEM. Tangential curvature is calculated perpendicular to the direction of maximum slope and affects the convergence or divergence of flow across the surface.
tcurve_10m_range	Range of tangential curvature values calculated from 10-m DEM. Tangential curvature is calculated perpendicular to the direction of maximum slope and affects the convergence or divergence of flow across the surface.
tcurve_10m_average	Average tangential curvature value calculated from 10-m DEM. Tangential curvature is calculated perpendicular to the direction of maximum slope and affects the convergence or divergence of flow across the surface.
tcurve_10m_stddev	Standard deviation of tangential curvature values calculated from 10-m DEM. Tangential curvature is calculated perpendicular to the direction of maximum slope and affects the convergence or divergence of flow across the surface.
tcurve_10m_var	Variance of tangential curvature values calculated from 10-m DEM. Tangential curvature is calculated perpendicular to the direction of maximum slope and affects the convergence or divergence of flow across the surface.
tcurve_10m_cv	Coefficient of variation of tangential curvature values calculated from 10-m DEM. Tangential curvature is calculated perpendicular to the direction of maximum slope and affects the convergence or divergence of flow across the surface.
tcurve_10m_number	Number of tangential curvature pixels in channel-like feature, calculated from 10-m DEM.
SDrange	Ratio of standard deviation of elevation to range of elevation, calculated from 10-m DEM.

* Indicates that the covariate is not computed as a zonal statistic but is a direct property of the channel-like polygon.

** Indicates that the covariate is not computed as a zonal statistic within the channel-like polygon but is computed from a 10 meter buffer around the channel-like polygon.

Appendix 3 - Upper Rappahannock field validation photos

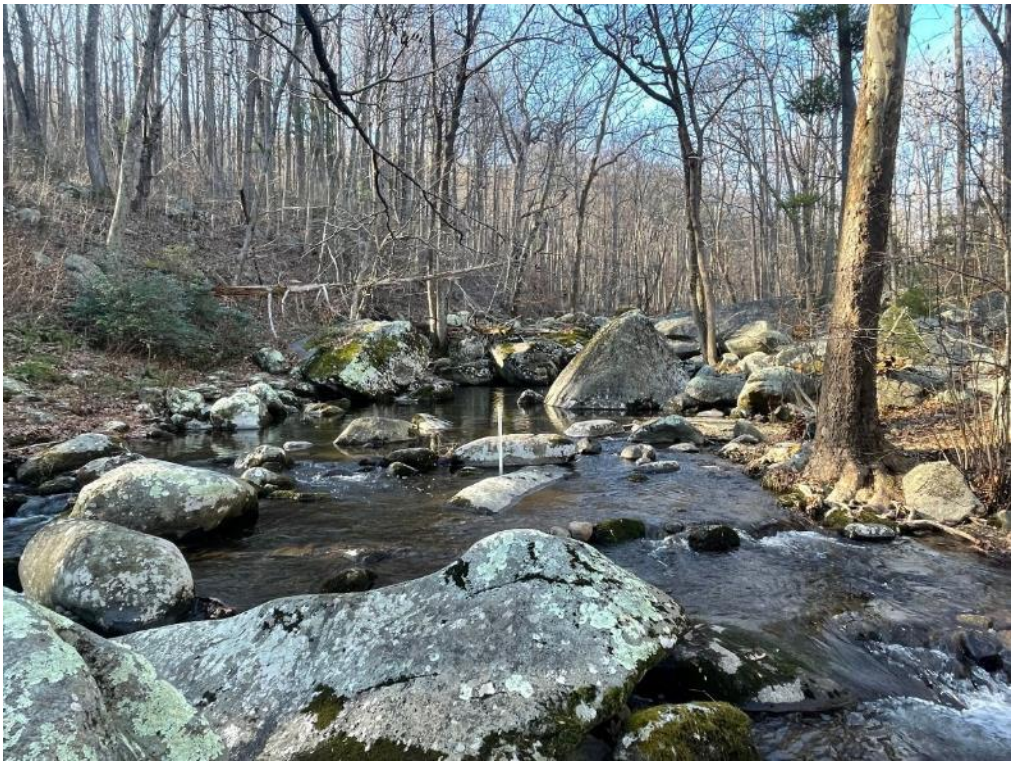
The following photos were collected by Friends of the Rappahannock as part of a field validation effort. Field sites were randomly selected from the hyper-resolution stream maps and data on presence/absence of a stream channel, presence/absence of flow, and channel dimensions were collected.

Site 174 - 3rd order, stream channel present, flowing water:

Site 2335 - 2nd order, stream channel present, standing water:

Site 2434 - 7th order, stream channel present, flowing water:

Site 4023 - 1st order, stream channel present, flowing water:



Site 4845 - 1st order, stream channel present, standing water:

Site 20877 - 1st order, stream channel absent, no water:

Site 22001 - 1st order, stream channel present, no water:

Site 23990 - 1st order, stream channel present, flowing water:

Site 24627 - 2nd order, stream channel absent, no water:

Site 24745 - 2nd order, stream channel present, flowing water:

Site 34178 - 4th order, stream channel present, flowing water:

Site 37786 - 1st order, stream channel present, no water:

Site 39015 - 1st order, stream channel present, no water:

Site 46216 - 4th order, stream channel present, flowing water:

Site 46772 - 2nd order, stream channel present, standing water:

Site 46998 - 1st order, stream channel absent, no water:

Site 54632 - 3rd order, stream channel present, no water:

Site 56849 - 5th order, stream channel present, flowing water:

Site 59932 - 1st order, stream channel present, flowing water:

Site 62698 - 1st order, stream channel present, no water:

Site 80010 - 5th order, stream channel present, flowing water: