Progress Report 21–01.0

PENNSYLVANIA HYDROGRAPHY DATASET

THE PENNSYLVANIA GEOLOGICAL SURVEY'S PROCESS FOR GENERATING FLOWPATHS

---

by Ellen R. Fehrs
Pennsylvania Geological Survey

PENNSYLVANIA GEOLOGICAL SURVEY
FOURTH SERIES
HARRISBURG

2021

REPORT DISCLAIMER

No Warranty

This report and any accompanying documentation are provided "as is," and the user assumes the entire risk as to their quality and performance.

The Commonwealth of Pennsylvania makes no guarantee or warranty concerning the accuracy of information contained in any geographic data or accompanying documentation. The Commonwealth of Pennsylvania further makes no warranties, either expressed or implied, as to any other matter whatsoever, including, without limitation, the completeness or condition of the product, or its fitness for any particular purpose. The burden for determining fitness for use lies entirely with the user.

Although the text has been processed successfully on a computer system at the Bureau of Geological Survey, the Commonwealth of Pennsylvania makes no warranty, expressed or implied, regarding the use of the report or any accompanying documentation on any other computer system, nor does the act of distribution constitute any such warranty.

Limitation of Liability

The user shall save the Commonwealth of Pennsylvania harmless from any suits, claims, or actions arising out of the use of or any defect in this report or any accompanying documentation.

The Commonwealth of Pennsylvania assumes no legal liability for the accuracy, completeness, or usefulness of this report and any accompanying documentation. In no event shall the Commonwealth of Pennsylvania have any liability whatsoever for payment of any consequential, incidental, indirect, special, or tort damages of any kind, including, but not limited to, any loss of profits arising out of use of or reliance on any geographic or geologic data contained herein.

Use and Access Constraints

Not for commercial resale.

This report is not designed for use as a primary regulatory tool in permitting and siting decisions. It is public information and, as such, it may be used as a reference source and may be interpreted by organizations, agencies, units of government, or others based on needs; however, each user is responsible for the appropriate application of the data. Federal, state, or local regulatory bodies are not to reassign to the Bureau of Geological Survey any authority for the decisions they make using this report.

COMMONWEALTH OF PENNSYLVANIA
Tom Wolf, *Governor*
DEPARTMENT OF CONSERVATION AND NATURAL RESOURCES
Cindy Adams Dunn, *Secretary*
OFFICE OF CONSERVATION AND TECHNICAL SERVICES
Lauren Imgrund, *Deputy Secretary*
BUREAU OF GEOLOGICAL SURVEY
Gale C. Blackmer, *Director*

Material from this report may be published if credit is given to the
Pennsylvania Geological Survey

This report has been reviewed by staff of the Pennsylvania Geological Survey.
It has not undergone external peer review.

Instructions for obtaining the complete report can be found at
https://www.dcnr.pa.gov/Geology/PublicationsAndData

Suggested citation: Fehrs, E. R., 2021, Pennsylvania Hydrography Dataset—The Pennsylvania Geological Survey's process for generating flowpaths: Pennsylvania Geological Survey, 4th ser., Progress Report 21–01.0, 73 p.

[BLANK PAGE]

CONTENTS

	Page
Acronyms and abbreviations	ix
Abstract	1
Introduction	1
Background	1
Flowline versus flowpath	1
Required data	2
Supplementary data	3
Field data	3
Existing hydrography data	3
Land cover	3
Stormwater	3
Roads	3
Culverts and bridges	4
Software	4
ArcGIS products	4
GRASS GIS	5
Data standards	5
Lidar base specification	5
Elevation-derived hydrography	5
Acquisition specifications	6
Representation, extraction, attribution, and delineation rules	11
PAHD flowpath-generation steps	13
1. Prepare data	13
2. Set up GRASS GIS	15
3. Run geomorphon analysis	16
Create a channel-extent geomorphon	19
Resample the DEM	19
Filter the resampled DEM	24
Import the filtered DEM into GRASS GIS	24
Define the geographic region	24
Run the geomorphon command	24
Export the raster to an output folder	30
Create a valley-extent geomorphon	30
Resample the DEM	30
Define the geographic region	30
Run the geomorphon command	30
Export the raster to an output folder	31
Create a buffered-channel-extent geomorphon	31
4. Reclassify the geomorphon outputs	31
Channel-extent geomorphon	31
Valley-extent and buffered-channel-extent geomorphons	31
5. Remove superfluous raster cells	32
Valley-extent and buffered-channel-extent ArcScan edits	33

Page
Constrain the channel-extent geomorphon 33
6. Manually delineate breaklines 34
7. Use land cover data to remove unnecessary features 38
8. Use a flow-accumulation model to remove additional unnecessary features 39
Notes on Arc Hydro tools 39
The Arc Hydro workflow 40
Use NHD to improve flow output 42
9. Perform simple topology edits 42
10. Manually edit flowpaths 43
11. Connect flowpath segments 44
12. Make final topology edits 49
Instructions for attribution and networking 51
1. Verify flow direction 51
2. Calculate stream order 55
3. Transfer attributes 57
A note about the transfer of attribute data and the NHD 57
PAHD transfer steps for NHD attributes 59
Strategies and examples for relating NHD flowlines to PAHD flowpaths 60
Example 1: NHD flowlines with multiple permanent identifiers 60
Example 2: Spatial placement 62
Example 3: Function 62
Example 4: Length/Extent 62
A note on intermittent/perennial/ephemeral FCodes 62
Example 5: Relationship to other features 62
Example 6: When NHD feature relationships are incorrect 67
Example 7: Difficult equivalency 67
4. Edit attributes 67
5. Assign additional attributes 70
6. Create a network 70
A note about networks 70
Network inclusivity 70
Discussion and collaboration 71
Automation 72
Feature identification 72
Data acceptance 72
Acknowledgments 72
References 72
Additional reading 73

FIGURES

Page

Figure 1. Diagrams showing differences between D8 and D-infinity raster analyses 6
2. Map showing the distribution of HU12 basins in Pennsylvania and the locations of the nine basins used to develop the methodology for extracting PAHD flowpaths 14
3. Hillshade and orthoimage of farmlands in Chester County demonstrate the gridded pattern produced when projecting a DEM using the Nearest Neighbor resampling technique 15
4. Flowcharts showing the GRASS GIS data organizational structure and examples of GRASS GIS project component names used when preparing PAHD flowpaths 16
5. Screenshots of the GRASS GIS Startup window and sample PAHD entries.. 17
6. Diagrams illustrating the 10 generalized landform interpretations that make up the geomorphon classification system 18
7. Sketches showing simplified examples of geomorphon classification analyses and the effects of varying search radii 19
8. Maps showing a channel-extent geomorphon, a buffered-channel-extent geomorphon, and a valley-extent geomorphon for part of a selected DEM. 20
9. Grids showing how resampling a DEM to a smaller cell size can affect the GRASS GIS geomorphon output 21
10. Maps and images showing how a reduced DEM cell size can improve the accuracy of a geomorphon output 23
11. Maps and images illustrating the effectiveness of using a smoothing filter to reduce noise in a geomorphon output 25
12. Maps showing how successive filtering of a DEM affects original and reclassified geomorphon outputs 26
13. Screenshots of the GRASS GIS Layer Manager and Map Display interfaces. 27
14. Screenshots of the GRASS GIS Command Prompt window and Region Settings window, which are used when defining the geographic region for any given Mapset 28
15. Screenshots of the GRASS GIS Geomorphon Analysis and Map Display interfaces showing parameter fields and an example of a geomorphon output, respectively 29
16. Maps showing reclassified buffered-channel-extent and valley-extent geomorphons 32
17. Hillshade showing a reclassified channel-extent geomorphon output overlaying the cell selections from the buffered-channel-extent and valley-extent geomorphons shown in Figure 16 34
18. Hillshades showing the effects of the Extract By Mask function on segregated geomorphon outputs from Figure 17. 35
19. Hillshade showing a vectorized geomorphon product created from the masked channel raster shown in Figure 18D using ArcMap's Thin and Raster To Polyline tools 36

Page

Figure 20. Hillshade showing an example of how vectorized flowpaths from geomorphon outputs fail in wider flow features 36
21. Hillshades showing an example of how failed geomorphon flowpaths in wider flow features can be improved 37
22. Map and hillshade showing an example of how land cover classes can be used to additionally filter out flowpath segments that are not related to hydrography features 39
23. Flow chart in ModelBuilder showing an Arc Hydro flow accumulation model that can be used to winnow out additional flowpath segments 41
24. Hillshades showing an example of how the output from a flow accumulation model can be used in conjunction with existing NHD flowlines to additionally filter out flowpath segments that are not related to hydrography features 43
25. DEMs showing small "loop" and "dangle" errors created when a geomorphon output was converted from a raster to a vector 44
26. Hillshades showing examples of manual edits that need to be performed prior to least-cost analysis. 45
27. Simplified illustration of ORC tool mechanics shows how the tool finds the least-cost paths on polygons and lines 46
28. Hillshade showing an example of ORC tool outputs overlying the original regions of interest 47
29. Hillshades showing flowpaths incorrectly cutting across higher elevations, and DEMs showing small erroneous kinks in flowpaths—both errors resulting from ORC tool use 48
30. DEM and screenshot of flowpath attributes from ArcGIS Pro shows how to sort on Shape_Length to mass delete all kinked features arising from use of the ORC tool 49
31. Hillshade showing an example of how topology edits can be used to correct network relationships. 50
32. Screenshot of ArcMap's topology Add Rule window showing the rule all junctions "must be covered by the endpoint of" dissolved edges 51
33. Sketch showing an example of how indeterminate flow can occur in a geometric network that uses both sink and source junction assignments to determine flow direction 52
34. Sketch showing how to check whether all flowpaths are digitized in the correct downstream direction by using the red vertex endpoint of the line 53
35. Screenshots showing how to use the elevation values in the flowpaths feature class attribute table to find and select flowpath lines that have been digitized with incorrect directions 54
36. Hillshade and screenshot showing an example of how the -1 value in the NextDownID attribute field can be used to identify flowpath segments that have been digitized in the wrong direction 56

Page

Figure 37. Map, hillshade, and screenshot showing an example of how a line segment could have a NextDownID value of -1 and not have been digitized in the incorrect direction if it is exiting an HU12 basin 57
38. Hillshade and screenshot showing an example of how a line segment could have a NextDownID value of -1 and not have been digitized in the incorrect direction if it is connects to a downstream geometry via subsurface flow .. 58
39. Hillshade showing an example of an inaccurate NHD attribute that was transferred to the PAHD 60

Figures 40–46. Hillshades showing—
40. An example of how NHDFlowline attribution does not always make sense without the context of additional feature layers 61
41. An example of how to relate an NHD flowline to PAHD flowpaths using spatial placement 63
42. The transfer of permanent identification values from NHD flowlines to PAHD flowpaths based on function 64
43. How to determine equivalency between an NHD flowline and a PAHD flowpath by comparing length/extent 65
44. An example highlighting network relationships between features 66
45. An example of when to ignore NHD flowline networking relationships 68
46. A difficult interpretation of flowline-flowpath equivalency 69

TABLES

Page

Table 1. Comparison of PAHD and LBS hydro-flattening requirements for water features 7
2. Comparison of line feature attributes in the NHD, EDH dataset, and PAHD 8
3. FCodes used for line and polygon features in the NHD, EDH dataset, and PAHD 10
4. Comparison of READ Rules and PAHD capture conditions for various hydrography features 12

ACRONYMS AND ABBREVIATIONS

BGS	Bureau of Geological Survey
DEM	Digital Elevation Model
EDH	Elevation-Derived Hydrography
EPSG	European Petroleum Survey Group
Esri	Environmental Systems Research Institute
FCode	Feature Code
FType	Feature Type
GIS	Geographic Information System
GNIS	Geographic Names Information System
GRASS	Geographic Resources Analysis Support System
GUID	Globally Unique Identifier
HU	Hydrologic Unit
HUC	Hydrologic Unit Code
LAS	Short for laser, LAS is the extension for an industry-standard binary format file used for storing lidar point cloud data.
LBS	Lidar Base Specification
Lidar	Light Detection and Ranging
MS4	Municipal Separate Storm Sewer System
NAIP	National Agriculture Imagery Program
NHD	National Hydrography Dataset
ORC	Optimal Region Connections
PAHD	Pennsylvania Hydrography Dataset
PEMA	Pennsylvania Emergency Management Agency
QA	Quality Assurance
QC	Quality Control
QL	Quality Level
READ	Representation, Extraction, Attribution, and Delineation
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey

PENNSYLVANIA HYDROGRAPHY DATASET

THE PENNSYLVANIA GEOLOGICAL SURVEY'S PROCESS FOR GENERATING FLOWPATHS

by

Ellen R. Fehrs

ABSTRACT

The Pennsylvania Department of Conservation and Natural Resources, Bureau of Geological Survey (BGS) is developing a workflow to produce elevation-derived geographical data that will comprise the Pennsylvania Hydrography Dataset (PAHD). This workflow relies on geomorphon classification of elevation data as the primary means of identifying flowpath geometries. Subsequent steps, such as least-cost paths and manual edits, winnow artifacts and other irrelevant features. The intent of this project is to use Quality Level 2 (QL2) light detection and ranging (lidar) data or better to produce a scale-equivalent and dynamic hydrography dataset for the state of Pennsylvania. This means that PAHD flowpaths will have a minimum horizontal accuracy of 1 m and a minimum vertical accuracy of 0.5 m.

INTRODUCTION

This document represents the most recent iteration (March 2021) of the workflow used by the BGS to create data for the PAHD. Still these processes are considered a work in progress; as such, these instructions, associated data, and any related user-produced data should not be considered a final product of PAHD. Rather, all should be considered a step toward the final dataset.

The BGS shares these methods to encourage collaboration within the greater hydrography community. They intend to start and sustain communication regarding the most efficient and effective means of producing hydrography data for Pennsylvania. To this end, this report provides an opportunity for

1. Readers to submit critiques of these methods and the resulting data/derivatives
2. Users to suggest improvements to the efficiency/efficacy of this workflow, particularly with regards to the automation of existing processes
3. Members of the greater community to use these methods for their own purposes and to contribute their work to PAHD

BACKGROUND

FLOWLINE VERSUS FLOWPATH

The current vocabulary associated with hydrography data is varied and occasionally inconsistent, which necessitates making a distinction between flowline and flowpath features.

The National Hydrography Dataset (NHD) uses a flowline feature class for which it provides the following definition (U.S. Geological Survey (USGS), 2010): "NHD Flowlines consist of routes that make up a linear surface water drainage network. Flowlines have a reach code and a measure, allowing for the establishment of upstream/downstream relationships. This network allows for powerful analysis and modeling capabilities. Each Feature type is attributed with descriptive information by an Fcode."

The NHD treats streams/rivers as a specific feature type (FType) of the NHDFlowline feature class (or the NHDArea feature class, depending on feature dimensions). A stream/river feature is defined as "A body of flowing water." In the NHD, a stream/river feature may be named or unnamed.

In this document, "flowpath" is being used to describe a linear hydrographic network that includes surface and some subsurface drainage features. Flowpaths are very much like a potentiometric surface: they occur where water can potentially accumulate but may not represent an actual channel or show evidence of erosion. This paper uses the term "flowpath" because—while the geomorphon technique, by definition, produces channelized features (i.e., flowlines)—the PAHD incorporates ancillary data (e.g., municipal separate storm sewer system (MS4) data) and techniques (e.g., flow accumulation and manual edits) that result in PAHD flow features that may not represent true channels.

Also, unlike the NHDFlowline feature class, PAHD flowpaths do not presently include Permanent Identifier values or ReachCodes—nor are these necessary to establish upstream/downstream relationships in a geometric network. PAHD flowpaths use the change in elevation from the start of a line segment to the end of a line segment to define upstream/downstream relations. Although each segment or "edge" in the PAHD flowpath network has an ObjectID, the ObjectIDs are by no means permanently assigned to their associated features. The BGS does understand the importance of standardized flow-feature identifiers, and they may incorporate these types of identifiers in a future iteration of the dataset.

REQUIRED DATA

The only data required for the methodologies described in the following sections are

1. The standard deliverables for QL2 lidar collection as outlined in the National Geospatial Program's Lidar Base Specification (LBS) (USGS, 2020):
   a. Metadata
   b. Classified point data
   c. Bare-earth surface (i.e., raster digital elevation model (DEM))
   d. Breaklines
2. Orthoimagery, preferably leaf-off

The DEMs, point data, breaklines, and orthoimagery used by the BGS are distributed through the Pennsylvania Spatial Data Access (PASDA) website. The orthoimagery shown in the figures in this Progress Report were flown leaf-off by the Pennsylvania Emergency Management Agency (PEMA) or leaf-on by the U.S. Department of Agriculture (USDA) National Agriculture Imagery Program (NAIP). Hillshades shown in the figures were produced from DEMs created by various agencies from QL2 lidar. The DEMs are organized and distributed by project name or region.

SUPPLEMENTARY DATA

There are supplementary data not necessary for the PAHD workflow that, when used, can greatly reduce the time and manual labor involved in generating accurate flowpaths. Descriptions of these data and their benefits follow.

Field Data

Reconnaissance and coordinate points taken in the field can serve a number of purposes, including the following:

1. Verifying the existence of a flowpath
2. Checking the existence and placement of a culvert or bridge
3. Determining the locations of flow-initiation points

This information can be used to make interpretations and inferences in settings where field-verified data do not exist.

Existing Hydrography Data

There are two classes of existing hydrography data—those which meet the standards for inclusion in the PAHD and those which do not meet those standards. For geometries to be included in the PAHD, the features must be scale equivalent to QL2 lidar or better. Features that do not have geometries that meet those standards can still be helpful; they can assist in manual delineation, quality assurance (QA), and quality control (QC). Even if a dataset is not suitable for incorporation, useful attributes in some existing datasets could be transferred to equivalent PAHD features.

Land Cover

Land cover data can be used to help constrain geomorphon outputs. While geomorphons can delineate channel-type features, they are prone to incorrect classification of some anthropogenic features as natural channels. Land cover classes can be used to rule out features that cannot possibly represent flowpaths, such as impervious roads or buildings.

Stormwater

MS4 datasets can be helpful in urbanized settings. These data allow the PAHD to be more than a surficial hydrographic network and incorporating subsurface drainage data contributes toward a more holistic hydrography system. Unfortunately, these data are difficult to find and vary greatly in spatial and temporal resolution and completeness.

Roads

Road datasets can be used to differentiate roads from linear flow features such as canals or ditches, which are similar in appearance and frequently run parallel to roadways. Intersections between roads and flow features can also help in the identification of culverts and bridges.

Culverts and Bridges

Culvert-type features are vital to the evolution and application of the PAHD. They are needed to achieve the ultimate goal of producing hydro-flattened, hydro-enforced, and hydro-conditioned[1] products. This is somewhat complicated by the fact that one of the major components in the BGS methodology is the manual delineation of culvert and bridge features. It would be better to have accurate culvert or bridge datasets that could be used to reduce the manual workload. Unfortunately, the majority of existing culvert datasets are based on intersections between NHD flowlines and Pennsylvania Department of Transportation roads, the former of which does not meet PAHD accuracy standards.

In a similar vein, lidar deliverables do not always meet the standard for accurate bridge deck removal. The LBS indicates that a bridge is "distinguished from a roadway over a culvert in that a bridge is an elevated deck that is not underlain with earth or soil." The LBS includes bridge decks in the LAS (lidar point cloud) classification and has required the removal of bridge features from bare-earth surface (DEM) deliverables since version 2.0 (USGS, 2018).

A bare-earth DEM is therefore not expected to include bridge deck points; however, in practical application, many of the bare-earth DEMs do have the occasional (sometimes multiple) bridge decks included. Correcting these oversights falls on the manual portion of this workflow.

SOFTWARE

This section covers the geographic information systems (GIS) software programs that are being used by the BGS to generate flowpaths, but many of the functions and tools described herein are available through alternative software programs. The efficacy of using programs such as QGIS, TauDEM, and SAGA GIS with the workflow outlined in this document has not been tested, but this does not preclude their use in deriving flowpath geometries that meet PAHD standards.

ArcGIS Products

While the BGS utilizes Esri's ArcMap (version 10.7.1) and ArcGIS Pro (version 2.5.0) to generate flowpath geometries, ArcMap is used for the majority of the workflow. Many of the ArcMap tasks outlined in this document can alternatively be accomplished using ArcGIS Pro—with the exception of those completed using ArcMap's ArcScan toolbar, which does not have an equivalent toolset in ArcGIS Pro.

The BGS employs ArcGIS Pro's Optimal Region Connections (ORC) tool (included with Esri's Spatial Analyst extension) to perform cost-based distance analysis on geomorphon

[1] Heidemann (2018), explained the concepts of hydro-flattening, hydro-enforcement, and hydro-conditioning as sequential steps. "Hydro-flattening" modifies DEMs using breaklines to level the surfaces of waterbodies and the cross-sectional surfaces of wide flow features while preserving their upstream-downstream gradients. "Hydro-enforcement" modifies the elevation values along flowpaths, which results in the removal of artificial barriers (e.g., culverts). Finally, "hydro-conditioning" requires all unnatural sinks to be filled to their pour points, so that the flow of water will be continuous across the surface. The goal of these three steps is to produce a DEM suitable for hydrographic modeling.

outputs. The ORC tool is used in lieu of ArcMap's Cost Connectivity tool (also included with Esri's Spatial Analyst extension) for the following reasons:

1. The ORC tool can process data while other tasks are performed concurrently in ArcMap.
2. The ORC tool provides the option of calculating distance using a planar (flat earth) or geodesic (ellipsoid) method. Selecting the ellipsoid method produces more accurate results for larger regions. Additionally, performing the distance calculation on the ellipsoid means that the output optimal paths will not change regardless of input raster and output path projections. The Cost Connectivity tool does not allow the user to choose this parameter.
3. The ORC tool appears to perform better than the Cost Connectivity tool when processing large quantities of data. The ORC tool produces results more quickly and is less prone to crashing than its legacy counterpart.
4. The ORC tool analyzes least-cost/optimal paths using the D-infinity (infinite directions) method, as opposed to the D8 (eight directions) method used by the Cost Connectivity tool. The D-infinity method avoids the distance distortions that arise from the D8 method (Figure 1).

Theoretically, flowpaths that meet the current standards for PAHD inclusion can be derived with the described methodology using either ArcMap or ArcGIS Pro.

GRASS GIS

Geographic Resources Analysis Support System (GRASS) GIS is a free and open-source software. In this report, GRASS is used to perform geomorphon analysis on DEMs prepared in ArcMap/ArcGIS Pro.

DATA STANDARDS

Lidar Base Specification

The LBS comprises formal guidelines that describe what constitutes acceptable lidar data for the expressed purposes of (1) retaining the complete source dataset to support a variety of geospatial applications, and (2) contributing toward more consistent industry practices and collaborative efforts (Heidemann, 2018, p. 1–2). As technology and industry standards have evolved, the LBS has undergone a series of edits and updates. Mentions of the LBS in this document refer to version 1.3, the most recent full-text publication by Heidemann (2018) in conjunction with the most recent set of online updates (USGS, 2020).

The LBS standards most relevant to this report are those addressing DEM surface treatments (i.e., hydro-flattening) and the products required to perform these treatments (e.g., breaklines). LBS and PAHD requirements and conditions for hydro-flattening are listed in Table 1.

Elevation-Derived Hydrography

The USGS has published two companion reports to address specifications for elevation-derived hydrography (EDH) data: *Elevation-Derived Hydrography Acquisition Specifications,*

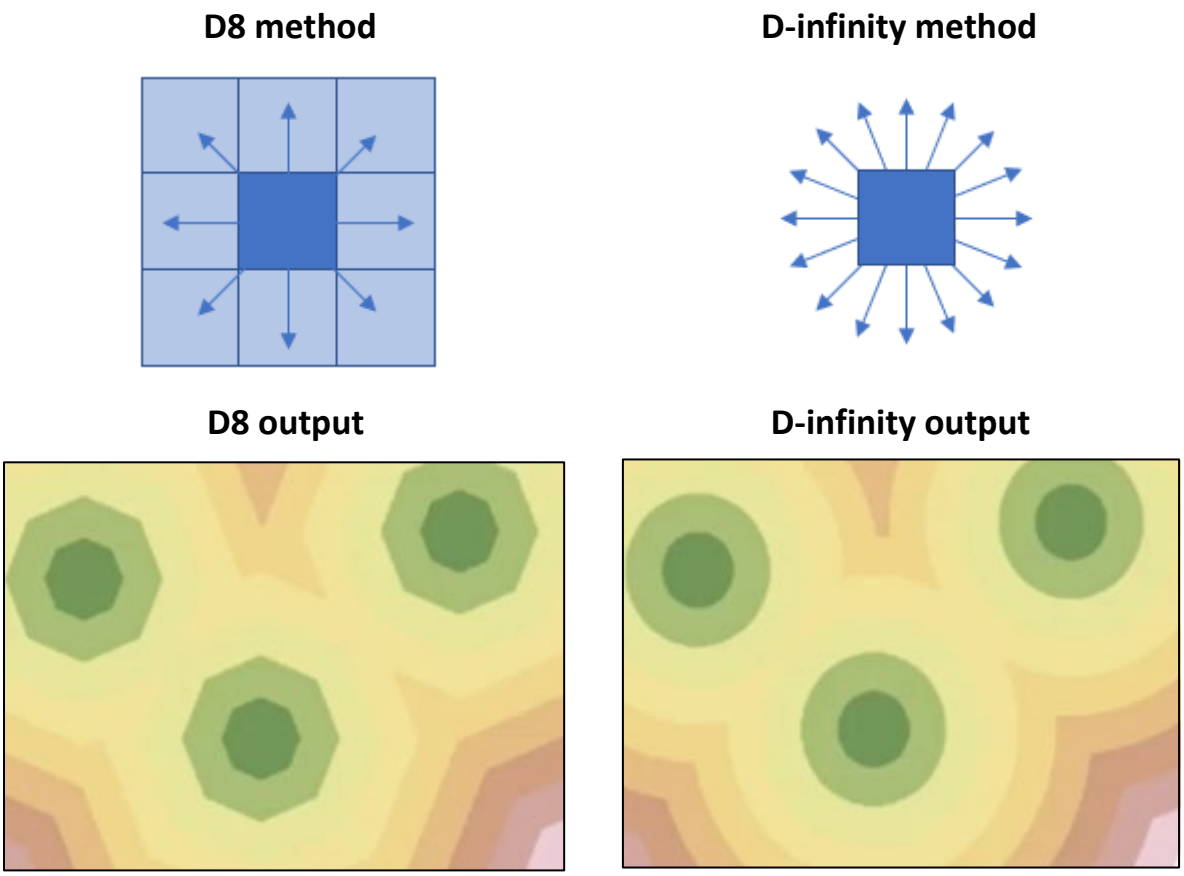


Figure 1. Differences between D8 and D-infinity raster analyses. The D8 method examines a cell's eight neighbors and determines a path to the neighboring cell with the lowest elevation. The D-infinity method examines all the surrounding directions from a cell, not just the central path to the eight neighbors. The D8 method can result in distortion errors, as seen in the D8 output image. (The examples of D8 and D-infinity outputs were taken from the Esri 2020 User Conference Session "Distance Analysis: Identifying Optimal Paths Using Rasters" by Kevin Johnston and Elizabeth Graham).

by Terziotti and Archuleta (2020); and *Elevation-Derived Hydrography—Representation, Extraction, Attribution, and Delineation Rules* (also known as READ Rules), by Archuleta and Terziotti (2020).

Acquisition Specifications

Terziotti and Archuleta (2020) covered the requirements and expectations for the collection of EDH data that are suitable as (1) input for lidar surface derivatives, and (2) a source layer for conflation with the NHD. Data requirements for inclusion in the PAHD generally follow the principles laid down by these specifications with a few notable exceptions:

- **Collection Area**: The EDH Acquisition Specifications require that features be collected within a minimum 10-digit hydrologic unit (HU10), whereas the PAHD requires that features be collected within a minimum 12-digit hydrologic unit (HU12).

Table 1. *Comparison of PAHD and LBS hydro-flattening requirements for water features.*

Features	PAHD hydro-flattening requirements	LBS hydro-flattening requirements
Inland Ponds and Lakes	Waterbodies with a surface area of 0.4 hectare (ha; 1 acre) or greater (approximately equal to a round pond 72 m in diameter) at the time of collection shall be flattened.	Waterbodies with a surface area of 0.8 hectare (ha; 2 acres) or greater (approximately equal to a round pond 100 m in diameter) at the time of collection shall be flattened.
	Same as LBS requirements.	Flattened waterbodies shall present a flat and level water surface (a single elevation for every bank vertex defining the waterbody's perimeter).
	Same as LBS requirements.	The entire water surface edge shall be at or below the immediately surrounding terrain.
	Same as LBS requirements.	Long impoundments such as reservoirs, inlets, and fjords, whose water surface elevations decrease with downstream travel, shall be treated as streams or rivers.
Inland Streams and Rivers	Breaklines will be manually delineated where geomorphon analysis fails. This includes streams and rivers of roughly a 10-m width or greater.	Streams and rivers of a 30-m or greater nominal width shall be flattened.
	Streams or rivers whose width varies above and below 10 m will not be broken into multiple segments; data producers will use their best professional cartographic judgment in determining when a stream or river has attained a nominal 10-m width.	Streams or rivers whose width varies above and below 30 m will not be broken into multiple segments; data producers will use their best professional cartographic judgment in determining when a stream or river has attained a nominal 30-m width.
	Same as LBS requirements.	Flattened streams and rivers shall present a flat and level water surface bank-to-bank (perpendicular to the apparent flow centerline).
	Same as LBS requirements.	Flattened streams and rivers shall present a gradient downhill water surface, following the immediately surrounding terrain.
	Same as LBS requirements.	In cases of sharp turns of rapidly moving water, where the natural water surface is notably not level bank-to-bank, the water surface will be represented as it exists while maintaining an aesthetic cartographic appearance.
	Same as LBS requirements.	The entire water surface edge shall be at or below the immediately surrounding terrain.
	Stream channels, including flow through culvert locations, shall be "burned" into the associated hydro-flattened DEM.	Stream channels shall break at culvert locations leaving the roadway over the culvert intact.
	Same as LBS requirements.	Streams shall be continuous at bridge locations.
	Same as LBS requirements.	Bridges in all their forms shall be removed from the DEM.
	Same as LBS requirements.	When the identification of a structure as a bridge or culvert cannot be made definitively, the feature shall be regarded as a culvert.
Non-Tidal Waterbodies	Same as LBS requirements.	Boundary waterbodies are waterbodies that contain some or all of the designated project area.
	Same as LBS requirements.	Boundary waterbodies may be any type of waterbody but are virtually always large in area or width.
	Same as LBS requirements.	A boundary waterbody shall be represented as a polygon that follows the shore throughout the project and is then closed using arbitrary line segments as needed across the waterbody. Boundary waterbodies do not include the natural far shoreline.
	Same as LBS requirements.	The water surface shall be flat and level, as appropriate for the type of waterbody (level for lakes, gradient for rivers, and so forth).
	Same as LBS requirements.	All landward water surface edges shall be at or below the immediately surrounding terrain.
	Same as LBS requirements.	Unusual changes in the water surface elevation that may take place over the course of the collection (for example, different river stages due to increased or decreased discharge from an upstream dam) shall be documented in the project metadata.
Islands	Islands within lake/pond or within a stream/river that are 5 m or greater along the shortest axis shall be delineated.	Permanent islands 0.4 ha (1 acre) (approximately equal to a round island 72 m in diameter) or larger shall be delineated within all waterbodies.

- **Attributes**: In terms of attribution, the PAHD schema is not as exhaustive as the NHD schema, but it is more detailed than the EDH dataset schema. Because the EDH dataset is intended as a conflation product for the NHD, it has considerably fewer attribute fields—most significant is the absence of network relationships and unique identifiers (e.g., Permanent_ID). See Table 2.
- **FCodes and FTypes**: The NHD attributes include both FTypes and FCodes. The EDH dataset and PAHD only have FCodes, which either equate to or are subcategories of FTypes. The differences between features included in the datasets are as follows: (1) the EDH dataset does not include subtypes of canal/ditch or stream/river features; (2) the NHD does not identify user-defined, drainageway, or culvert features; (3) neither the EDH dataset nor the PAHD include subcategories for pipelines; and (4) the EDH dataset does not identify underground conduit features. See Table 3.
- **Headwaters at Roads**: The EDH Acquisition Specifications outline four general rules for identifying and delineating headwaters. PAHD inclusion requirements differ from the second of these four rules, which states, "If a stream channel is visible (in imagery or a lidar surface) upstream from the road, extend the stream/river through the road using the rules described for delineation of culverts, and extend the stream/river at least 100 ft or 30 m upstream from the road intersection." (Terziotti and Archuleta, 2020, p. 30). When delineating an initiation point for a PAHD flowpath at or near a road, it is not necessary to extend the geometry any further upstream from its intersection with the road than there is evidence of its existence.
- **Collection of Canal/Ditch Features**: The EDH Acquisition Specifications and READ Rules both refer to the collection of canal/ditch features; specifically, these documents provide instructions to "avoid the over-collection" of canal/ditch features in order to avoid "too detailed" features that would result in overly complicated hydrography networks. The current PAHD methodology for flowpath collection does not adhere to this rule. The PAHD approach to incorporating "too detailed" features into a greater network is discussed under "Network Inclusivity" on pages 70–71.

Table 2. *Comparison of line feature attributes in the NHD, EDH dataset, and PAHD.*

(Presence of feature in each dataset indicated by checkmark.)

Attribute Field	Description	NHD	EDH	PAHD
Object ID	Value maintained by ArcGIS that guarantees a unique ID for each row in the table (based on NHD ObjectID).	✓		✓
Shape	Describes feature geometry; calculated automatically.	✓		✓
Shape_Length	Gives length of edge/line segment.	✓		✓
FClass	Defines whether a feature is intended for use as an NHD feature, a non-NHD feature, or a nonhydrography feature.		✓	
EClass	Indicates whether a feature is not used for elevation derivatives; used for hydro-flattening or for other nonspecified elevation purposes; used for hydro-flattening; or an elevation dataset limitation feature.		✓	
FType	Three-digit integer value; unique identifier of a feature type.	✓		

Table 2. *(Continued)*

Attribute Field	Description	NHD	EDH	PAHD
FCode	Five-digit integer value; comprised of the feature type and combinations of characteristics and values.	✓	✓	✓
Desc	Text that describes FCode or user-defined features not included in the domain list.		✓	
Permanent_Identifier/ Permanent_ID	40-character GUID value that uniquely identifies the occurrence of each feature in The National Map—National Database primary key.	✓		✓
FDate	Date of last feature modification.	✓		✓
Resolution	Source resolution. Currently NHD is available in several resolutions. Plans are to develop a single-resolution database using the highest resolution data that can be generalized.	✓		
GNIS_ID	Unique identifier assigned by GNIS, 10-character value.	✓		✓
GNIS_Name	Proper name, specific term, or expression by which a particular geographic entity is known, 65-character value.	✓		✓
EntityName	Proper name, specific term, or expression by which a particular geographic entity is known but that is not included in the GNIS database or presents as a NULL value in the NHD GNIS_Name field. May be carried over from ancillary datasets, e.g., the DCNR Bureau of Forestry stream dataset and county stream datasets.			✓
LengthKm	Length of linear feature based on Albers Equal Area, 8-character value.	✓		
ReachCode	Unique identifier. The first eight digits are the WBD_HUC8. The next six digits are randomly assigned, sequential numbers that are unique within a HUC8, 14-character value.	✓		✓
Enabled	Describes whether a geometric network feature allows flow ("enabled") or whether the geometric network feature blocks flow ("disabled").			✓
FlowDir/FLOWDIR	Direction of flow relative to coordinate order, 4-character value.	✓		✓
WBArea_Permanent_ Identifier	Permanent_Identifier of the waterbody through which the flowline flows.	✓		
MainPath	Indicates whether a flowline is considered the main path of a divergence.	✓		
HYDROID	Unique ID assigned by Arc Hydro Attribute Tools to features in a geometric network.			✓
NextDownID	HydroID of next downstream flowpath; defined by flow direction.			✓
StrahlerOrder	Stream order calculated using the Strahler method.			✓
InNetwork	Indicates whether the feature is included in the network.	✓		
VisibilityFilter	Allows for filtering of vector data features at eight approximate scales. A given VisibilityFilter coded value indicates that the feature is appropriate for use at approximately the defined scale and all larger scales.	✓		
Source/DataSource	Text description of the elevation source data used for deriving the hydrography. Lidar-source data should include collection name, date, and quality level.		✓	✓
Method	Text description of the method used for deriving the hydrography.		✓	✓
UserCode	A code designated by the acquisition entity to identify features collected outside the scope of features described in the "EDH Acquisition Specifications" document. It is intended to be used as a key to join tables with attributes outside of this specification.		✓	
Comments	Field for user comments.		✓	
FeatureSource	Describes source of feature(s)/source of feature submission.			✓

Table 3. *FCodes used for line and polygon features in the NHD, EDH dataset, and PAHD.*

(Presence of FCode in each dataset indicated by checkmark.)

FCode[1]	Feature[1]	Description[1]		NHD	EDH	PAHD
0	User-defined feature	Feature type only: no attributes			✓	✓
34300	Dam/Weir	Feature type only: no attributes		✓	✓	✓
33400	Connector	Feature type only: no attributes		✓	✓	✓
33600	Canal/Ditch	Feature type only: no attributes		✓	✓	✓
33601		Aqueduct		✓		✓
33603		Stormwater		✓		✓
36100	Playa	Feature type only: no attributes		✓	✓	✓
37800	Ice mass	Feature type only: no attributes		✓	✓	✓
39000	Lake/Pond	Feature type only: no attributes		✓	✓	✓
42000	Underground conduit	Feature type only: no attributes		✓		✓
42001		Positional accuracy	Definite	✓		✓
42002			Indefinite	✓		✓
42003			Approximate	✓		✓
42800	Pipeline	Feature type only: no attributes		✓	✓	✓
42801		Aqueduct	At or near surface	✓		
42802			Elevated with respect to surface	✓		
42803			Underground	✓		
42804			Underwater	✓		
42805		General case	At or near surface	✓		
42806			Elevated with respect to surface	✓		
42807			Underground	✓		
42808			Underwater	✓		
42809		Penstock	At or near surface	✓		
42810			Elevated with respect to surface	✓		
42811			Underground	✓		
42812			Underwater	✓		
42813		Siphon	Unspecified relation to surface	✓		
42814		General Case		✓		
42815		Penstock		✓		
42816		Aqueduct		✓		

Table 3. *(Continued)*

FCode[1]	Feature[1]	Description[1]		NHD	EDH	PAHD
43600	Reservoir	Feature type only: no attributes		✓	✓	✓
44500	Sea/Ocean	Feature type only: no attributes		✓	✓	✓
45000	Sink/Rise	Feature type only: no attributes		✓	✓	✓
46000	Stream/River	Feature type only: no attributes		✓	✓	✓
46003		Hydrographic category	Intermittent	✓		✓
46006			Perennial	✓		✓
46007			Ephemeral	✓		✓
46800	Drainageway	Feature type only: no attributes			✓	✓
53700	Area of complex channels	Feature type only: no attributes		✓	✓	✓
55800	Artificial path	Feature type only: no attributes		✓	✓	✓
56600	Coastline	Feature type only: no attributes		✓		✓
Assigned based on adjacent features.	Culvert	Feature type only: no attributes			✓	✓
991	Low-confidence area (predetermined)	Feature type only: no attributes			✓	✓
992	Low-confidence area (sparse bare earth)	Feature type only: no attributes			✓	✓
993	Low-confidence area (snow covered)	Feature type only: no attributes			✓	✓

[1]FCodes, features, and descriptions are from the U.S. Geological Survey (2010) and Terziotti and Archuleta (2020).

The EDH Acquisition Specifications refer to some collection and delineation criteria for specific hydrography features that differ significantly from PAHD requirements. These criteria are more clearly described in the EDH READ Rules and are discussed in the following section.

Representation, Extraction, Attribution, and Delineation Rules

Archuleta and Terziotti (2020) described the fields, domains, and minimum collection requirements for EDH features. Disparities between the READ Rules and PAHD feature standards are identified in Table 4. These divergences stem from the READ Rules' defined capture conditions, which strictly limit the addition of hydrography features that are not already included in the NHD. A stated goal of the READ Rules is to generate features suitable for conflation with the NHD—it is not unreasonable to conclude that this goal strongly influenced the requirements for the EDH capture conditions. As PAHD does not have the same conflation goal, it can be less restrictive about what hydrography features to include (Table 4).

Table 4. *Comparison of READ Rules and PAHD capture conditions for various hydrography features.*

Feature	READ Rules capture conditions	PAHD capture conditions
Area of Complex Channels	If area of complex channels contains at least five active subchannels (connected to the main channel with water in them, not scarring because of river channel migration) and is greater than or equal to 320 ft (98 m) along the shortest axis and greater than or equal to 984 ft (300 m) along the longest axis, then capture.	The EDH definition of an Area of Complex Channels is, "An area where a stream or river flows in an intricate network of interlacing channels with no permanent, primary channel." The PAHD captures any feature that meets this definition and whose channel morphology is definite enough to be recognized by geomorphon analysis performed on lidar derivatives of spatially and temporally appropriate resolution.
Artificial Path	Artificial paths shall be placed in all polygons except isolated reservoirs, isolated lake/ponds, and isolated basins (reservoirs, lake/ponds, and basins not connected to the stream network). Artificial paths shall represent the shortest path from the inflow to the outflow without crossing through banks or islands.	Artificial paths shall be placed in any polygon that is connected to a flowpath, regardless of whether this flowpath is connected to the greater flowpath network. Artificial paths shall represent the most reasonable interpretation of real-world flow from the inflow to the outflow without crossing through banks or islands. This can be based on channels that become visible during periods of extreme dryness or where flow appears to be concentrated based on the shape of the waterbody. Artificial paths are an abstract concept whose purpose is to maintain flow connectivity and will likely vary in depiction from one creator to another.
Canal/Ditch	If canal/ditch is named, or if canal/ditch is ≥984 ft (≥300 m) along the longest axis, then capture. If canal/ditch is needed to provide network connectivity, then capture. If canal/ditch is within agricultural fields and drains to another canal/ditch or other hydrologic feature, then do not capture (see following exception). The exception is if a project has a special need for canals and ditch features that are within agricultural fields and drain to another canal/ditch or other hydrologic feature, then these features should be coded with a separate FCode (user defined) so that they can be excluded from the features to conflate to NHD. Note that a hydrologic network should not be broken if features are excluded.	If a canal/ditch has channelized morphology significant enough to be identified by geomorphon analysis performed on lidar derivatives of spatially and temporally appropriate resolution, it shall be captured.
Island	The minimum size for an island within lake/pond is 60 ft (18 m) along the shortest axis. The minimum size for an island within stream/river is 60 ft (18 m) along the shortest axis.	The minimum size for an island within lake/pond or within a stream/river is 5 m along the shortest axis.
Lake/Pond	For required lidar base specification breakline acquisition, waterbodies with a surface area of ≥2 acres (0.8 hectare) (approximately equal to a round pond 328 ft [100 m] in diameter) at the time of collection shall be flattened. For EDH feature collection, if lake/pond is ≥100 ft (≥30 m) along the shortest axis (or approximately 0.4 hectare), then capture (for required hydro-flattening breaklines, not applicable).	The PAHD minimum lake/pond dimension requirement for feature collection is approximately 15 m along the shortest axis.

Table 4. (*Continued)*

Feature	READ Rules capture conditions	PAHD capture conditions
Reservoir	For required lidar base specification hydro-flattening breakline acquisition, waterbodies with a surface area of ≥2 acres (≥0.8 hectare) or greater (approximately equal to a round pond 328 ft [100 m] in diameter) at the time of collection shall be flattened. For EDH feature collection, if lake/pond is ≥100 ft along the shortest axis (or approximately 0.4 hectare), then capture (for required hydro-flattening breaklines, not applicable).	The PAHD reservoir dimension requirement for feature collection is approximately ≥15 m along the shortest axis.
Stream/River	If stream/river flows from lake/pond or spring/seep, or if stream/river is ≥984 ft (≥300 m) along the longest axis, then capture.	Due to regulatory use, PAHD does not define all flowpaths as stream/river. A flowpath may be stream/river or it may be a "zero-order" flow feature. The only requirement for flowpath inclusion in the PAHD is that the channel feature or features be recognized by geomorphon analysis using lidar derivatives that have spatially and temporally appropriate resolution.

PAHD FLOWPATH-GENERATION STEPS

The BGS plans to release the PAHD as a group of data packages bundled by HU12 (subwatershed) basins. Each dataset package will be released on the Pennsylvania Spatial Data Access website at ftp://ftp.pasda.psu.edu/pub/pasda/PAHydrographyDataset/ as it becomes available.

The BGS started with nine subwatershed areas as test cases for this methodology (Figure 2). The nine HU12s represent an array of physiographic provinces, geologies, land covers, drainage patterns, anthropogenic influences, and existing datasets. The steps used to create these individual datasets are explained below.

1. PREPARE DATA

In order to successfully import data into GRASS, data must have a projected coordinate system with an associated European Petroleum Survey Group (EPSG) code. The BGS uses the North American Datum 1983 State Plane Pennsylvania South FIPS 3702 coordinate system, which has an EPSG code of 3651. The BGS uses ArcMap's Project Raster tool to convert the DEMs into this projection. This tool requires a selection of a resampling technique.

When performing any kind of resampling on a DEM in ArcMap, the BGS selects the Bilinear Interpolation option. For continuous surfaces, the Bilinear Interpolation technique is generally better than the default Nearest Neighbor technique. Projecting a DEM with the Nearest Neighbor technique can result in gridded artifacts (Figure 3). The Cubic Convolution method is another option for avoiding these artifacts; but unlike the Bilinear Interpolation method, it might not keep the resulting raster pixel values within the original DEM's pixel value range.

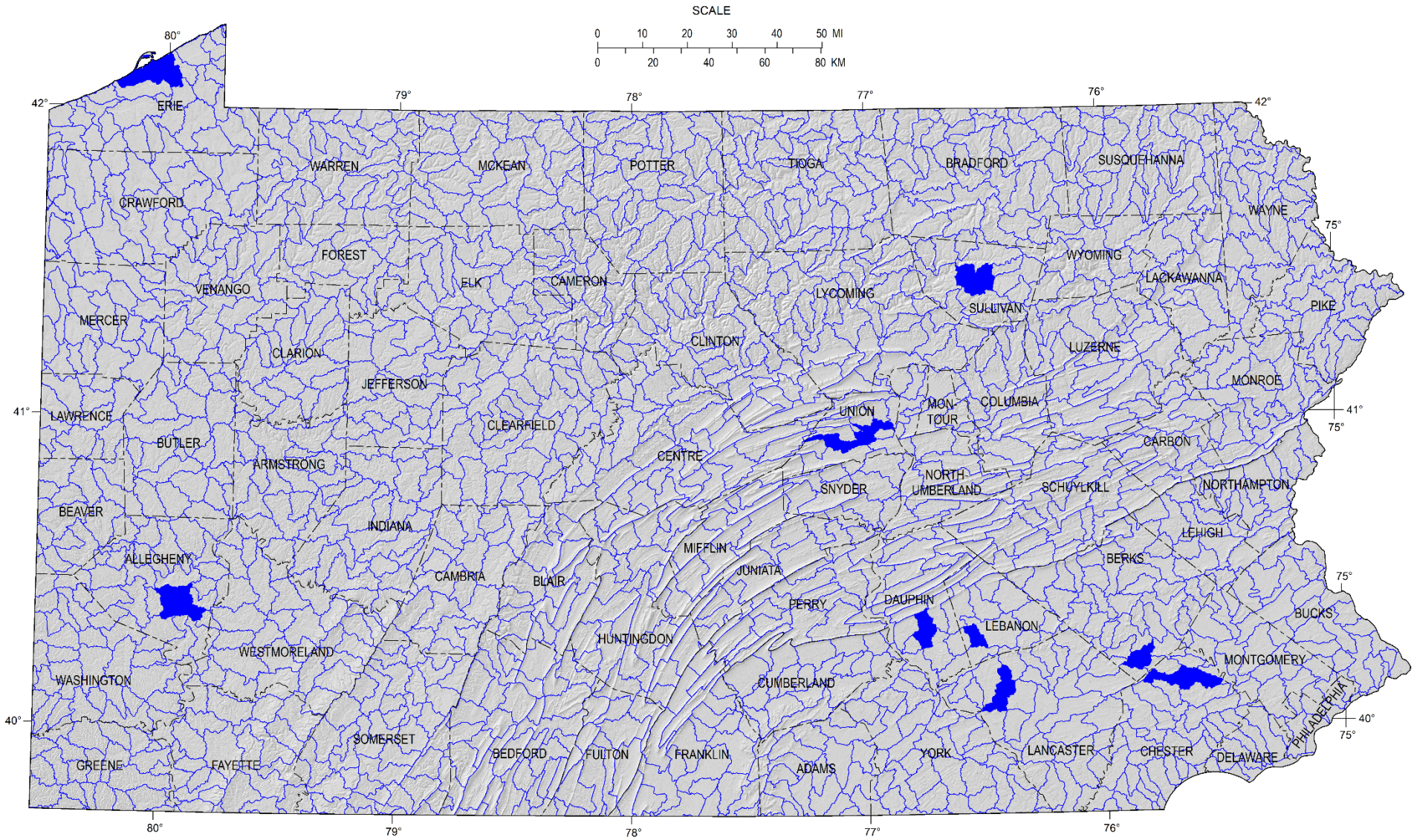


Figure 2. Distribution of HU12 basins in Pennsylvania. Blue lines represent basin boundaries, and blue areas indicate the initial nine basins used to develop the methodology for extracting PAHD flowpaths. The hillshade shows the varied landscape across the state.

Figure 3. The top image shows a hillshade of farmlands in Chester County produced from a DEM that was projected using the Nearest Neighbor resampling technique. The bottom image shows the orthoimage for the area. Comparing the two images confirms that the gridded pattern in the top image is not a realistic representation of ground features—rather it is an artifact resulting from the use of the Nearest Neighbor technique.

2. SET UP GRASS GIS

A GRASS GIS database directory contains Locations, and Locations contain Mapsets (Figure 4). The database directory is essentially a folder that can contain multiple GRASS projects. Locations can be individual projects or, if the project is particularly large or complex, they can be subregions within a project. All data within a Location must have the same coordinate reference system, and this coordinate system must have an associated EPSG code in order to be recognized by GRASS. Mapsets contain GIS data related to their parent Location and can be used to organize the outputs of multiple tasks or subregions.

The first step in preparing data in GRASS is to create a database directory with an appropriate name (e.g., "GrassData"). Before the first Location is created in this folder, a red text warning will appear in the Startup window. Once an initial Location is created, this warning will disappear (Figure 5).

To create a GRASS Location, select "New" under the Startup option "2. Select GRASS Location." GRASS will guide users through a step-by-step process for setting up a Location based on the data's EPSG code.

After a Location is created, select "New" under the Startup option "3. Select GRASS Mapset." Mapsets can be named in reference to a study area, data-analysis technique, or whatever else is logical for any given project. In the context of the PAHD process, Mapsets are being used

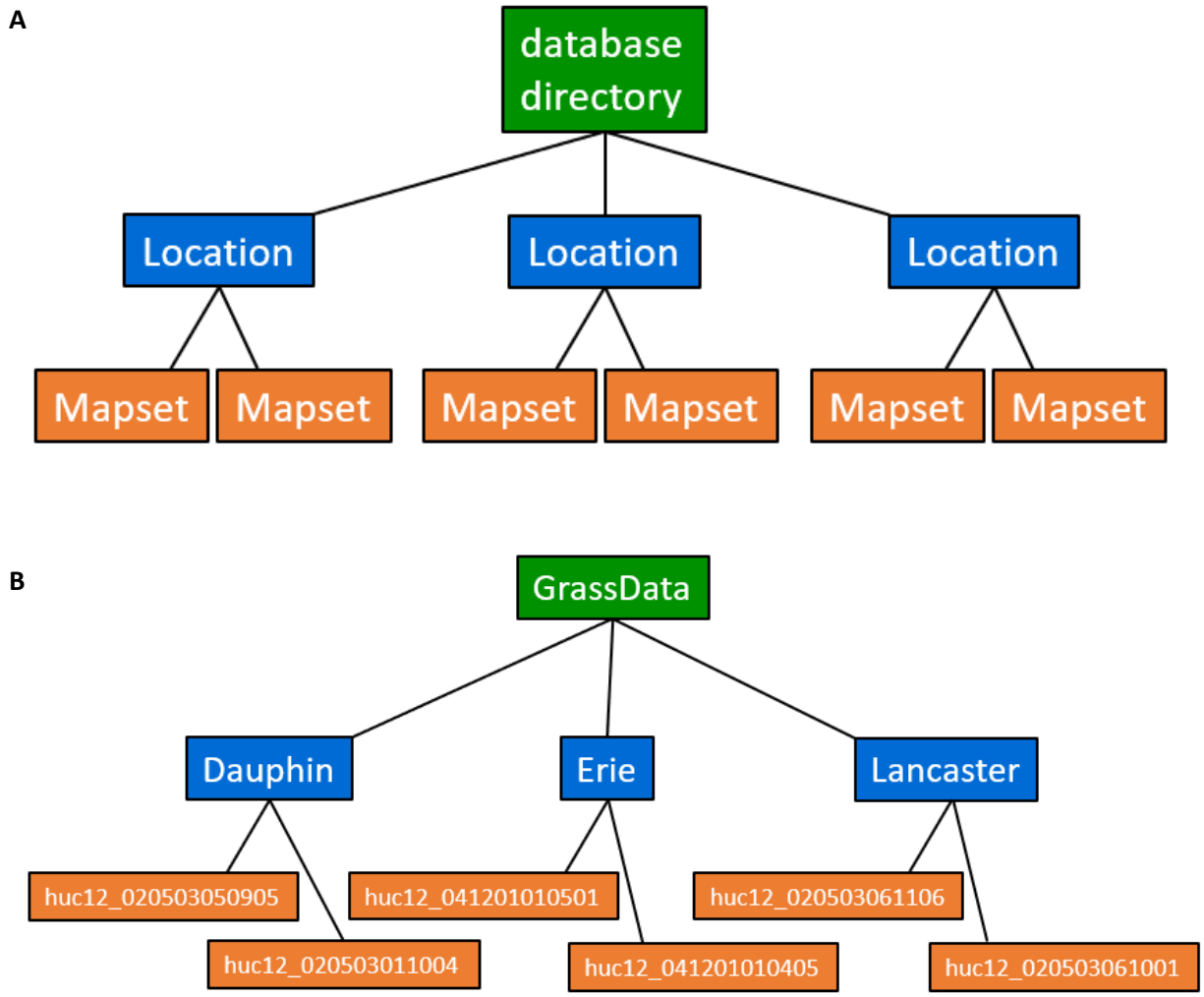


Figure 4. A. Relationships between a GRASS GIS database directory (green), Locations (blue), and Mapsets (orange). **B.** Examples of GRASS GIS project component names used when preparing PAHD flowpaths.

to represent individual HU12s and are identified by the HUC12 number. The example shown in Figure 5B has a Mapset named huc12 #020503050905. It also shows a Mapset named "PERMANENT"—GRASS GIS automatically generates a PERMANENT Mapset for every Location. These Mapsets cannot be deleted, and in the PAHD process, they are ignored.

3. RUN GEOMORPHON ANALYSIS

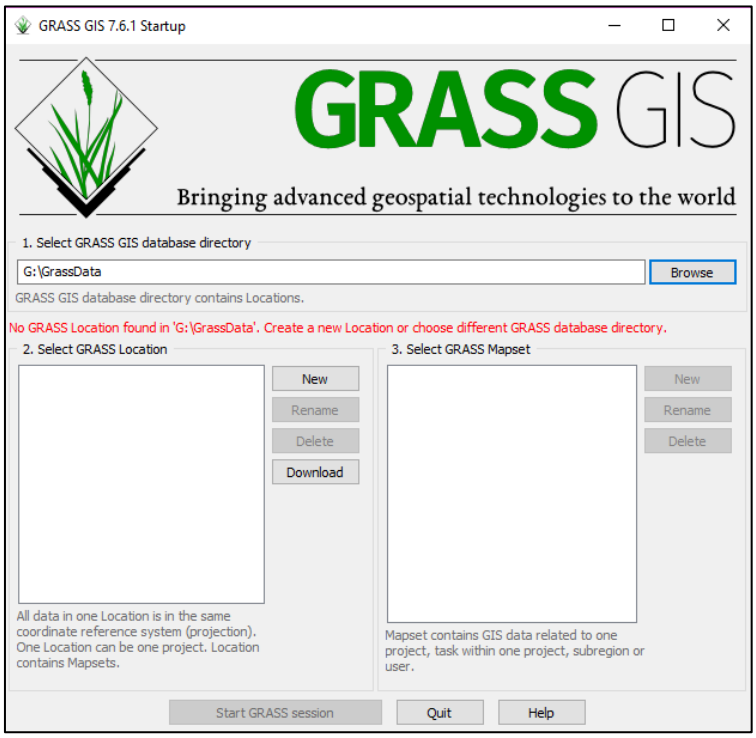
The concept of "geomorphons" (a portmanteau for "geomorphologic phonotype") was first introduced by Stepinski and Jasiewicz in their 2011 paper *Geomorphons—A New Approach to Classification of Landforms*. The authors provided additional information in their paper *Geomorphons—A Pattern Recognition Approach to Classification and Mapping of Landforms* (Jasiewicz and Stepinski, 2013). The central idea is that any surface can be simplified down

to a ternary pattern with a central pixel and eight surrounding pixels. There are 6,561 possible patterns with this three-by-three schematic, but this can be reduced to 10 distinct patterns if redundancies due to rotation or reflection are eliminated (Figure 6). These patterns, or geomorphons, are applied to an elevation raster via classification. In GRASS GIS, the 10 landform classifications are assigned to cells on an elevation surface based on what "shape" (i.e., landform) the elevation surface most closely resembles.

The GRASS geomorphon tool allows for the manipulation of a number of parameters, but this methodology only focuses on the input elevation raster, the outer search radius, and the inner search radius. The outer and inner search radii refer to the distance around a cell that the tool will analyze.

An outer search radius of three, for instance, indicates that the geomorphon tool will examine all neighboring cells within a three-cell radius (Figure 7A). It is important to pay attention to raster cell size: if the raster cells have 1-m radii and an outer search radius value of three, the outer search radius will represent a distance of 3 m; but if the

A



B

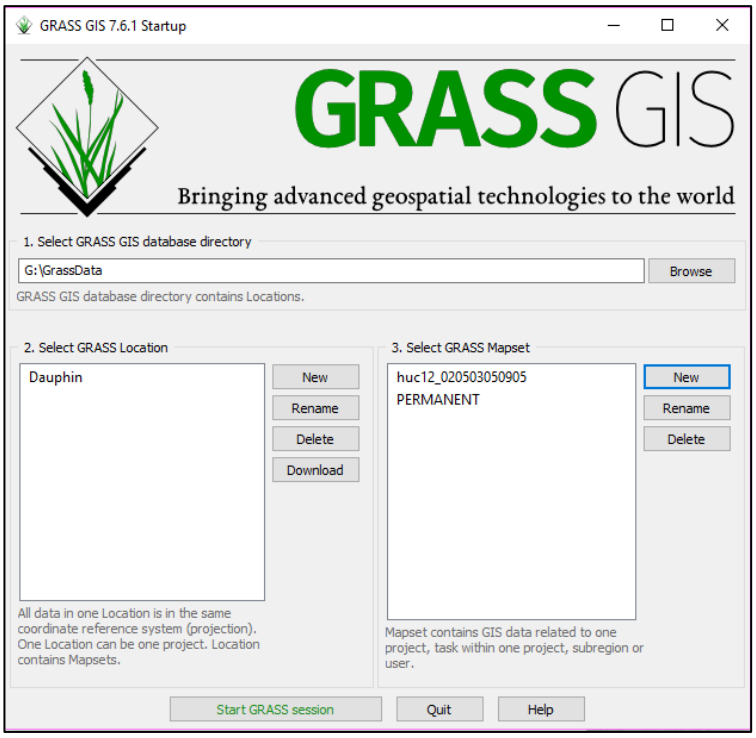


Figure 5. A. GRASS GIS Startup window showing the location of the database directory "GrassData." A red text warning is displayed because there are no Locations assigned to the directory. **B.** GRASS GIS Startup window showing the GRASS database directory "GrassData," the GRASS Location "Dauphin," the default primary GRASS Mapset "PERMANENT," and the user-defined GRASS Mapset "huc12_020503050905."

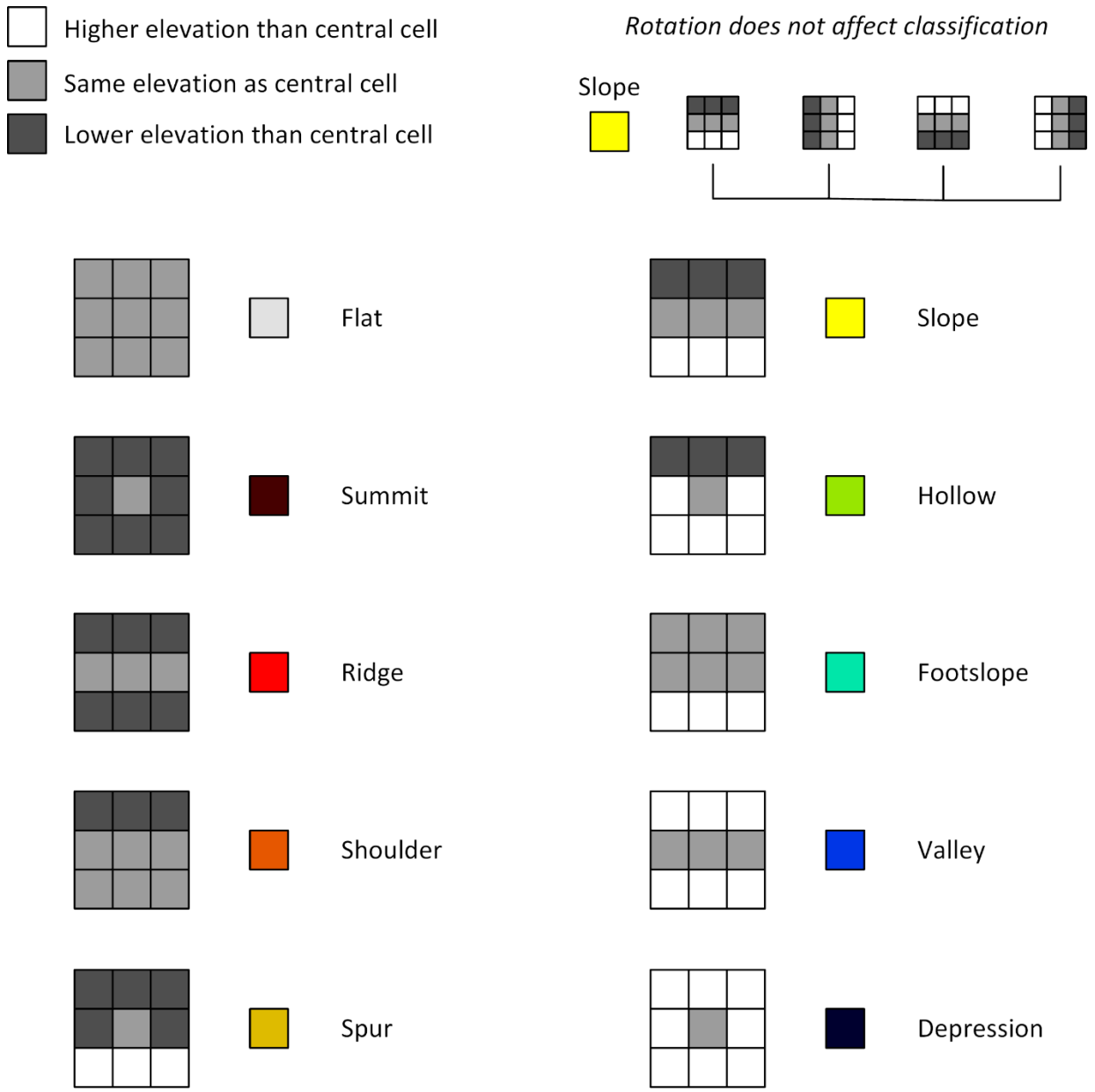


Figure 6. Ten generalized landform interpretations that make up the geomorphon classification system. Note that rotating the patterns does not affect the classification.

raster cells have 10-m radii and an outer search radius value of three, the outer search radius will represent a distance of 30 m.

The default inner search radius in GRASS is zero cells. Changing this value results in the geomorphon tool "skipping over" cells. For example, if the outer search radius is set to three cells and the inner search radius is set to one cell, the geomorphon tool will analyze all neighboring cells within a three-cell radius except for those neighboring cells that fall within a one-cell radius (Figure 7B).

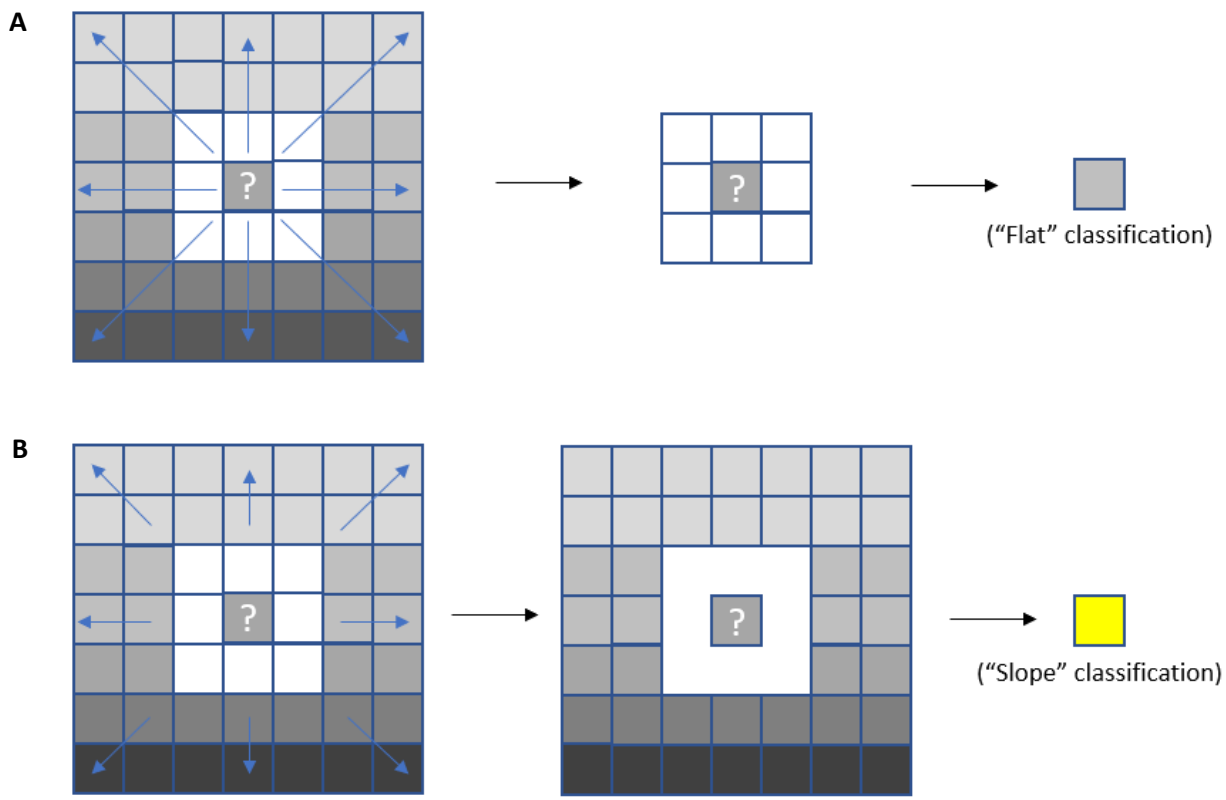


Figure 7. Simplified examples of geomorphon classification analyses and the effects of varying search radii. Boxes represent DEM cells where elevation decreases with increasing darkness (e.g., white is the highest elevation). **A.** A single cell with an outer search radius of three and a default inner search radius of zero. In this example, the eight cells immediately surrounding the central cell all have higher elevation values than the other cells. With line-of-sight analysis, this will result in a "flat" classification for the interior cell. **B.** A single cell with an outer search radius of three and an inner search radius of one. Since the inner search radius of one cell results in the tool "skipping over" the eight cells immediately surrounding the central cell, line-of-sight analysis results in a geomorphon classification of "slope" for the interior cell.

The BGS follows the suggestion of Baker and others (2018) of using geomorphon outputs to "winnow out" noisy or otherwise undesirable data. To do this, the BGS utilizes three geomorphon outputs for each study area: (1) a "channel-extent" geomorphon, (2) a "valley-extent" geomorphon, and (3) a "buffered-channel-extent" geomorphon (Figure 8). The terms "channel-extent" and "valley-extent" are from Baker and others (2018); the buffered-channel-extent is a BGS concept originated for the PAHD process.

Create a Channel-Extent Geomorphon

Resample the DEM

The first step in producing a channel-extent geomorphon is to resample the projected DEM—which should have a cell size of 1 m or less—to a 0.5-m cell size using Bilinear Interpolation in ArcMap. Resampling to a smaller cell size does not improve the resolution of the

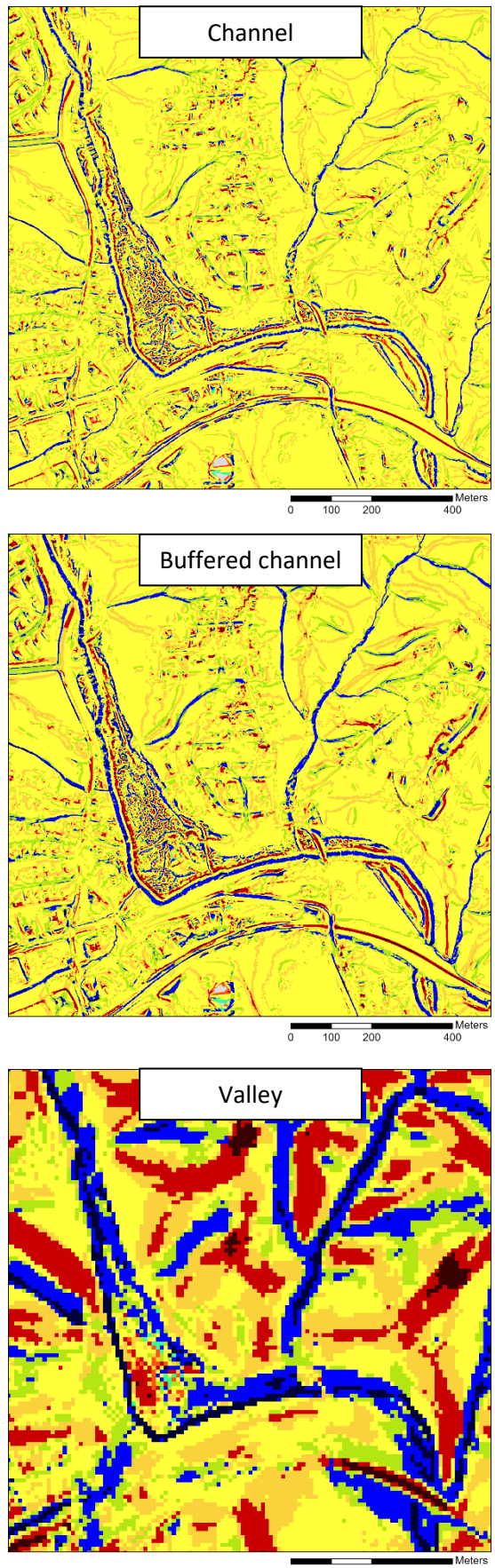


Figure 8. A channel-extent geomorphon, a buffered-channel-extent geomorphon, and a valley-extent geomorphon for part of a selected DEM from Dauphin County. All images show the same spatial extent. The left column features a hillshade and an orthoimage of the same area.

elevation raster, but it does change how the GRASS GIS geomorphon function interprets or classifies individual pixels (Figure 9).

Using a raster with a smaller cell size allows GRASS to make a more detailed geomorphon classification (compare geomorphon outputs in Figures 10A and 10B). While this does not

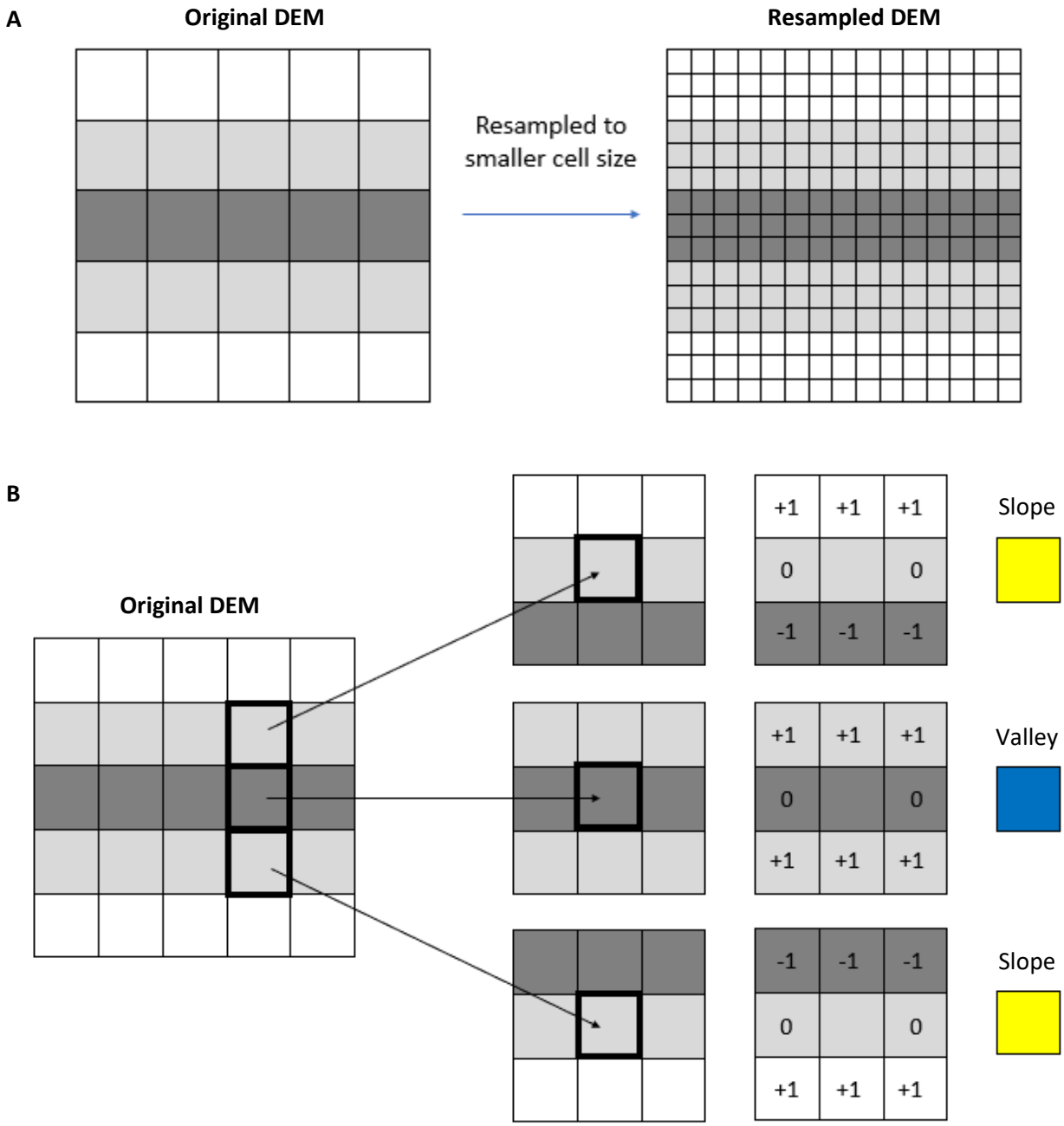


Figure 9. Grids showing how resampling a DEM to a smaller cell size can affect the GRASS GIS geomorphon output. **A.** A DEM resampled to a smaller cell size. **B.** GRASS GIS geomorphon classification of cells in the original DEM. **C.** GRASS GIS geomorphon classification of cells in the resampled DEM. **D.** Comparison of GRASS GIS geomorphon outputs for the original DEM and the resampled DEM.

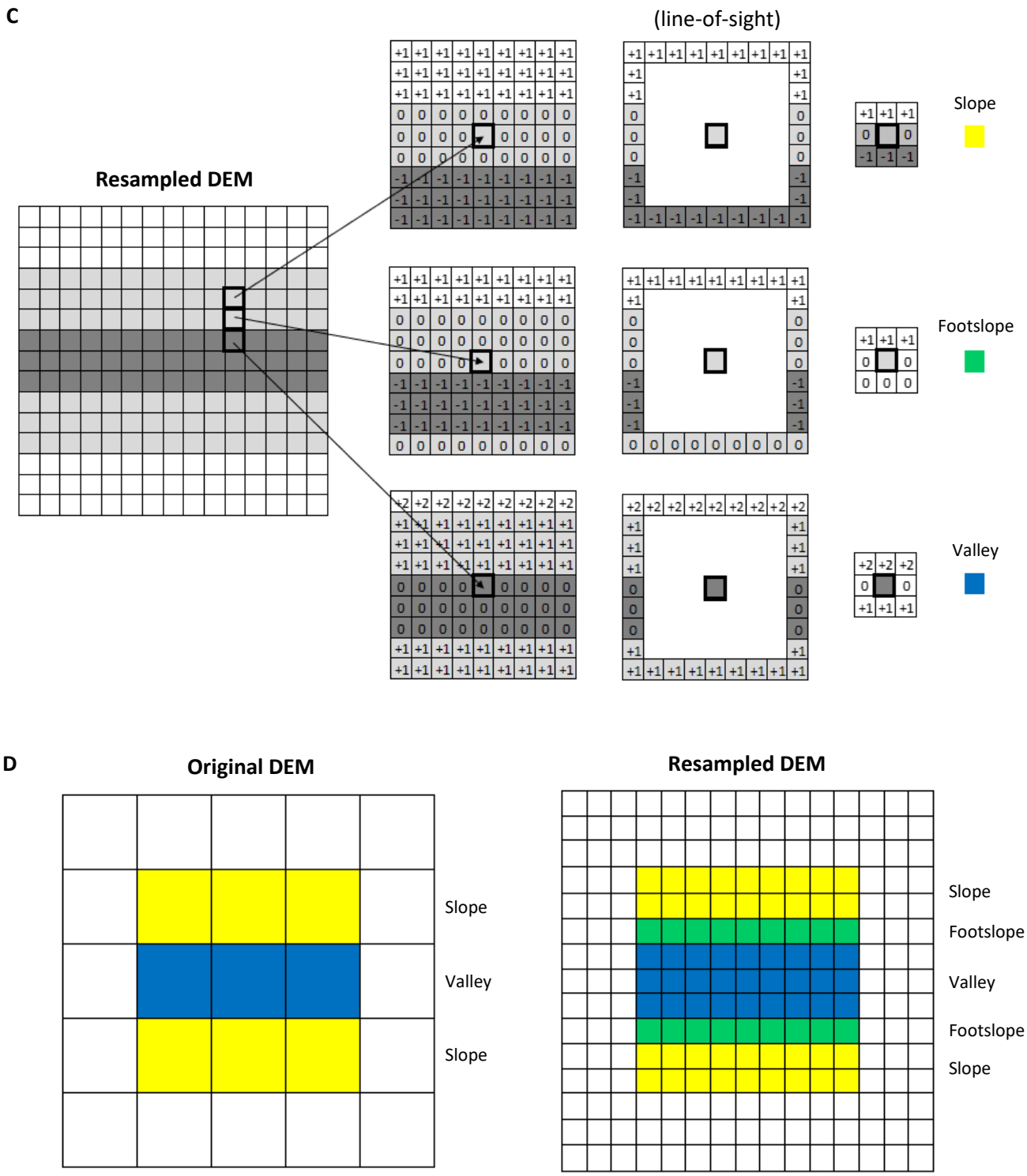


Figure 9. Continued.

introduce "more accuracy" to the original DEM, a smaller cell size ultimately results in geomorphon-derived flowpaths that more accurately reflect the morphology of the water feature they represent (Figures 10C and 10D).

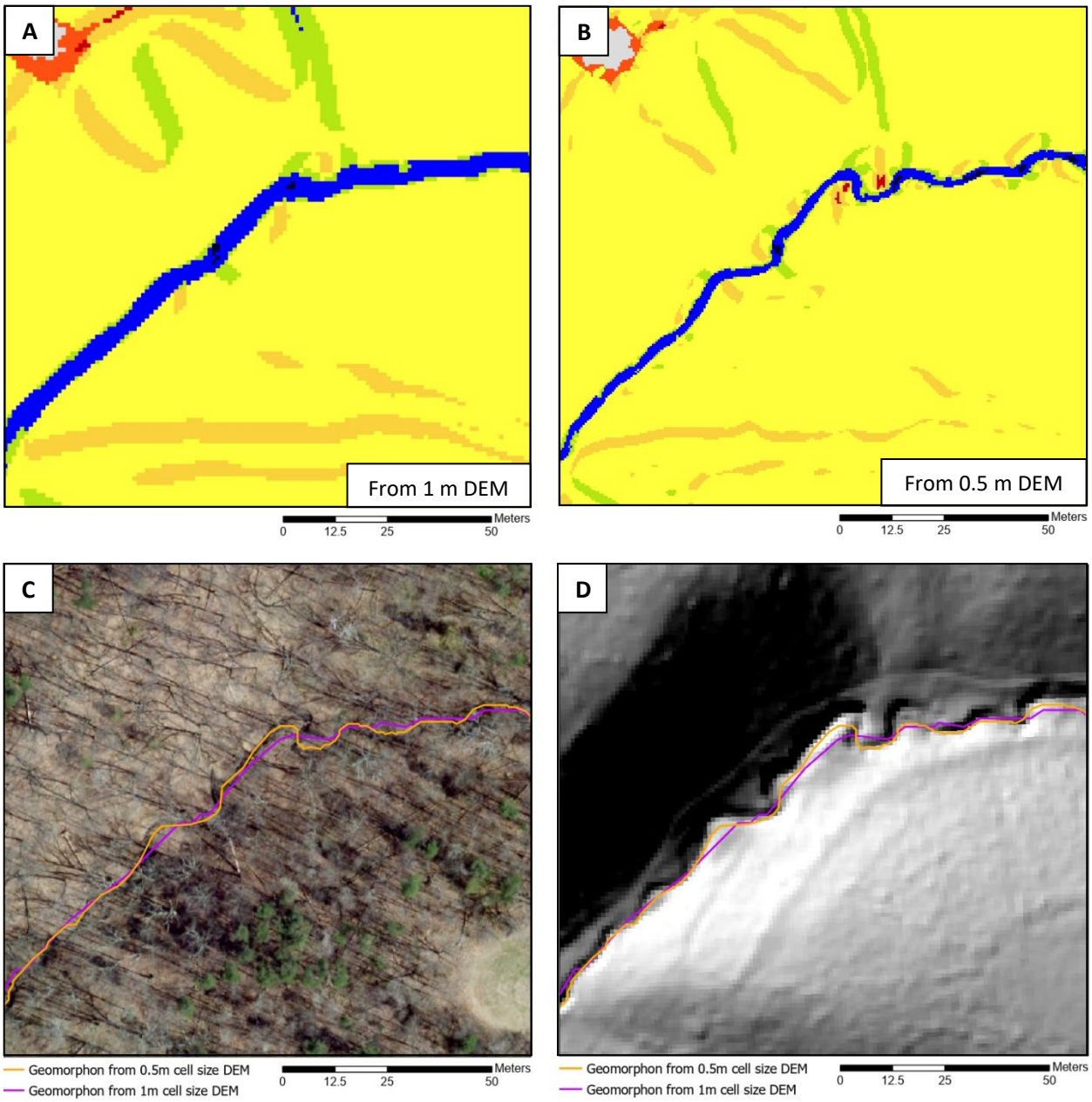


Figure 10. Maps and images showing how a reduced DEM cell size can improve the accuracy of a geomorphon output. **A.** GRASS GIS geomorphon output for a DEM with a 1-m cell size. **B.** GRASS GIS geomorphon output for a DEM with a 0.5-m cell size. **C.** Output vector lines derived using the 0.5-m cell size (orange) and the 1-m cell size (purple) channel-extent geomorphons overlying associated orthoimagery. **D**. The output vector lines from C overlying the associated hillshade, which was produced from the original DEM (1-m cell size).

Filter the Resampled DEM

The second step in producing a channel-extent geomorphon is to smooth the resampled DEM using either the ArcMap Focal Statistics tool with a 3 x 3 neighborhood and a mean statistics type or the ArcMap Filter tool with a low-pass filter. These two processes are effectively the same, and either tool will produce the same result. The tools calculate the average of each cell neighborhood and assign the value to the central cell, reducing the more extreme values in the data, such as those caused by anthropogenic features (e.g., roads), and smoothing noisy data (e.g., floodplains) (Figure 11).

The BGS applies the filter 10 times successively to the resampled DEM. Gaussian filters are known to be better at reducing spurious data in rasters; however, low-pass filters run more quickly than Gaussian filters, and the repeated application of an averaging or low-pass filter approximates the output of a Gaussian filter (Smith, 1999; Lukin, 2007; Kovesi, 2010). While researching the methodology presented here, the author found that filtering HU12-sized elevation rasters more than 10 times did not result in significant improvements to geomorphon outputs (Figure 12); therefore, in order to save on processing time, the number of suggested successive runs was limited to 10.

In preparation for the third step, the filtered DEM is exported in a TIF format to a folder that will be accessed through GRASS. To avoid confusion, the BGS uses an "in" folder and an "out" folder for its GRASS products.

Import the Filtered DEM Into GRASS GIS

In the GRASS GIS Startup window, select the appropriate user-defined Mapset. This will cause the Startup window to close and two new windows will open—the Layer Manager and the Map Display (Figure 13). In the Layer Manager window, select File > Import raster data > Import of common raster formats. This will open the raster import window, from which one can browse to and select the TIF of the filtered DEM. (The window can also be brought up by typing the command "r.in.gdal" in the Command window, which is always open in GRASS GIS.) After selecting the TIF file, the DEM can be added to the Layer Manager and Map Display windows by selecting the "Add raster map layer" option from the ribbon command in the Layer Manager. The raster may not display immediately in the Map Display window, selecting "Zoom to selected map layer(s)" from the ribbon command in the Map Display window will rectify this.

Define the Geographic Region

The fourth step in generating the channel-extent geomorphon is to define the geographic region. Type the geographic region set command (g.region) beside the GRASS command prompt (Figure 14A). This will open the Region Settings window (Figure 14B) where the geographic region for a Mapset can be defined.

Run the Geomorphon Command

The fifth step is to run the geomorphon command. Type "r.geomorphon" into the GRASS command prompt. This will open the window for the Geomorphon Analysis tool. The parameters for running this tool are defined under the Required and Patterns tabs (Figure 15A). For most of the PAHD methodology, the other tabs are ignored; however, the author did find the Command output tab useful in the event of output errors or troubleshooting.

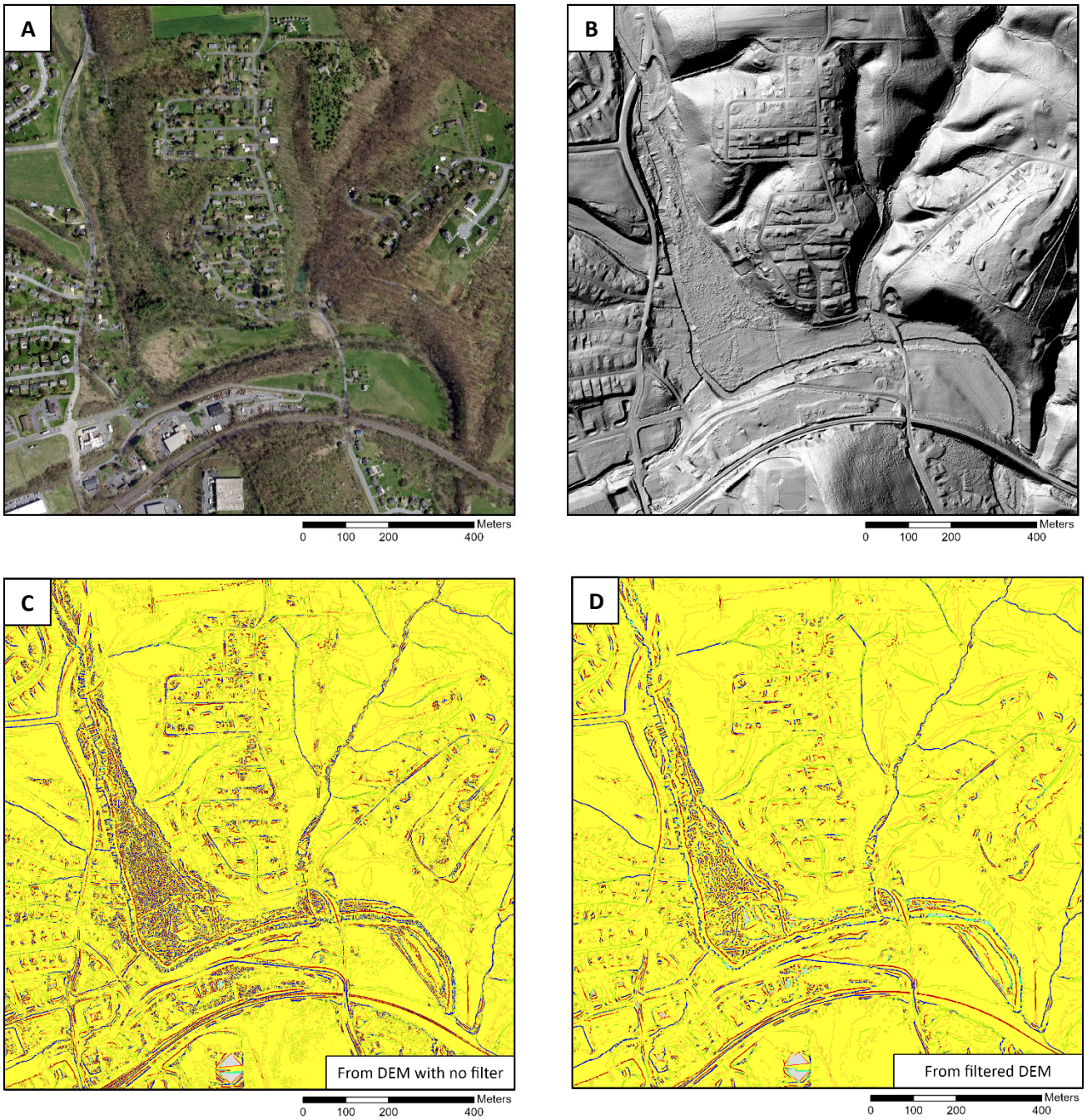


Figure 11. Maps and images illustrating the effectiveness of using a smoothing filter to reduce noise in a geomorphon output of an area in Dauphin County. **A.** Orthoimage of the area. **B.** Hillshade produced from DEM created from QL2 lidar. **C.** Geomorphon output produced from the unfiltered DEM. **D.** Geomorphon output produced from the DEM after a smoothing filter was applied 10 times.

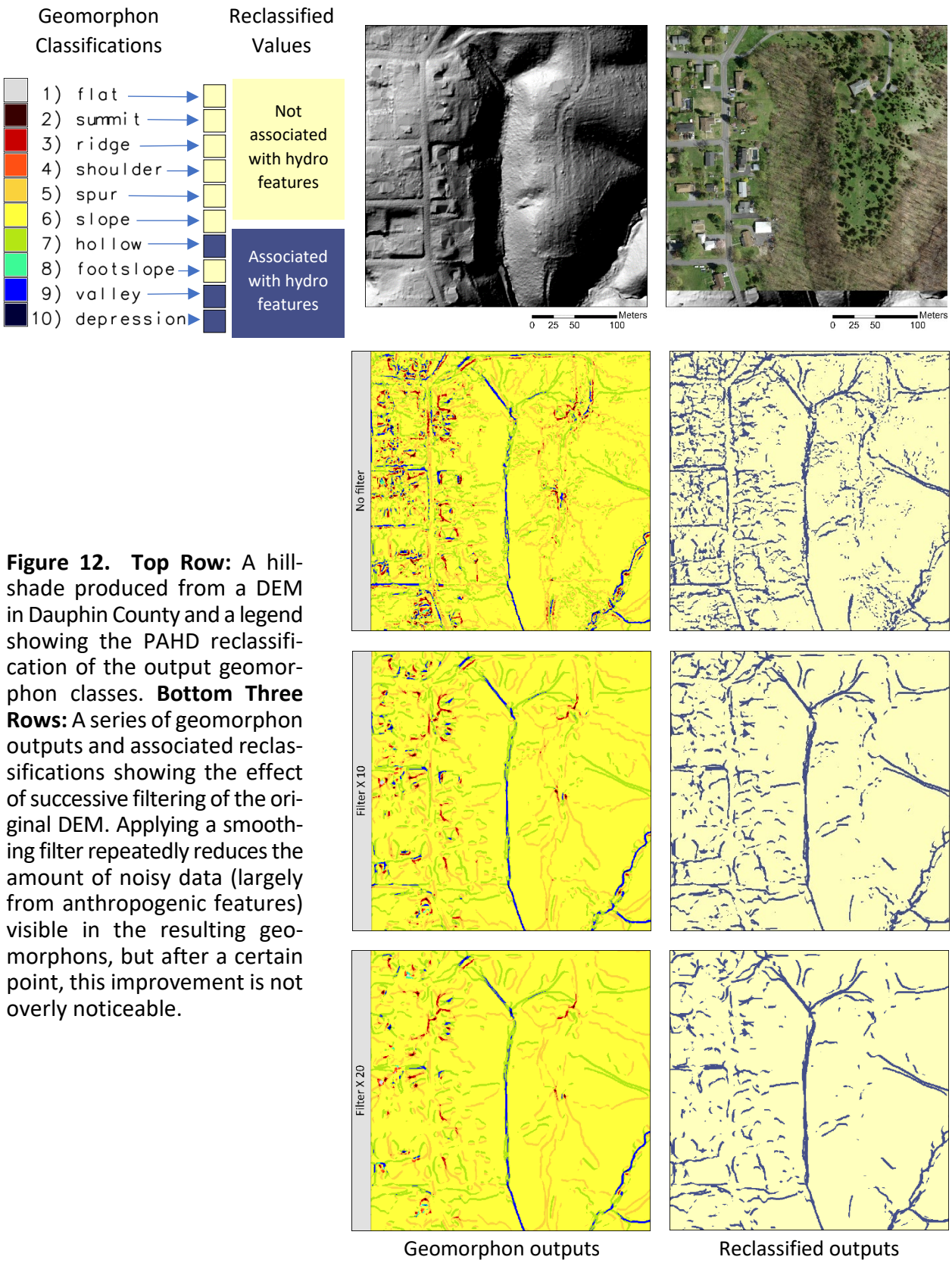


Figure 12. Top Row: A hillshade produced from a DEM in Dauphin County and a legend showing the PAHD reclassification of the output geomorphon classes. **Bottom Three Rows:** A series of geomorphon outputs and associated reclassifications showing the effect of successive filtering of the original DEM. Applying a smoothing filter repeatedly reduces the amount of noisy data (largely from anthropogenic features) visible in the resulting geomorphons, but after a certain point, this improvement is not overly noticeable.

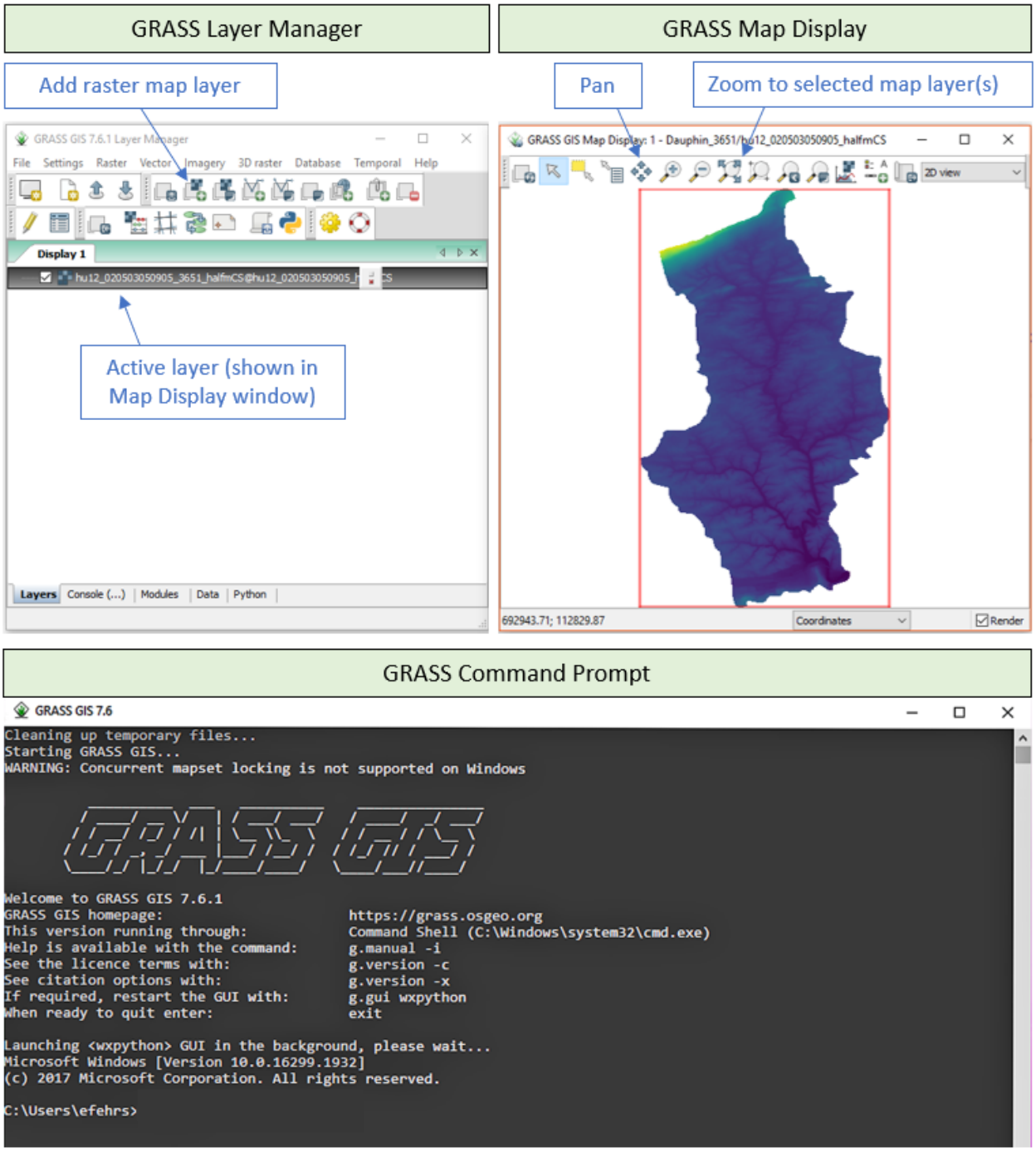


Figure 13. The GRASS GIS Layer Manager and Map Display interfaces.

The Required tab is where all the tool parameters are entered. The entries used by the BGS for a channel-extent geomorphon are the input elevation raster that was created in the previous steps, an outer search radius of 20 cells (10 m), and an inner search radius of 0 cells. The other values (flatness threshold and flatness distance) are left to their default settings.

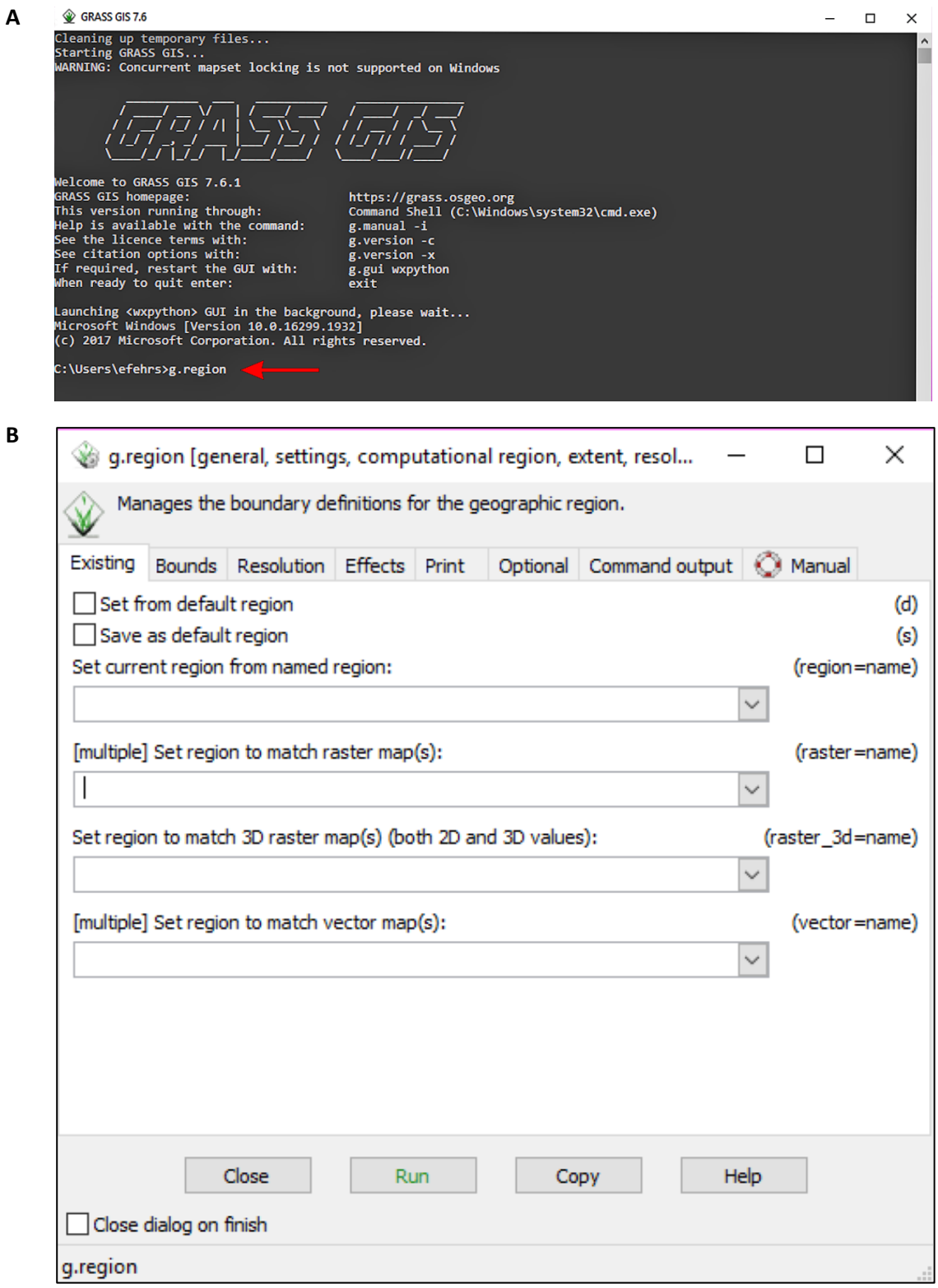
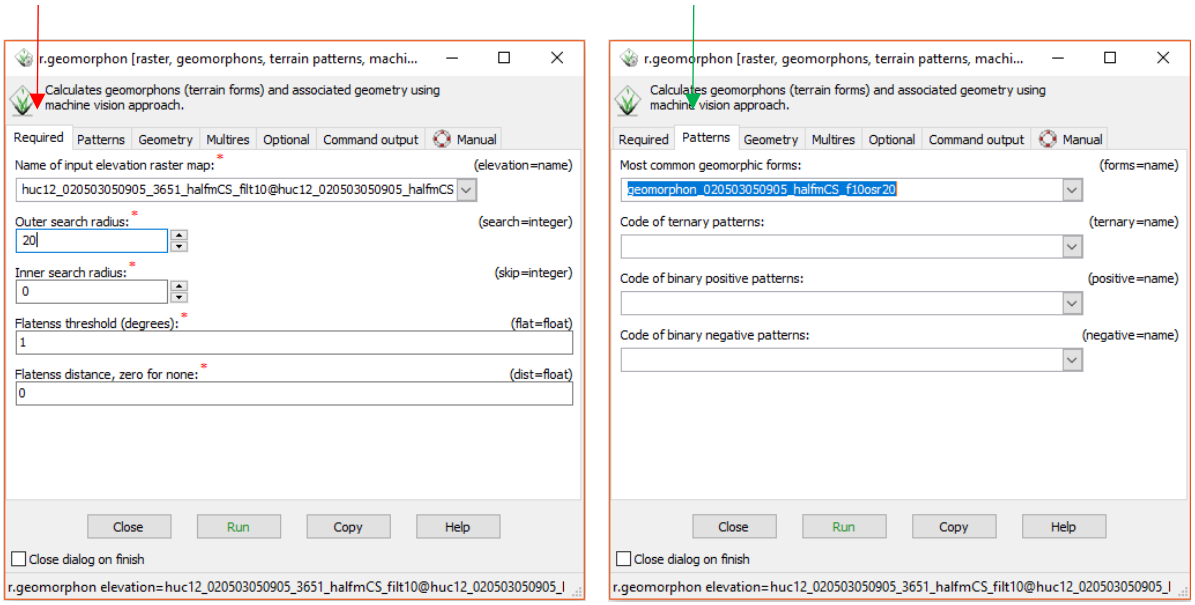


Figure 14. A. GRASS GIS Command Prompt window showing the command "g.region." This command is used to open the Region Settings window. **B**. GRASS GIS Region Settings window. This is where the geographic region for any given Mapset gets defined.

A



B

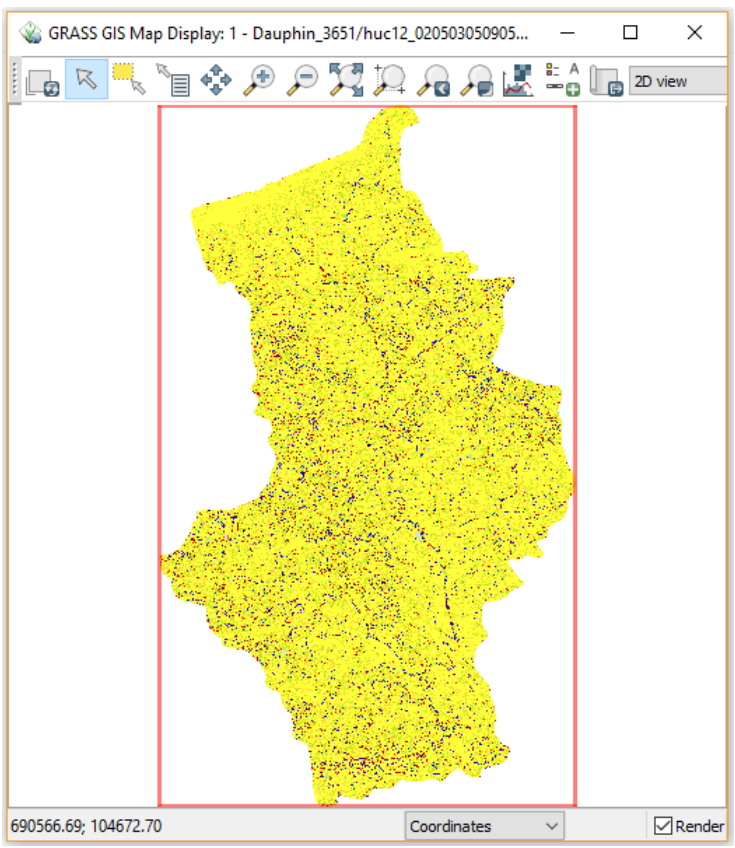


Figure 15. A. Example of inputs for geomorphon analysis. Any time the geomorphon tool is run in GRASS, the following parameters must be specified under the "Required" tab (red arrow): an input elevation raster, an outer search radius, and an inner search radius. The name of the output geomorphon classification is entered under the "Patterns" tab (green arrow). **B.** Example of a geomorphon output.

The name of the output geomorphon classified raster is set under the Patterns tab. It is considered best practice to choose a name that reflects the original raster's location (i.e., HUC12 #), resolution (cell size), any manipulations preprocessing, the outer-search-radius value, and the inner-search-radius value (if not the default). For example, "geomorphon_

020700040605_halfmCS_f10osr20" indicates a geomorphon output encompassing the HUC12 #020700040605 area, which has been generated from an elevation raster with a 0.5-m cell size, filtered 10 times, and analyzed by the geomorphon tool using a 20-cell (10-m) outer search radius and 0-cell (0-m) inner search radius.

The use of an outer search radius of 10 m and inner search radius of 0 m to generate a channel-extent geomorphon are taken from Baker and others (2018, p. 14).

Export the Raster to an Output Folder

Once the geomorphon tool finishes processing, the r.geomorphon window should be closed. The geomorphon output can be viewed by selecting the "Add raster map layer" button from the Layer Manager window (keyboard shortcut Ctrl + Shift + R) and choosing the geomorphon output from the dropdown menu that appears. This adds the geomorphon to the Map Display (Figure 15B). In order to make geomorphon outputs available for analysis in other programs, they must be exported to a predetermined folder. After adding the raster map layer in GRASS, it can be exported by selecting the Layer Manager options File > Export raster map > Common export formats (or command "r.out.gdal").

Create a Valley-Extent Geomorphon

The goal of creating a valley-extent geomorphon classification is to produce a raster that targets valley bottoms. At a later point, these valley bottoms will be used to constrain (using masking techniques) the channel-extent geomorphon classes. This will effectively discard channel data that is not contained within a valley bottom, which includes the majority of the inaccurately classified anthropogenic features.

Resample the DEM

The scale of valley features is generally larger than that of channel features, and therefore a DEM with a relatively small cell size (e.g., 0.5 m) is not necessary to produce a valley-extent geomorphon. Resampling the unfiltered DEM to a larger cell size reduces the processing time of the geomorphon tool. The BGS resamples the DEM for the valley-extent geomorphon to a cell size of 10 m. At this scale, successive filtering prior to geomorphon processing is unnecessary.

Define the Geographic Region

The DEM input is now a different cell size than the channel-extent geomorphon (10 m as opposed to 0.5 m) and this necessitates rerunning the g.region tool (explained under "Create a Channel-Extent Geomorphon," page 24). If the user does not redefine the geographic region using the new 10-m cell size DEM in the Regions Setting window, an error or warning will be returned when the geomorphon function is performed.

Run the Geomorphon Command

After setting the 10-m cell size DEM as the new geographic region, the r.geomorphon tool should run successfully. In their work generating valley-extent geomorphons, Baker and others (2018) described using an outer search radius equivalent to 200 m and an inner

search radius equivalent to 20 m; in researching the current PAHD methodology, better results are achieved with an outer search radius equivalent to 100 m (10 cells) and an inner search radius equivalent to 20 m (2 cells).

Export the Raster to an Output Folder

After the geomorphon tool finishes processing, name the geomorphon-classified raster and export it to the same folder as the previously created channel-extent raster following the instructions under "Create a Channel-Extent Geomorphon," page 30).

Create a Buffered-Channel-Extent Geomorphon

The use of a buffered-channel-extent geomorphon is specific to the current PAHD methodology. The concept of this classification is similar to the channel-extent geomorphon; however, instead of targeting only channel-type features with specific geomorphon classes (footslope, valley, depression), those geomorphon classes more associated with channel and bank morphology (hollow, valley, depression) are extended "up" the associated banks of the channels by using a larger outer search radius (Figure 16).

The BGS generates the buffered-channel-extent geomorphon using the same filtered DEM that served as input for the generation of the channel-extent geomorphon but with the following geomorphon tool parameters: an outer search radius of 30 cells (15 m) and a default inner search radius of 0 cells (0 m). The buffered-channel-extent product helps to filter out much of the noisy data associated with floodplains that would not have been filtered out with the valley-extent geomorphon alone. The buffered-channel-extent geomorphon also improves cell connectivity compared to the channel-extent geomorphon and facilitates the ArcScan editing described later in this workflow.

4. RECLASSIFY THE GEOMORPHON OUTPUTS

From this point forward, the BGS uses ArcGIS to process the geomorphon outputs. The first step in processing the channel-extent, valley-extent, and buffered-channel-extent geomorphons is to reclassify each of the rasters. The individual classes that are associated with hydrography features will vary slightly from one geomorphon output to another.

Channel-Extent Geomorphon

The channel-extent geomorphon raster is the only geomorphon output that will not be edited with ArcScan tools. The BGS uses the ArcMap Reclassify tool to reassign the values of the classes 8/9/10 (footslope/valley/depression) to 1. All other values are designated as NoData. This reclassification extracts all cells associated with channel (flow) features.

Valley-Extent and Buffered-Channel-Extent Geomorphons

In the PAHD methodology, the valley-extent and buffered-channel-extent geomorphons serve to limit or constrain the cells associated with flow features in the channel-extent geomorphon; as such, these "limiting" rasters need to undergo additional edits using ArcScan.

Before the geomorphon rasters can be edited inArcScan, they must be reclassified so that there are two main classes: cells potentially associated with hydrography features and cells not likely to be associated with hydrography features. Both geomorphon outputs are reclassified so that the classes 7/9/10 (hollow/valley/depression) equal 1 and all else equals 0.

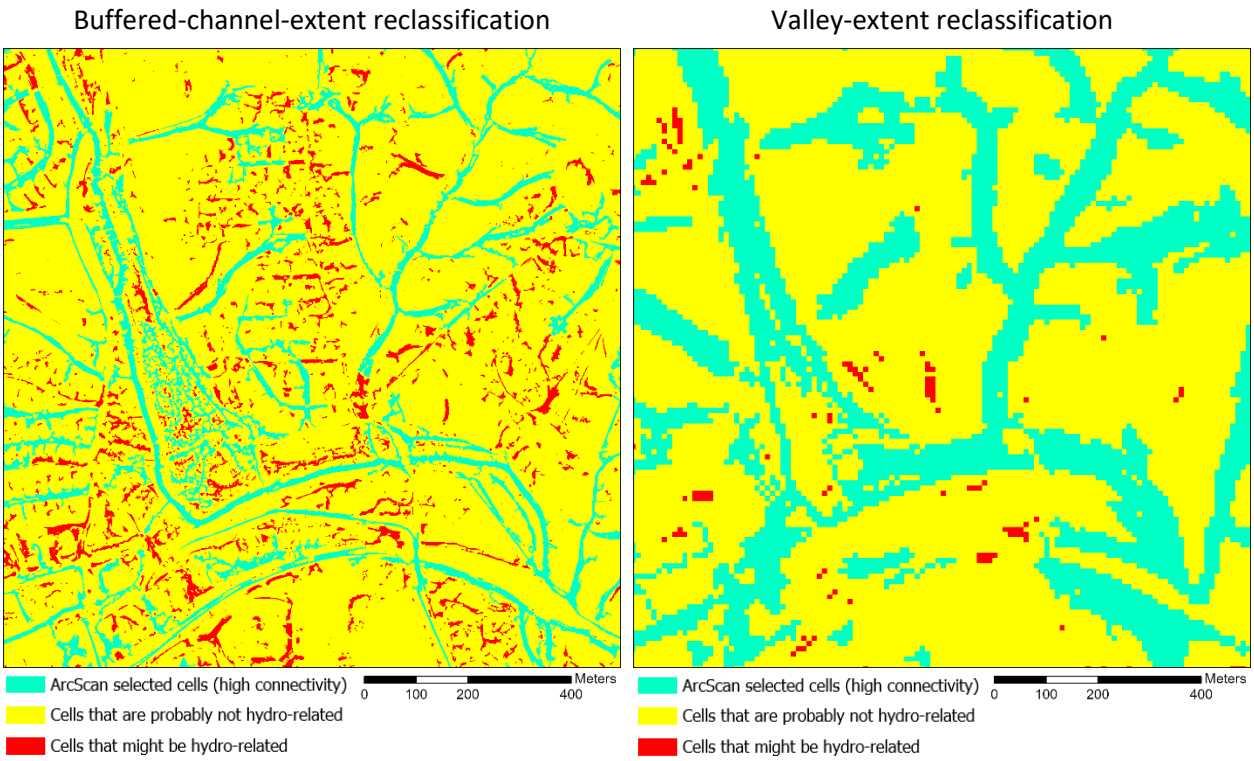


Figure 16. Reclassified buffered-channel-extent (left) and valley-extent (right) geomorphon outputs. Based on the analysis, the yellow cells represent areas that are likely not related to hydrography, and the red cells represent areas that are likely hydrography features. The cyan cells are selected areas from the red, hydrography-related cells that are meant to highlight cell clusters with greater connectivity. In the buffered-channel-extent geomorphon output, touching hydrography-related cells in groups of 5,000 or more are highlighted in cyan. In the valley-extent geomorphon output, touching hydrography-related cells in groups of 10 or more are highlighted in cyan. Any red cells that are not highlighted will be considered "noisy" data, likely the result of anthropogenic features.

5. REMOVE SUPERFLUOUS RASTER CELLS

The following edits are intended to maintain the central network of valley bottoms and buffered channels as represented by their respective geomorphon outputs while at the same time excluding superfluous features. This is exemplified in Figure 16, which shows the results of an ArcScan edit on the reclassified buffered-channel-extent and valley-extent geomorphons. In the edited rasters, the cyan (selected) cells represent pixels that are classified as potential hydrography features *and* maintain a certain degree of connectivity. Red cells represent those pixels that are classified as potential hydrography features *but do not* main-

tain a certain degree of connectivity. These remaining red cells should now be considered unrelated to hydrographic features and need to be edited out.

Valley-Extent and Buffered-Channel-Extent ArcScan Edits

Cell connectivity in the reclassified valley-extent geomorphon can be highlighted by opening the ArcScan toolbar and selecting Cell Selection > Selected Connected Cells, and entering the following (mostly default) settings:

- Total area of connected cells
- Foreground
- Greater than 10 raster pixels
- Create a new selection

This selects all the cells that potentially represent hydrographic features and are clumped in groups of more than 10 cells.

Once the cells are highlighted, the selection should be examined to make sure that all the included features are appropriate. There may be some small patches of pixels that do not appear to be related to hydrographic features. Where possible, any features that are obviously not flowpath related (e.g., buildings and railroads) should be removed from the selection. After unselecting these unrelated or irrelevant cell groups, the remaining cell selection should be saved and reclassified so that there is only one class.

The ArcScan toolbar can now be used to edit the buffered-channel-extent geomorphon in the same manner. The only difference between the two geomorphon editing sessions is that instead of selecting cells in groups greater than 10, the BGS selects cells in groups greater than 5,000 for the reclassified buffered-channel-extent geomorphon. The saved cell selections for the valley-extent and buffered-channel-extent geomorphons can now be used to constrain the reclassified channel-extent geomorphon (Figure 17).

Constrain the Channel-Extent Geomorphon

Masking techniques using the cell selections from the previous section make it possible to limit the resulting geometries of the reclassified channel-extent geomorphon. To do this, the BGS runs the Extract By Mask tool twice:

- For the first run, the reclassified channel-extent geomorphon is used as the input raster and the valley-extent geomorphon's selected cells are used as the confining/limiting raster. Under Environment Settings > Raster Analysis > Cell Size, "Minimum of Inputs" is selected. This extraction isolates the channel features within the valley bottoms, which represent generalized floodplains.
- The output of the previous step is then entered as the Extract By Mask tool's input raster and the buffered-channel-extent geomorphon's selected cells are entered as the confining/limiting raster. Since these inputs have the same cell size, it is not necessary to alter the Environment Settings under Raster Analysis.

The result of these steps is that the cells from the channel-extent geomorphon that are potentially associated with hydrography features are now confined within valley bottoms

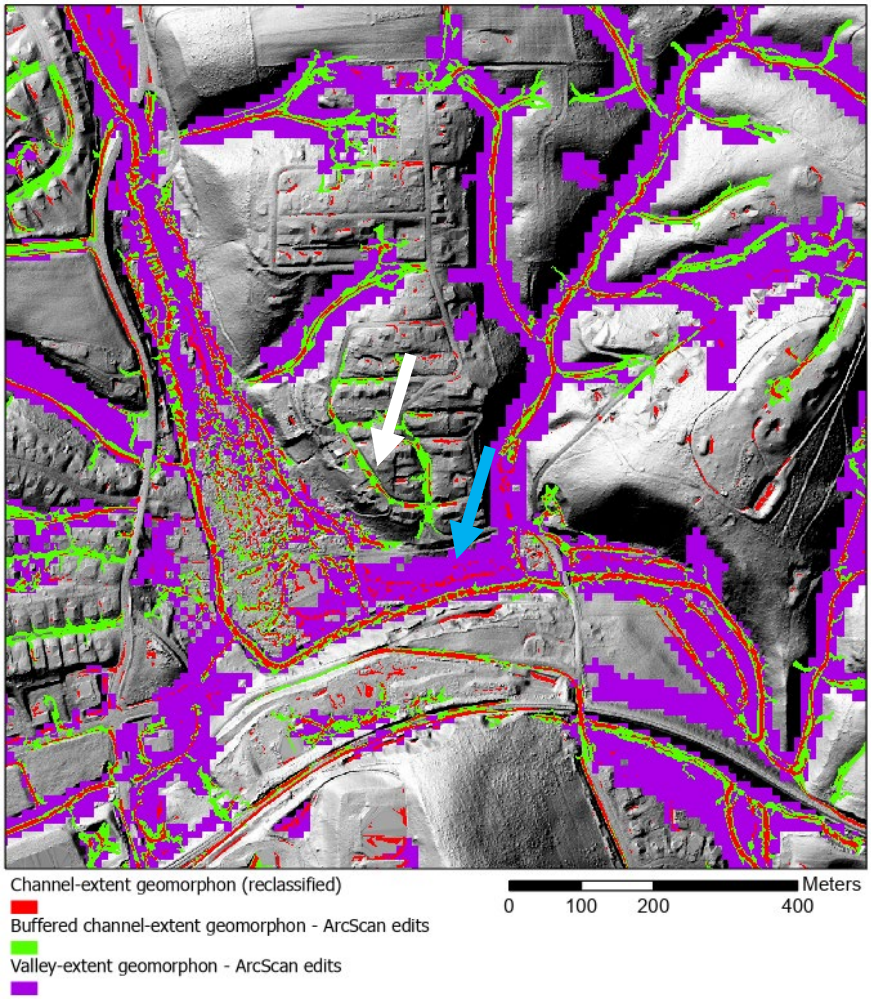


Figure 17. Reclassified channel-extent geomorphon output overlying the cell selections from the buffered-channel-extent and valley-extent geomorphon outputs shown in Figure 16. These outputs will be used to constrain one another with the intent of filtering out superfluous data. For instance, the valley-extent geomorphon selection excludes some anthropogenic features like buildings and roads picked up in buffered channels (see white arrow), whereas the buffered-channel-extent geomorphon selection excludes much of the extraneous data picked up in the floodplain (see blue arrow).

(therefore excluding the majority of anthropogenic features) and do not include the noisy data that is often picked up by GRASS's geomorphon tool in floodplains (Figure 18).

This is the extent of the flowpath editing prior to vectorization. ArcMap's Thin tool is now used to simplify this final constrained channel raster, and then ArcMap's Raster To Polyline tool can be used to convert the thinned raster into a vector (polyline) suitable for topology editing (Figure 19).

6. MANUALLY DELINEATE BREAKLINES

Breaklines need to be delineated for flow features where the vectorized flowpaths produced from the previous steps fail (usually for flow features wider than about 10 m; see Figure 20). For features with widths greater than or equal to 30 m, breaklines were to have been included with the lidar deliverables—as such, these breaklines do not need to be manually delineated. The BGS manually digitizes breaklines for features with widths between 10 m and 30 m.

Once breaklines are digitized, ArcMap's Collapse Dual Lines to Centerline tool should be used to generate a centerline (Figure 21A). The breaklines will not be used additionally in

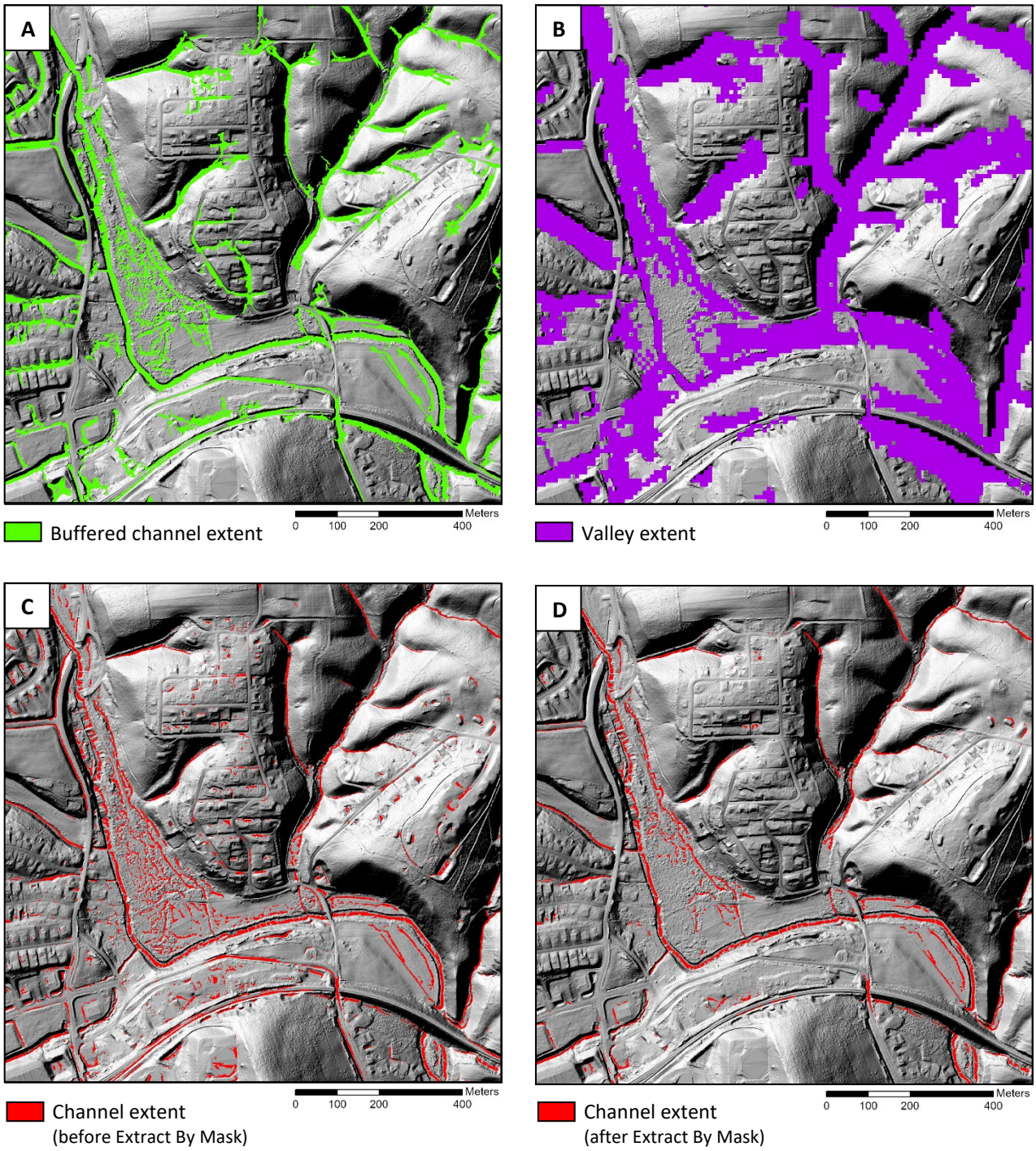


Figure 18. Segregated geomorphon outputs from Figure 17 and the resulting output of the Extract By Mask function. **A.** Hydrography-related cells with connectivity greater than 5,000 cells from the buffered-channel-extent geomorphon. **B.** Hydrography-related cells with connectivity greater than 10 cells from the valley-extent geomorphon. **C.** Hydrography-related cells from the channel-extent geomorphon. **D.** The result of using the Extract By Mask function on the channel-extent geomorphon (18C), first with the valley-extent selection (18B) and then additionally with the buffered-channel-extent selection (18A). The degree to which data have been filtered out can be observed by comparing the original channel-extent output (18C) with the masked channel-extent output (18D).

Figure 19. Vectorized geomorphon product created from the masked channel raster shown in Figure 18D using ArcMap's Thin and Raster To Polyline tools.

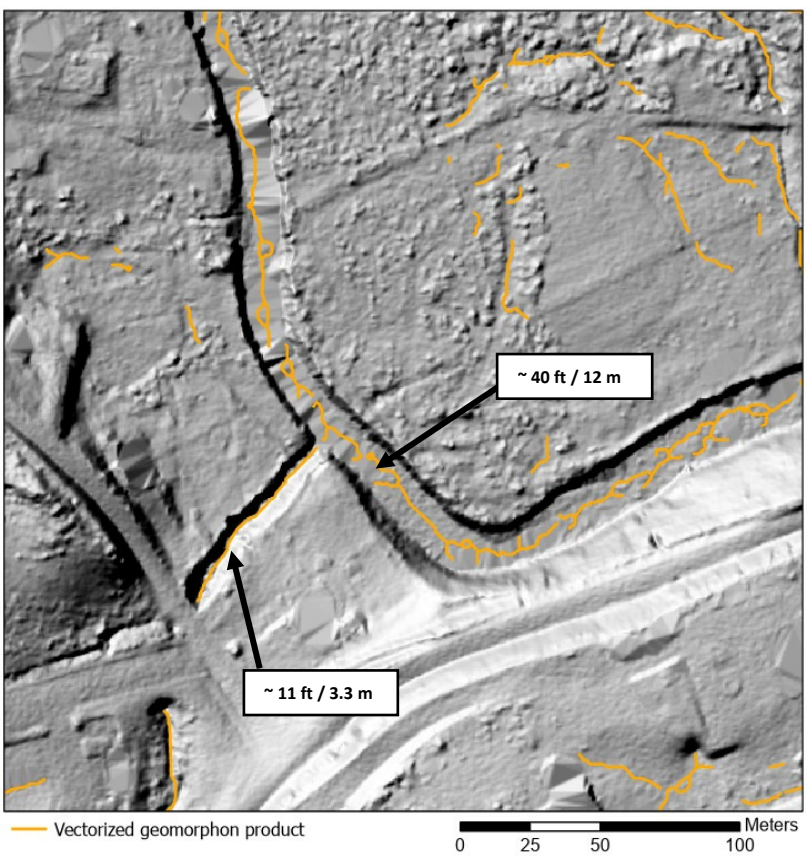


Figure 20. Example of how vectorized flowpaths from geomorphon outputs fail in wider flow features. More specifically, the geomorphon outputs do not produce cohesive flowpaths for features greater than about 10 m wide.

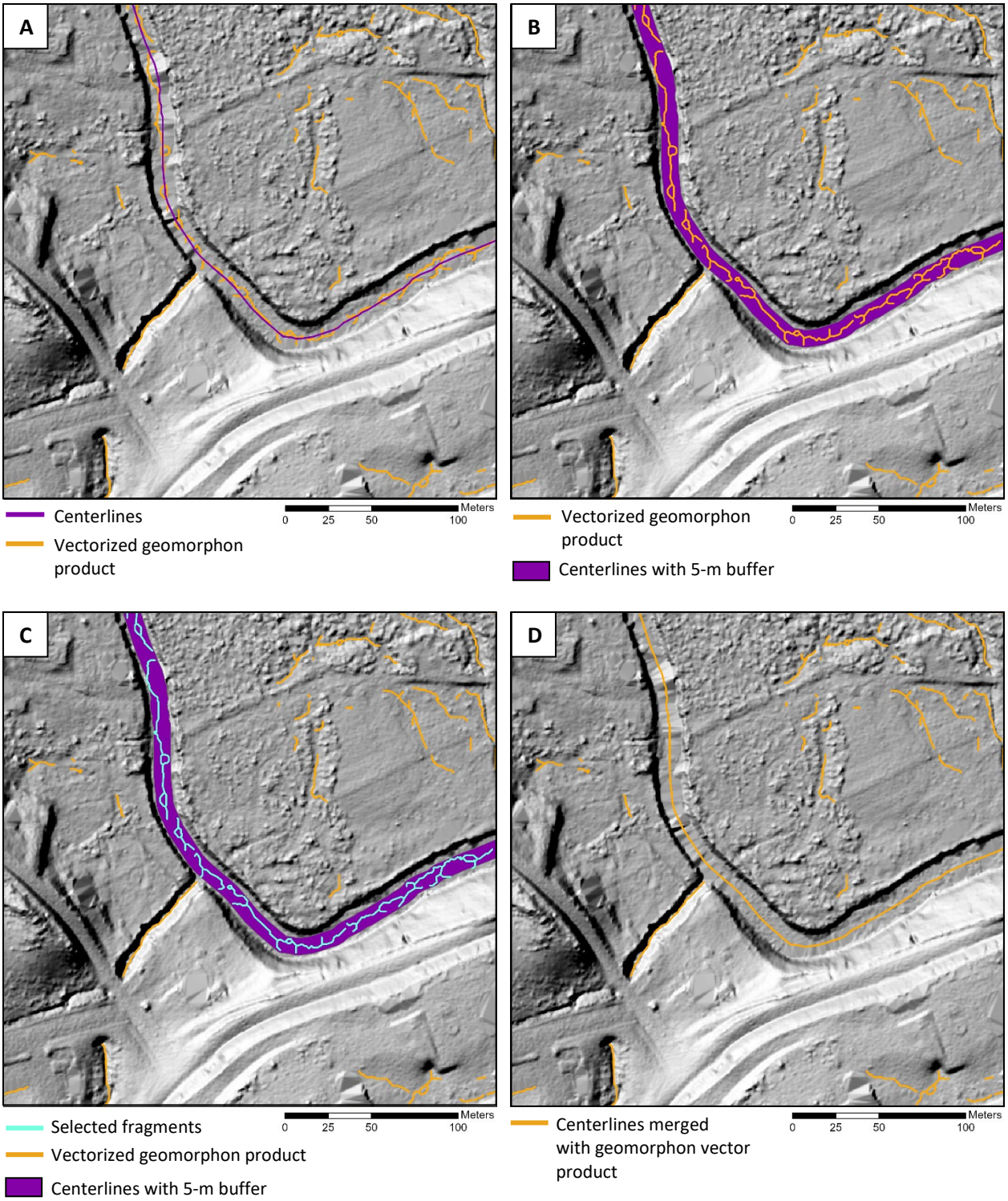


Figure 21. Example of how failed geomorphon flowpaths in wider flow features can be improved. **A.** A centerline is produced by manually digitizing breaklines and collapsing them. **B.** Centerline is buffered by 5 m in order to encompass the fragmented geomorphon flowpaths. **C.** Select By Location is used to highlight all the fragmented lines associated with wider flow features, and these fragments are then deleted. **D.** The remaining flowpaths are merged with the centerline.

deriving flowpath geometries, but they will prove useful later for hydro-flattening and hydro-enforcement purposes. Minor editing of the centerlines is sometimes necessary in areas where the tool fails. All centerlines—those manually delineated, and those provided with lidar deliverables—should be merged into one layer.

The fragmented line segments that appear in "failed" flow features (Figures 20 and 21) need to be removed. This can be accomplished by buffering the generated centerlines by 3–5 m and choosing Selection > Select By Location in order to identify where the geomorphon flowpaths overlap the buffered centerlines (Figure 21C). After checking that none of the well-defined flowpaths are highlighted, the selected segments can be deleted. This will likely not remove all of the smaller fragmented segments from wide flow features, but it will prevent any of these smaller fragmented segments from intersecting the original centerline. The centerline layer should now be merged with the remaining flowpath segments (Figure 21D).

7. USE LAND COVER DATA TO REMOVE UNNECESSARY FEATURES

This section is an optional component to the overall PAHD methods for generating flow-path geometries, but it can significantly reduce the amount of time spent doing manual edits and later QA/QC. The concept here is to use data for pervious surfaces (e.g., shrubland, low vegetation) to select flowpath segments that are likely to represent existing flow features.

The first step is to acquire land cover data that have a resolution of 1 m or better. A good example of such data is the Chesapeake Conservancy's 2013/2014 Land Cover dataset, which is available for public download. Unfortunately, this dataset does not cover the entirety of Pennsylvania. Two of the initial eight study areas for this project, an HU12 in Erie County and an HU12 in Allegheny County, do not have the Chesapeake Conservancy's land cover data.

Having acquired land cover with an appropriate resolution, the next step is to reclassify the data to "pull out" those land cover types that are more likely to include flow features, that is, land covers that belong to pervious/permeable classes. In the Chesapeake Conservancy's 2013/2014 dataset, for example, classes are assigned the following values:

Water	1
Wetlands	1
Tree Canopy	1
Shrubland	1
Low Vegetation	1
Barren	1
Structures	NoData
Impervious Surfaces	NoData
Impervious Roads	NoData
Tree Canopy over Structures	NoData
Tree Canopy over Impervious Surfaces	NoData
Tree Canopy over Impervious Roads	NoData

The pervious/permeable raster data are then converted to a polygon layer and ArcMap's Selection > Select By Location is used to find all flowpath segments that intersect it. The selected flowpaths are then exported as a new layer—leaving behind all those flowpath segments that overly "impervious" features (Figure 22).

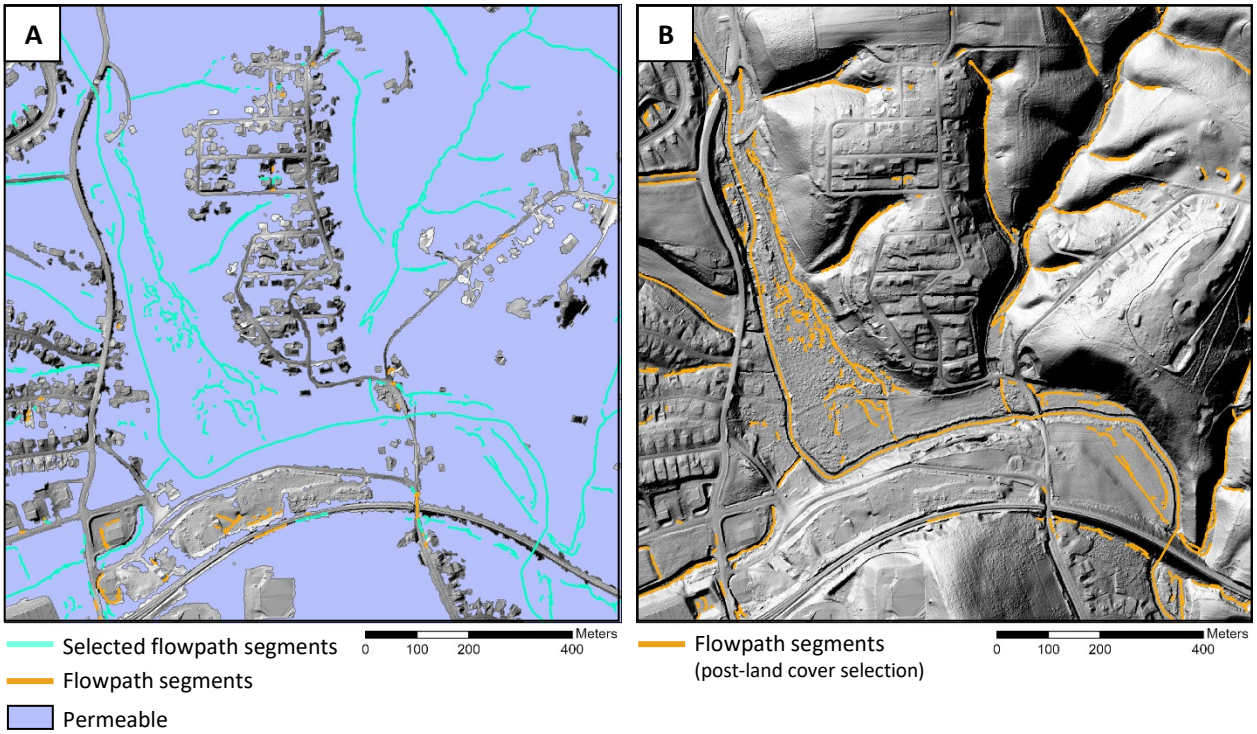


Figure 22. Example of how land cover classes can be used to additionally filter out flowpath segments that are not related to hydrography features. **A.** Land cover classes that are considered permeable are extracted and vectorized. Select By Location is then used to highlight all overlapping flowpath segments. In this example, the land cover data are from Chesapeake Conservancy's 2013/2014 dataset, and the permeable classes are water, wetlands, tree canopy, shrubland, low vegetation, and barren. **B.** The flowpath segments that overlapped permeable land cover classes have been exported to a new layer.

8. USE A FLOW-ACCUMULATION MODEL TO REMOVE ADDITIONAL UNNECESSARY FEATURES

Using a flow accumulation model comprised of Arc Hydro tools will not produce geometries that are accurate enough to meet PAHD inclusion standards, nor is this the goal in utilizing Arc Hydro tools. At this point in the process, Arc Hydro tools represent a quick and easy way of further winnowing down the line segments produced from geomorphon analysis—much like the use of land cover data in the previous step. If, at this point, an HU12's flowpaths do not have abundant extraneous segments (i.e., noisy data), then this step should be considered unnecessary and likely to impede flowpath connectivity. Generally, this step is found to be more helpful in rural areas and less helpful in urbanized regions. There are exceptions to this; for example, directed flow in farmlands is not well represented by flow accumulation outputs. Creator judgment should be used to determine whether or not to proceed with this step.

Notes on Arc Hydro Tools

The Arc Hydro Toolbox is not included with a standards ArcGIS license, but it can be downloaded for free from Esri. There are currently versions of the Arc Hydro Toolbox avail-

able for ArcMap and ArcPro, but as Esri continues to promote ArcPro usage, it is unlikely that Esri will continue to maintain the ArcMap versions of Arc Hydro.

In ArcMap, it is highly recommended to access and use the Arc Hydro tools from ArcToolbox rather than from the Arc Hydro toolbar. Not all tools and commands are accessible through the Arc Hydro toolbar. The same tools and commands accessed through the toolbar—as opposed to ArcToolbox—provide far less control over what data types can be entered, where the resulting data ends up, how it is named, whether the tool functions as intended, and whether the tool functions at all. For these reasons, the Arc Hydro functions described in this methodology are taken from ArcToolbox and linked appropriately in ArcMap's ModelBuilder (Figure 23).

Also, despite Arc Hydro's use of the term "stream," neither the output of the flow accumulation model nor the final dataset produced from the PAHD methodology are actually considered streams. This distinction is intended to respect the existing standards of regulatory factions, many of whom have extensive documentation and practices that rely on specific definitions of "stream."

The Arc Hydro Workflow

The Arc Hydro tools collected in ModelBuilder (in order of operation) and their input/output values are as follows:

1. **DEM Reconditioning** (Arc Hydro Tools > Terrain Preprocessing > DEM Manipulation > DEM Reconditioning)

   - Input Raw DEM (the original unfiltered DEM)
   - Input AGREE Stream (the most recent flowpath geometries)
   - Number of Cells for Stream Buffer: 5 (default value)
   - Smooth Drop in Z Units: 10 (default value)
   - Sharp Drop in Z Units: 10
   - Output AGREE DEM – Choose a location for and name the output raster

   Purpose: This tool "burns" the current flowpath geometries into the original DEM. Wherever flowpath geometries exist, the DEM is edited so that the elevations are lowered by a uniform vertical distance.

2. **Fill Sinks** (Arc Hydro Tools > Terrain Preprocessing > DEM Manipulation > Fill Sinks)

   - Input DEM (output from previous step)
   - Output HydroDEM – Choose a location for and name the output raster

   Purpose: This tool analyzes the DEM for cells that are surrounded by higher elevation values—these "sinks" prevent water from leaving the central cell. Fill Sinks modifies the central cell's elevation value to allow subsequent tools to calculate continued flow, rather than showing the water as being "trapped" by the sink.

3. **Flow Direction** (Arc Hydro Tools > Terrain Preprocessing > Flow Direction)

   - Input HydroDEM (output from previous step)
   - Output FlowDirGrid – Choose a location for and name the output raster

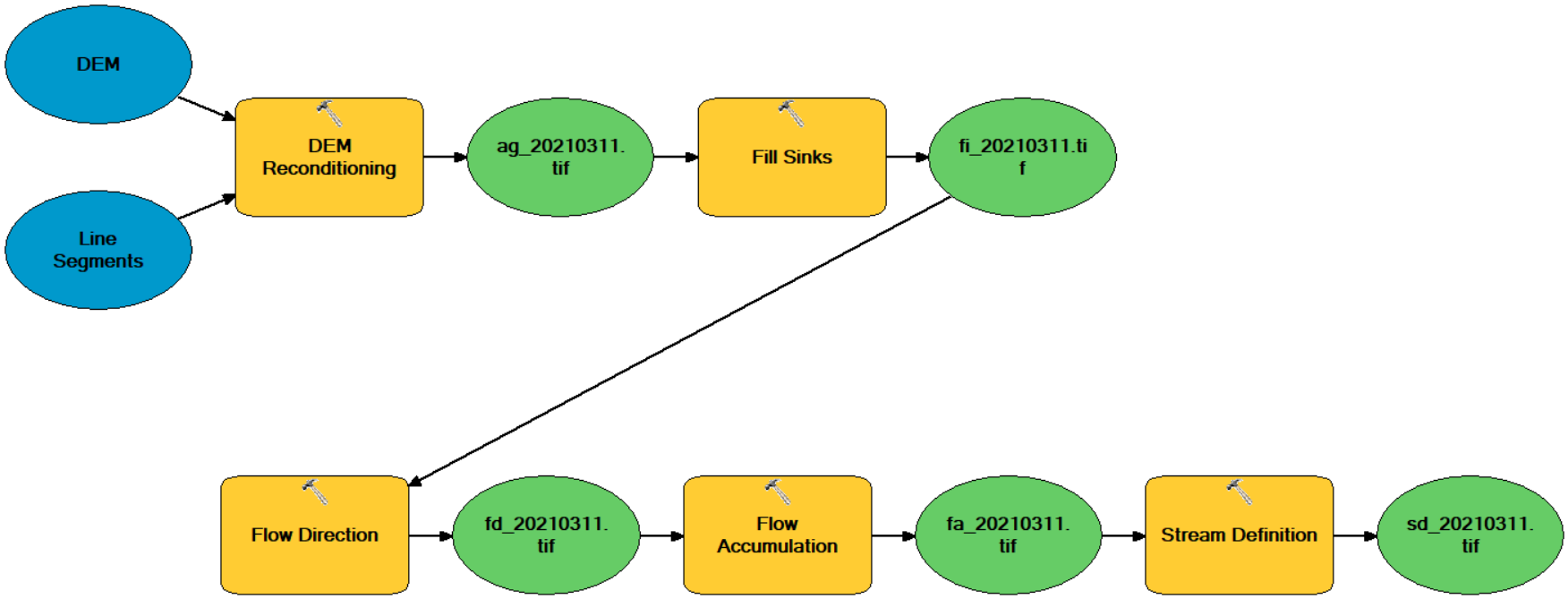


Figure 23. Arc Hydro tools laid out in ModelBuilder. This flow accumulation model can be used to winnow out additional flowpath segments, much like the land cover classes in the previous figure. The success of this step varies significantly with regional setting; sometimes using Arc Hydro-defined stream features removes too many of the geomorphon-generated flowpath segments and resulting connectivity. Generally, the flow accumulation model is more helpful in areas with less anthropogenic influence. This step is considered detrimental in urban settings and farmlands, both of which have had considerable human impact on flow features.

Purpose: This tool calculates flow direction based on each raster cell's elevation value. The Flow Direction tool treats the direction of steepest descent from any given cell as that cell's flow direction. The Arc Hydro Flow Direction tool is only capable of computing the D8 method (Figure 1). Because the output of this model is an ancillary product and does not need to be of the same quality as the final flowpath geometries, Arc Hydro's use of D8 over D-infinity methodology is not concerning.

4. **Flow Accumulation** (Arc Hydro Tools > Terrain Preprocessing > Flow Accumulation)

   - Input FlowDirGrid (output from previous step)
   - Output FlowAccGrid – Choose a location for and name the output raster

   Purpose: The Flow Accumulation tool calculates the total number of cells that contribute flow to each cell in a raster based on the flow direction output.

5. **Stream Definition** (Arc Hydro Tools > Terrain Preprocessing > Stream Definition)

   - Input FlowAccGrid (output from previous step)
   - Number of cells to define stream (specified by user)
   - Output StreamGrid – Choose a location for and name the output raster

   Purpose: This tool generates a stream grid based on the flow accumulation output and a user-set "number of cells to define stream." The number of cells to define a stream, also referred to as a threshold value or an accumulation threshold, defines how many cells must contribute flow to a single cell for that individual cell to be considered part of a "stream."

Use NHD to Improve Flow Output

Like the flow accumulation output, the NHD is not accurate enough to meet PAHD inclusion standards. Despite this, the NHD flowline geometries can be used to improve upon the connectivity and some of the main geometries of the flow accumulation output, both of which are impaired by the presence of false dam-type features in the DEM (e.g., bridges and culverts). That is why, after the flow accumulation model finishes running, it is necessary to merge the resulting polyline with the NHDFlowline layer (Figure 24A).

Selection > Select By Location is then performed on the merged flow accumulation output and NHD layer where it intersects the geomorphon-generated flowpaths. This data can then be exported to a new layer (Figure 24B).

9. PERFORM SIMPLE TOPOLOGY EDITS

Next, the BGS creates a new topology with a cluster tolerance of 1 m. Since the cell size of the geomorphon output is 0.5 m, this effectively highlights all small-scale "loops" and "dangles" without having to run the No Self Intersect and No Dangles rules (Figure 25). At this point, it is not actually necessary to add any topology rules beyond defining the cluster tolerance. When prompted, the topology should be validated. Finally, the Topology Error Inspector can be used to mass-delete all errors resulting from a defined cluster tolerance of 1 m. This removes all small-scale errors.

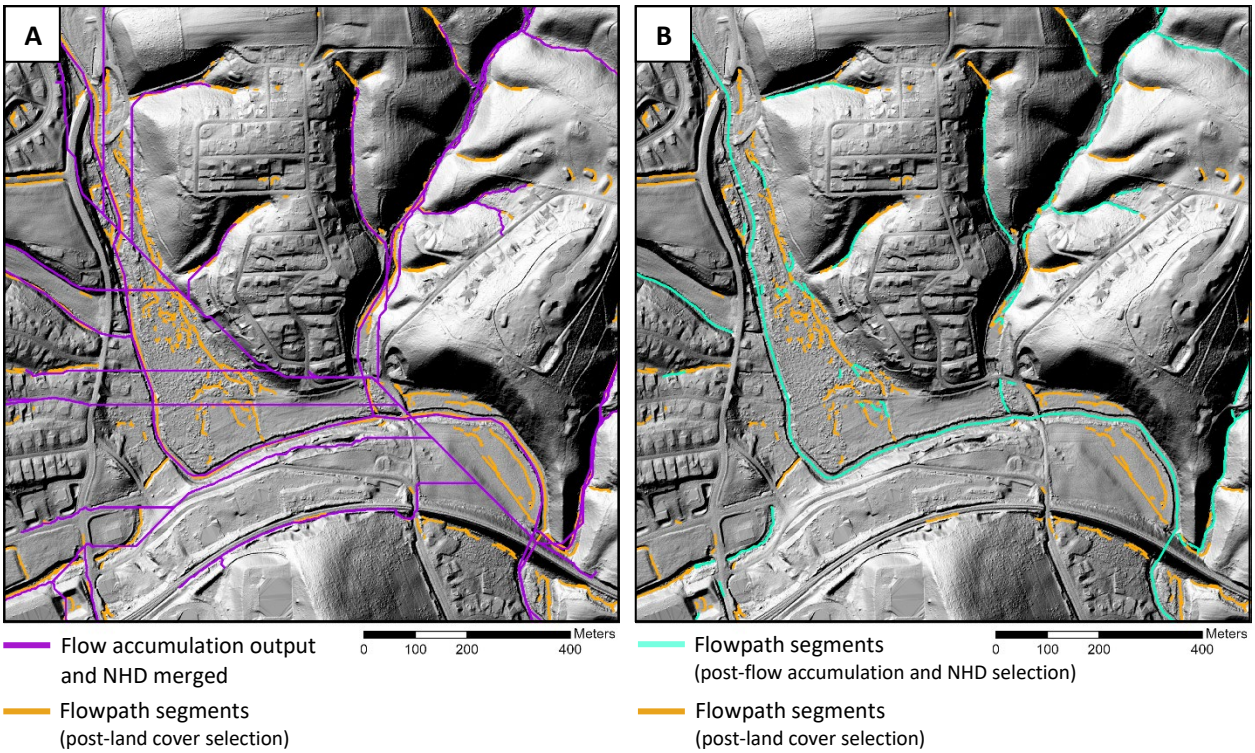


Figure 24. Example of how the output from a flow accumulation model using Arc Hydro tools can be used in conjunction with existing NHD flowlines to additionally filter out flowpath segments that are not related to hydrography features. **A.** The output from the flow accumulation model merged with the NHD flowlines overlying geomorphon-generated flowpath segments. Using a flow accumulation model alone is not enough to produce geospatially accurate flowpaths. The output in this figure is shown to cut large swaths across existing floodplains. The inaccuracies here are partially due to the fact that the Arc Hydro tools have been applied to a DEM that has not been hydro-conditioned, which is one of the goals of this workflow. **B.** The flowpath segments that overlap the flow accumulation model output and NHD flowlines exported to a new layer overlying the original geomorphon-generated flowpath segments.

10. MANUALLY EDIT FLOWPATHS

Before performing a least-cost, also known as a cost-distance, analysis with the unfinished flowpath segments, some manual edits are necessary (Figure 26). These include the following:

- Deleting erroneous lines and segments
- Extending lines over or across culvert and bridge features
- Delineating lines where flowpaths are missing along large or complicated sections due to poor data return; this includes missing sections longer than about 100 m and sections with either multiple junction points or significant sinuosity

Perhaps this is the most time-consuming step in the entire PAHD workflow. Theoretically, with a true least-cost analysis, the only manual work necessary at this juncture *should* be the extension of flowpath segments across culvert and bridge features. Practically, how-

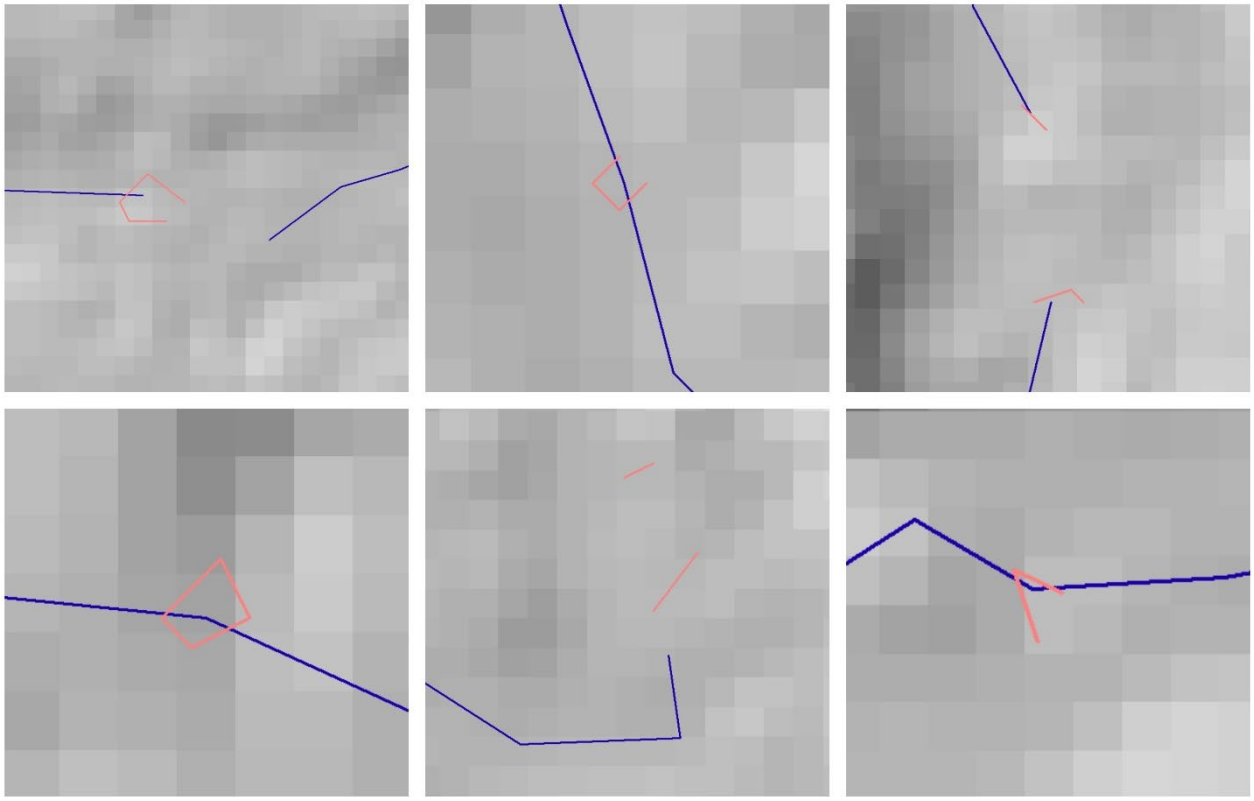


Figure 25. Small "loops" and "dangles" (pink features) are errors created when a geomorphon output was converted from a raster to a vector. The vectorized geomorphon output (navy lines) are shown overlying the gray pixels of a 1-m-resolution DEM. The small-scale errors were generated from a raster with a cell size of 0.5 m, which means that their vertices are less than 1 m apart from one another. The BGS created a topology with a cluster tolerance of 1 m to select these errors, which can be easily mass deleted.

ever, Arc Pro's ORC tool (or, alternatively, ArcMap's Cost Connectivity tool) does not perform as well with lines as it appears to perform with polygons.

11. CONNECT FLOWPATH SEGMENTS

To connect and join the flowpath segments, the BGS runs a least-cost analysis. This computation can be done in either Arc Pro (ORC tool) or in ArcMap (Cost Connectivity tool). While the tools in both programs are capable of connecting flowpath segments based on a "least-cost" path calculated from the DEM, the ORC tool is preferable in this case. The Cost Connectivity function is a legacy tool that lacks many of the improvements associated with the ORC tool; for instance, the ORC tool has D-infinity capabilities (Figure 1). In practice, the ORC tool also has a much faster processing time than its legacy equivalent and is far less prone to crashing.

For the PAHD methodology, the ORC tool requires two inputs: (1) Input raster or feature region data (the flowpath segments), and (2) Input cost raster (the original unedited DEM).

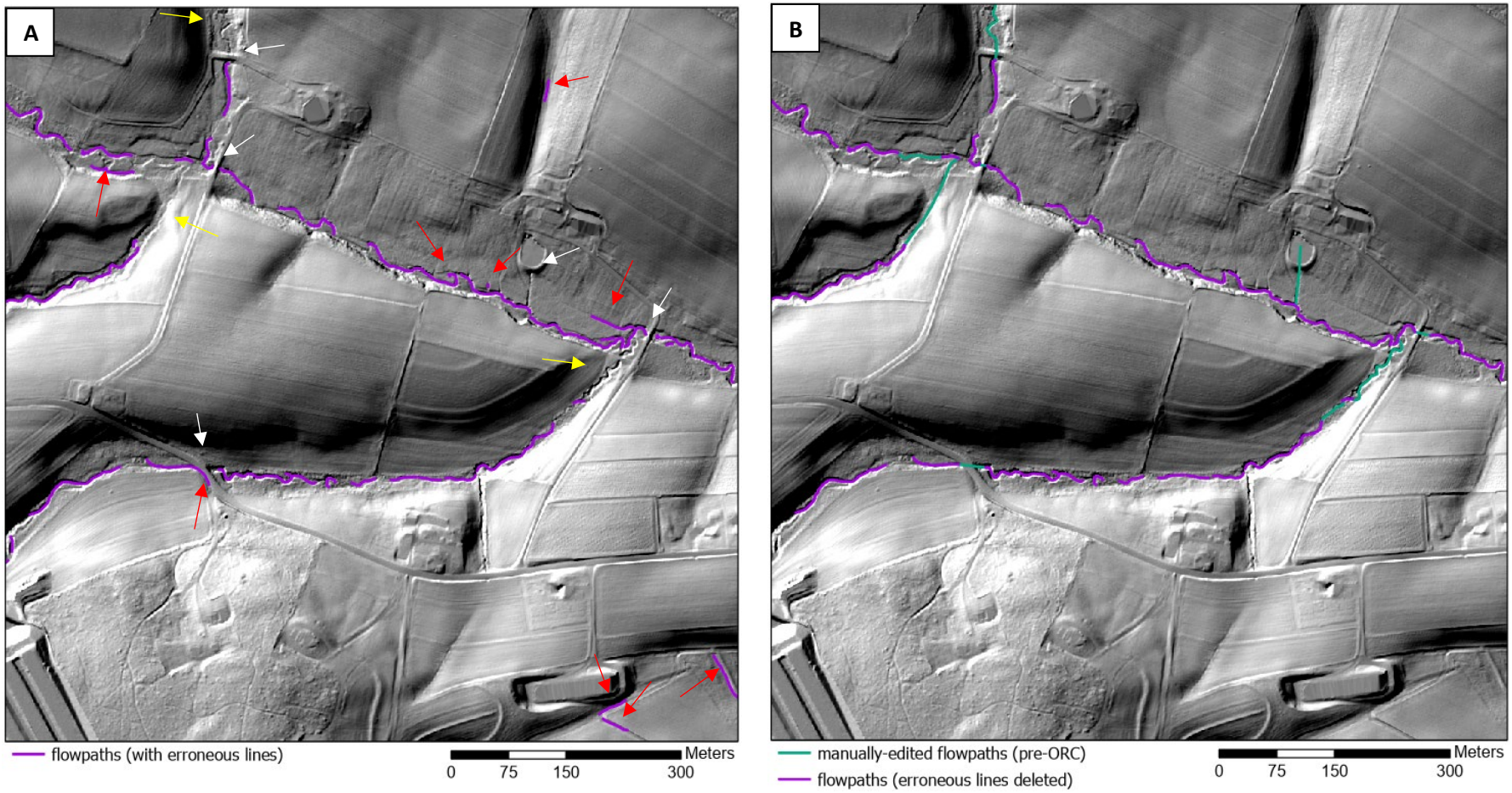


Figure 26. Examples of manual edits that need to be performed prior to least-cost analysis. **A.** Before edits are made. The red arrows show erroneous lines that need to be deleted, the white arrows show where flowpath segments need to be extended across culvert-type features, and the yellow arrows show where flowpaths need to be digitized along large or complicated missing sections. **B.** After the edits have been performed. The teal lines are manually digitized segments that either breach culvert-type features or fill in missing flowpath segments that are either too long or too complicated for the ORC tool to fill. The purple lines are geomorphon-generated flowpaths that have had all erroneous lines deleted.

The Distance Method type is another input, but it does not make much of a difference to the output. It is considered best practice to select the Geodesic method when working with large areas, so that is what the BGS does. The Geodesic method requires more processing time, but it will never make the output less accurate.

Because the ORC tool was designed to work with polygons, as opposed to polylines, the output of this tool necessitates additional edits. Figure 27 shows a simplification of the ORC tool mechanics: the paths between polygons that have the "lowest cost" (i.e., the lowest elevation values) are demarcated. These least-cost paths begin at the center of a polygon and end at the center of a neighboring polygon. When this technique is applied to polylines, the connecting least-cost paths end halfway through the outermost line segments (Figures 27 and 28).

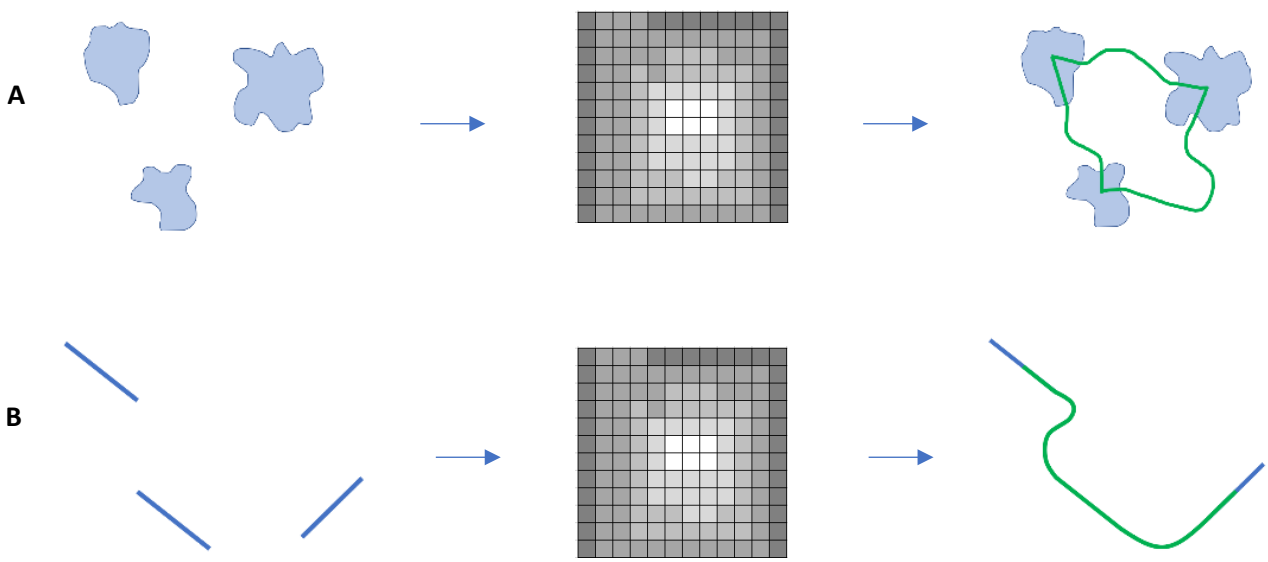


Figure 27. Simplified illustration of ORC tool mechanics. **A.** The tool attempts to find the least-cost paths (green line) between regions of interest (blue polygons). The "costliness" of these paths is determined by elevation: lower elevation values are lower cost. The tool starts a path at the center of each polygon and follows the lowest elevation values until it reaches the next polygon. **B.** The same ORC mechanics are illustrated here, but with lines instead of polygons. As a result of starting each path in the center of a "region" (blue lines), the final least-cost path (green line) will not fully cover all starting regions.

The reason the ORC tool does not appear to work as well with polylines as it (reportedly) does with polygons is unclear. Spatial Analyst production engineers at Esri have suggested that it may be due to a developer issue and that a patch may be created for ArcGIS version 2.6 or 2.7. The author shared the following ORC-tool-output problems with the Spatial Analyst production engineers during and after the 2020 Esri User Conference:

- Incorrect junction relationships that lead to abnormally long and incorrect least-cost path segments (Figure 29A)

- Kinks in the output lines that did not previously exist in the input line segments. These kinks were not generated as least-cost paths; they did not occur at line-segment gaps, but rather they were generated where the input flowpath already existed. These kinks effectively made the existing geometries of the flowpath segments less accurate (Figure 29B).

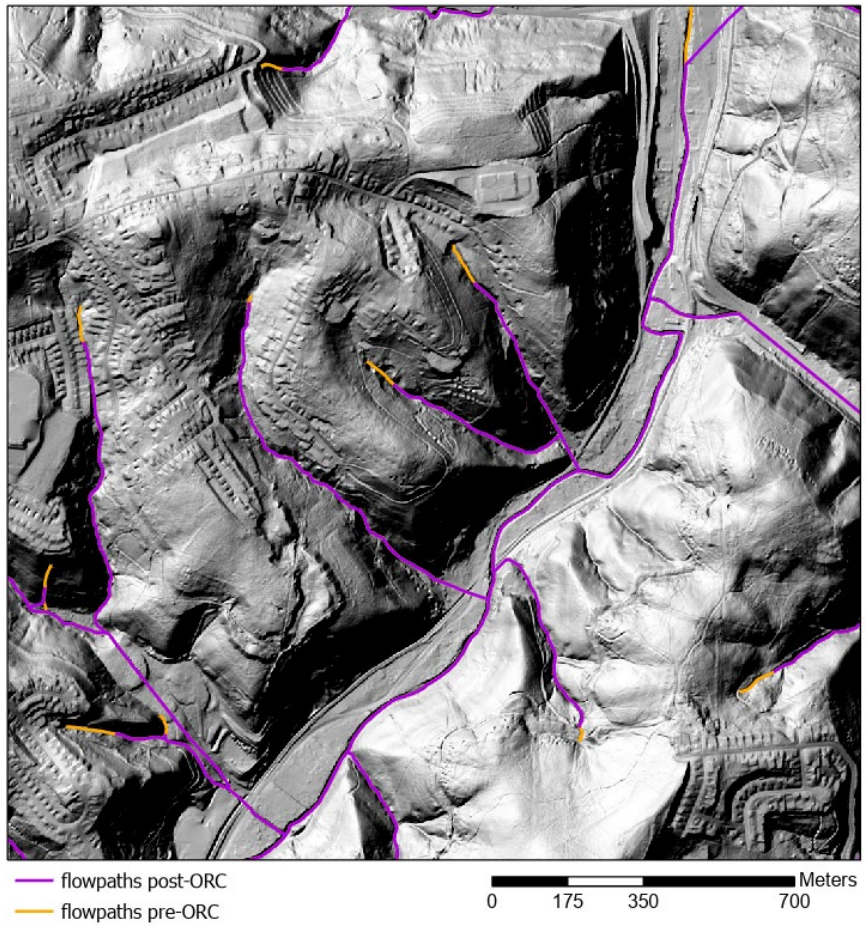


Figure 28. Example of ORC tool outputs overlying the original regions of interest. Due to the nature of the ORC tool, the outputs do not entirely replicate the extent of the original flowpaths.

To begin the process of producing a connected and cohesive set of flowpaths, the BGS uses the Erase tool to subtract the pre-ORC flowpaths from the post-ORC flowpaths. This leaves all the ORC-generated least-cost flowpaths that do not overly pre-ORC flowpaths (Figure 30), that is, the very short kink errors, the long, straight segments connecting incorrect junctions, and the flowpaths that fill in gaps from the pre-ORC flowpath vectors.

The kinks are then removed by opening the attribute table for the modified least-cost paths, sorting the Shape_Length field in ascending order, and deleting all lines less than about 2.9 m in length. When performing this operation, it is recommended to check a few of the lines close to the 2.9-m length to verify whether or not they are artifacts. The checkpoint length value will vary regionally; in more urban settings, the divide between "good" ORC outputs and "bad" ORC outputs was less clear.

Next, the BGS manually deletes the least-cost flowpath segment errors resulting from incorrect junction relationships. These errors tend to be obvious, so this bit of manual work is easy and goes very quickly.

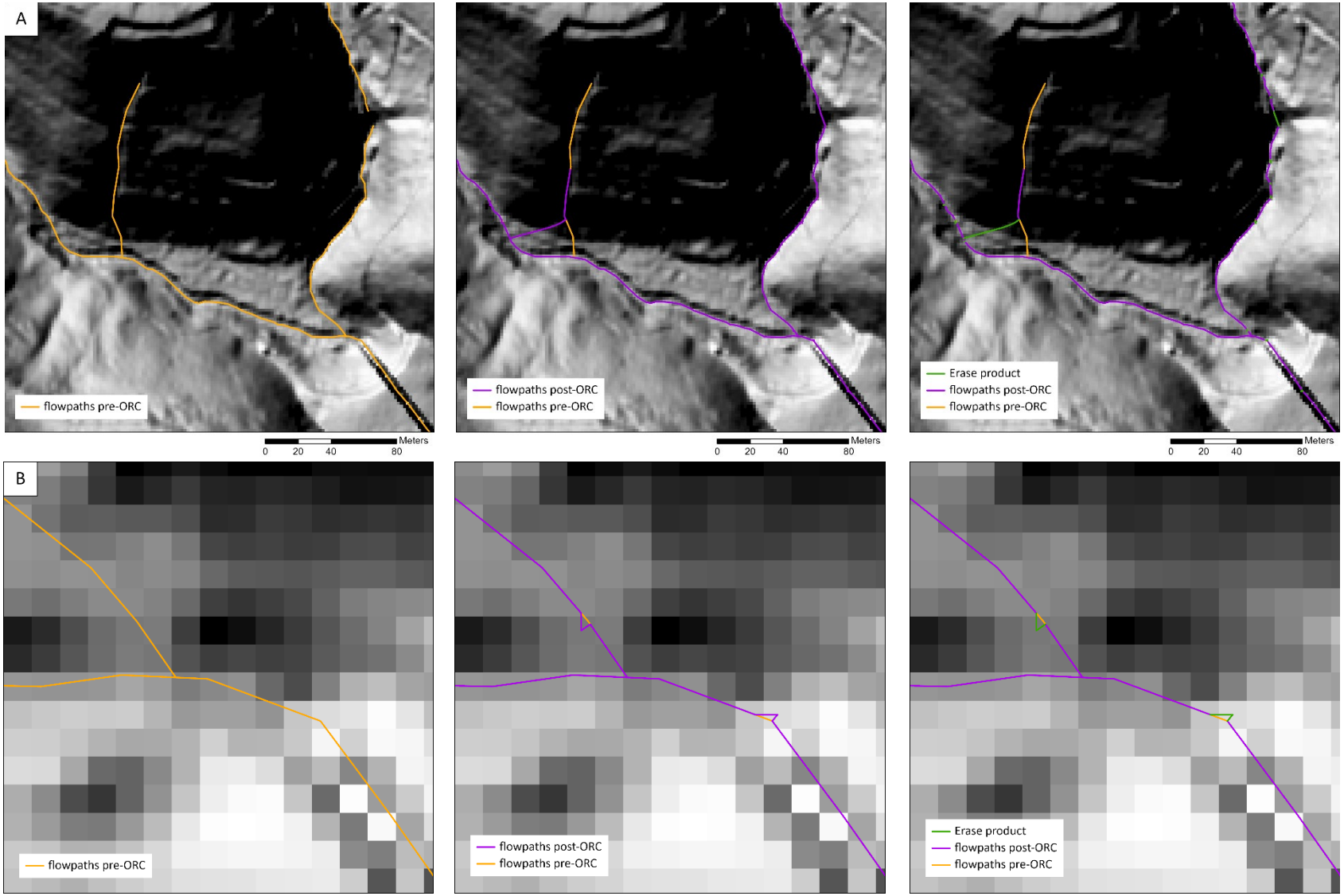


Figure 29. Errors resulting from ORC tool use. **A.** Incorrect junction relationships lead to lines incorrectly cutting across higher elevations from one flowpath to another, which result in abnormally long straight segments. **B.** Small kinks in the ORC tool output lines that are not present in the input flowpaths.

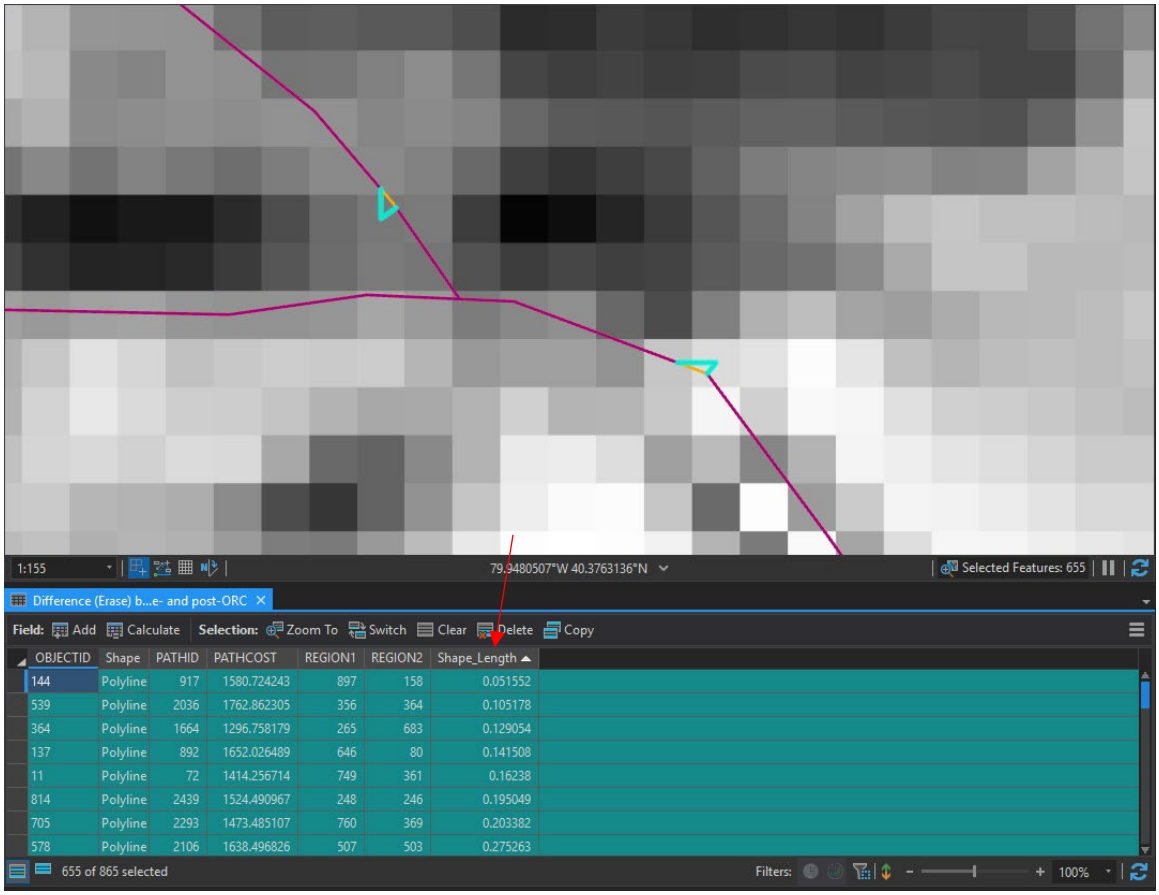


Figure 30. Example of how using the Shape_Length field to highlight all lines less than about 2.9 m in length will make it possible to mass delete all kinked features arising from use of the ORC tool.

After the modified least-cost flowpaths have been appropriately edited, they are merged with the pre-ORC flowpaths. This solves the problem of the ORC-generated flowpaths stopping halfway through the outermost pre-ORC line segments.

12. MAKE FINAL TOPOLOGY EDITS

The methodology described in the previous section does not extend or combine lines; rather, it creates a series of new lines. It also creates a lot of unnecessary breaks in the output line segments (Figure 31). To remove these breaks, the BGS creates a geometric network using the flowpaths as simple edges. This is done by creating a new Feature Dataset and populating it with the line segment polyline. Right clicking the Feature Data and selecting New > Geometric Network will open up the New Geometric Network wizard. The wizard requires a network name and asks whether features should be snapped within a specified tolerance (choose yes and use the default value). The line segments polyline should be selected to build the network and existing enabled values should be preserved. The result of this is a network that will be treated as "temporary." The network is comprised of edges (the line segment polyline) and newly generated junctions.

The flowpath edges and created junctions get exported to a new feature dataset for topology edits. (If the network is deleted before the newly created junctions are exported, it will also delete all the junctions. Exporting to a new feature dataset saves the junctions as a feature class that is independent from the original network.) The edges are then dissolved into one continuous line.

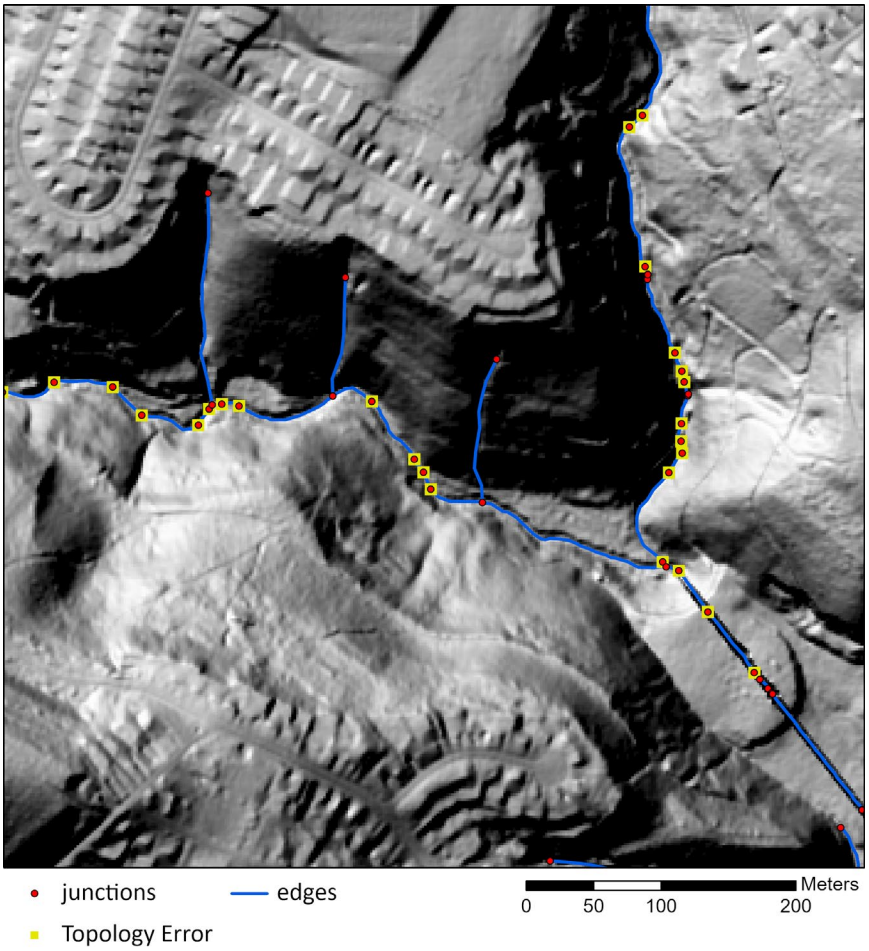


Figure 31. Example of how topology edits can be used to correct network relationships. Here the topology rule that all junctions "must be covered by the endpoint of" dissolved edges has highlighted all the unnecessary breaks in the original line segments.

Next, the BGS creates a topology using the dissolved edge(s) and junctions with the following topology rule: junctions "must be covered by endpoint of" dissolved edge(s) (Figure 32). This topology highlights all the unnecessary breaks in the original line segments. All unnecessary junctions can now be deleted by selecting the Topology tool bar > Error Inspector, leaving behind the real network junctions.

The network junction points can then be used in conjunction with the dissolved lines as input for the Split Line At Point tool, which correctly splits lines at points of divergence or convergence and nowhere else. It also serves to take care of other topology errors, including overlaps, self-overlaps, and pseudo nodes.

Lastly, to make sure that all lines were merged correctly and that no mistakes have slipped through, features in the flowpaths attribute table should be sorted by length, and the shortest lines should be checked (Right click > "zoom to") visually.

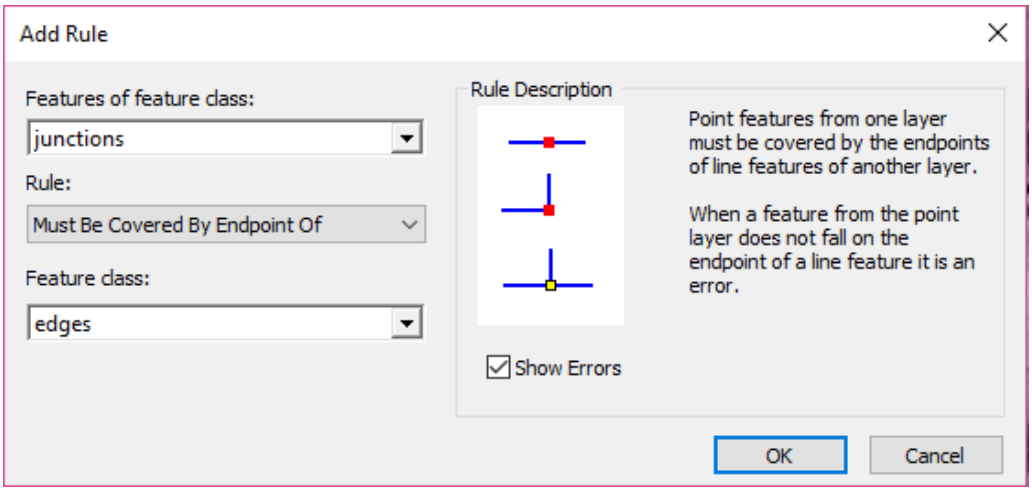


Figure 32. ArcMap's topology Add Rule window shows the rule used by the BGS for all junctions "must be covered by the endpoint of" dissolved edges.

INSTRUCTIONS FOR ATTRIBUTION AND NETWORKING

The order of operations in this section is paramount to successful attribution and networking. Creating a network before assigning stream order causes a myriad of problems with the Arc Hydro toolset. Transferring attributes (e.g., from the NHD) to equivalent features (in the PAHD) before establishing stream order means that this data is being passed from one dataset to another with an incomplete understanding of flowpath relations. Furthermore, it is not possible to assign stream order prior to establishing flow direction.

1. VERIFY FLOW DIRECTION

Setting and storing flow directions are Network Tools within the Arc Hydro toolbar; as such, it is necessary to create another "temporary" geometric network using the previously edited edges and junctions prior to using these tools. Flow direction in a geometric network can be established by either (1) assigning all junction points the role of "source" or "sink," or (2) the digitized direction of the network edges.

Sources and sinks determine flow direction. A junction point that is a "source" will push flow away from itself along the edges of a network; a junction point that is a "sink" will pull flow toward itself along the edges of a network. Esri (2016) gives the following examples of sources and sinks:

- Pump stations can be modeled as sources because they push (pump) water away from themselves through pipes
- River mouths can be modeled as sinks because water moves toward river mouths due to gravity

Flow direction can be determined with either sources or sinks, so it is generally recommended to specify only one or the other in a single network—this avoids (but does not entirely prevent) creating edges with indeterminate flow (Figure 33).

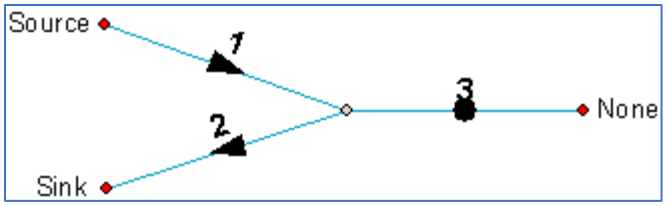


Figure 33. Example of how indeterminate flow can occur in a geometric network that uses both sink and source junction assignments to determine flow direction. Taken from Esri (2016).

There are a few problems with using source or sink assignments to determine flow direction; for instance, leaving the responsibility of determining source versus sink up to the creator allows for a greater possibility of human error. More significantly, this is an indirect method of analyzing elevation data, assuming the elevation data are analyzed at all. The main directive of the PAHD project is to produce elevation-derived hydrography data, so it does not make sense to leave something as influential as flow direction (and therefore stream order and network relationships) up to subjective creator interpretation. Additionally, unlike the source-sink method of determining flow direction (Figure 33), assigning flow direction based on the digitized direction of network edges should never result in indeterminate flow.

The PAHD method of determining flow direction is therefore based on the digitization direction of edge features. Before setting flow direction to follow the direction of digitization ("WithDigitize"), it is necessary to assess whether all line segments are digitized in the correct direction. When generating geometries with Arc Hydro, lines are automatically digitized in the correct direction. When generating geometries from geomorphons, it is necessary to go back and check and correct the direction of delineation.

In an open editing session, the digitized direction of line segments can be seen by selecting the Edit Tool (black arrow) and double-clicking individual line segments. The "endpoint" of the feature will have a red vertex (Figure 34).

To assure that the lines have been digitized in the correct direction, the elevation (Z-value) of the starting vertex has to be higher than the elevation of the ending vertex. To accomplish this, the BGS adds three new float fields to the edge attribute table: "Z_start," "Z_end," and "Z_difference." The Interpolate Shape tool can then be used to interpolate elevation values for the line segments based on the unaltered DEM.

Once elevations have been defined for the edge features, individual elevation values are assigned to the "Z_start, "Z_end," and "Z_difference" fields. The start and end values are calculated by opening the edges attribute table and selecting Calculate Geometry > Property: Z Coordinate of Line Start (or, alternatively, Z Coordinate of Line End). The "Z_difference" field is calculated using the Field Calculator to find the difference between the start and end elevation values of the edge features.

If a line's starting vertex has a lower elevation than its ending vertex, the Z_difference will have a negative value—this represents an edge feature that was digitized in the wrong direction (Figure 35A). Conversely, if a line's starting vertex has a higher elevation than its ending vertex, the Z_difference will have a positive value—representing an edge feature that was digitized in the correct flow direction (Figure 35C).

To fix all the incorrectly digitized flowpaths, (i.e., the flowpaths that were digitized in a direction opposite to expected flow), the Z_difference values should be sorted in ascending order. This bumps all of the negative values of the incorrectly digitized lines to the top of the

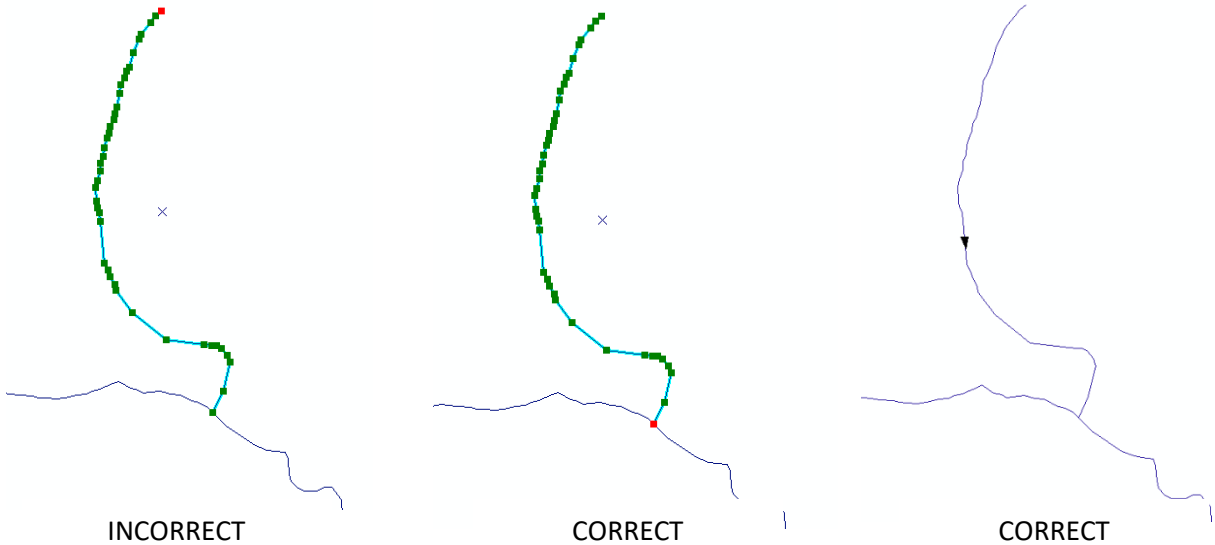


Figure 34. Because flow direction is matched to the direction in which each line is digitized ("WithDigitize"), it is necessary to check that all flowpaths are digitized in the correct downstream direction. The "endpoint" of a line feature in an editing session will have a red vertex. The left-most image shown here has the red vertex at the initiation point/headwater of this tributary, which is incorrect. The middle image shows the result of applying the Flip tool. The rightmost image shows the flipped line with the correct flow direction assigned.

Z_difference column. Select all the features with a negative Z_difference and export them to a new line feature class. This feature class will serve as the input for the Flip Lines tool (Figure 35B).

After flipping the incorrectly digitized lines (Figure 35C), the Erase tool can be used to remove their incorrect counterparts from the original set of edge features. The newly flipped lines (with positive Z_difference values) can then be merged back with the original feature class.

The flowpaths, which are now all correctly digitized from a higher elevation to lower elevation direction, are entered into a "temporary" geometric network. Arc Hydro Tools (toolbar) > Network Tools > Store Flow Direction is selected and applied to the "temporary" geometric network, after which Arc Hydro Tools (toolbar) > Network Tools > Set Flow Direction (select flow direction WITH_DIGITIZED) is selected for the same geometric network. As a final check, the Utility Network Analyst toolbar can be used to show flow direction (Flow > Display Arrows For) for the geometric network.

Whenever a geometric network is generated, the features within the network are checked to make sure their geometries are not invalid within the context of the network. Examples of Network Build Errors include incompatibilities such as features with empty geometries, edge features that form a closed loop, or standalone junctions that do not match to any edges. A Network Build Errors table is created automatically whenever a geometric network is generated, and this table can be used to identify and correct these network errors. The caveat of this table is that it is not automatically updated: any changes made to the features participating in the geometric network will not be reflected in the table unless the user manually

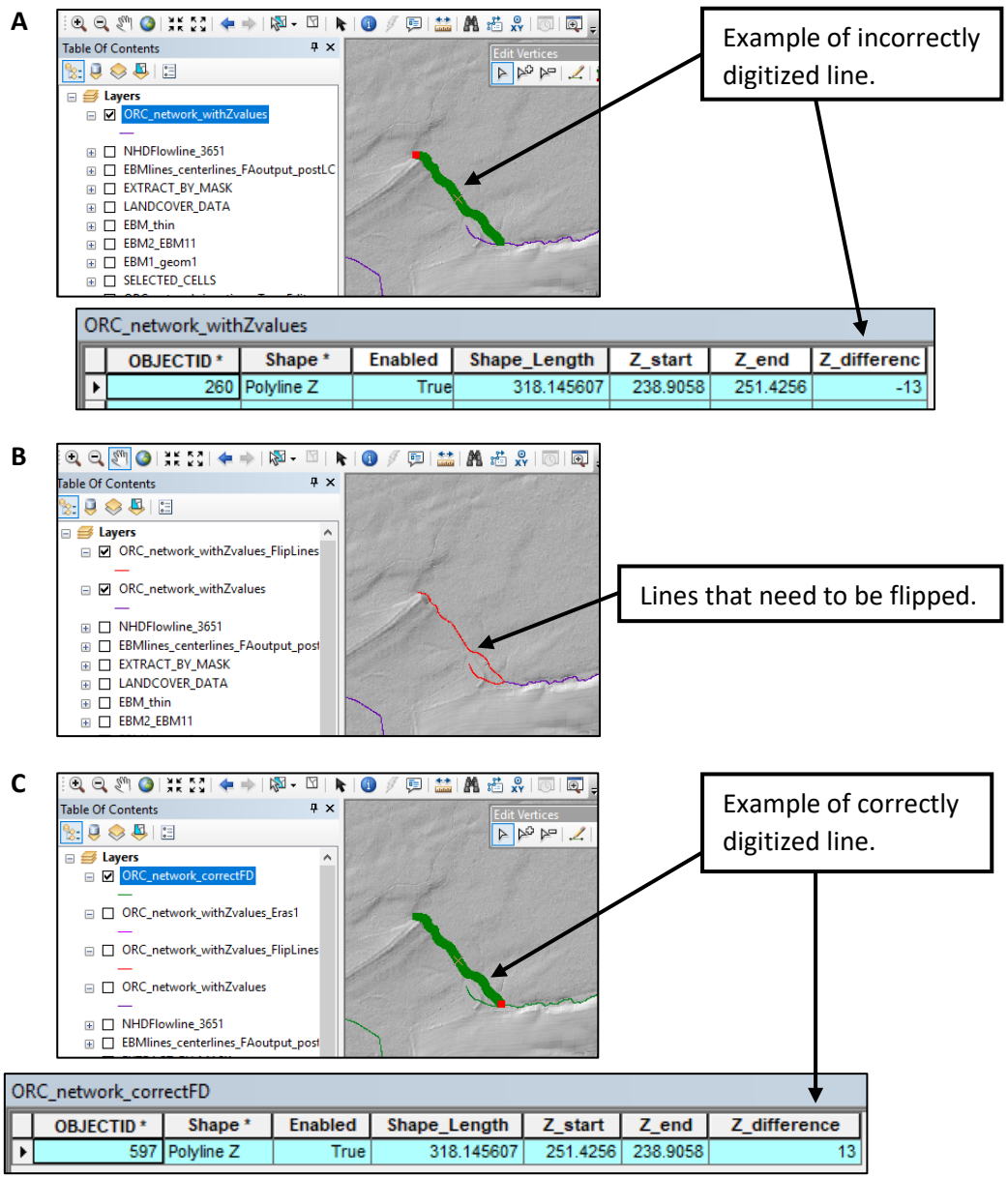


Figure 35. Finding and selecting flowpath lines that have been digitized with incorrect directions can be done quickly using the associated elevation values in the flowpaths feature class. **A.** Example of a line that is digitized incorrectly, that is, the ending vertex (red point) is shown at the flowpath's initiation point. In the flowpaths attribute table, the elevation values of this line also show that the incorrect "starting point" has a lower elevation than the line's "ending point" (i.e., negative difference in Z-values). **B.** All flowpaths in the example that have negative Z_difference values were selected and exported to a new layer (shown by red lines). At this point, the digitization direction of these features can be corrected with the Flip tool. **C.** This new image shows the result of deleting the incorrectly directed lines from the original layer and replacing them with the correct "flipped" lines. Now all flowpaths are digitized in the direction of higher elevation to lower elevation (resulting in positive Z_difference values), and the matching assigned flow directions will correctly reflect the flow of water from a higher to lower elevation.

updates the table or generates a new geometric network. Because creating "temporary" geometric networks is an important aspect of this workflow, regularly checking the Network Build Errors table is a useful way of checking for feature-based network errors.

Once the original elevation surface has been used to correct line flow direction, the geometric network should be deleted to avoid tool errors during stream order computations.

2. CALCULATE STREAM ORDER

Once elevation-based flow directions have been assigned to the flowpaths, it is possible to calculate stream-order values. Again, some of the tools described in this section do not function as reliably when the flowpaths are a part of an existing geometric network; in particular, the Assign River Order tool is especially prone to failure when flowpaths are participating in a geometric network. It is also important to note that (excepting Assign HydroID) all of these tools work much better when accessed from ArcToolbox, *not* the Arc Hydro toolbar. Assign River Order is not even accessible through the Arc Hydro toolbar.

First, HydroID's are assigned to all flowpath segments (Arc Hydro toolbar > Attribute Tools > Assign HydroID). The next tool used in this process is the Generate From/To Node For Lines tool, which helps define line relationships based on flow direction. Following the assignment of From/To Nodes, the Find Next Downstream Line tool is used to further define the connections between line segments. The Find Next Downstream Line tool uses these relationships to retrieve the HydroID of the next downstream line and store it in each feature's NextDownID field.

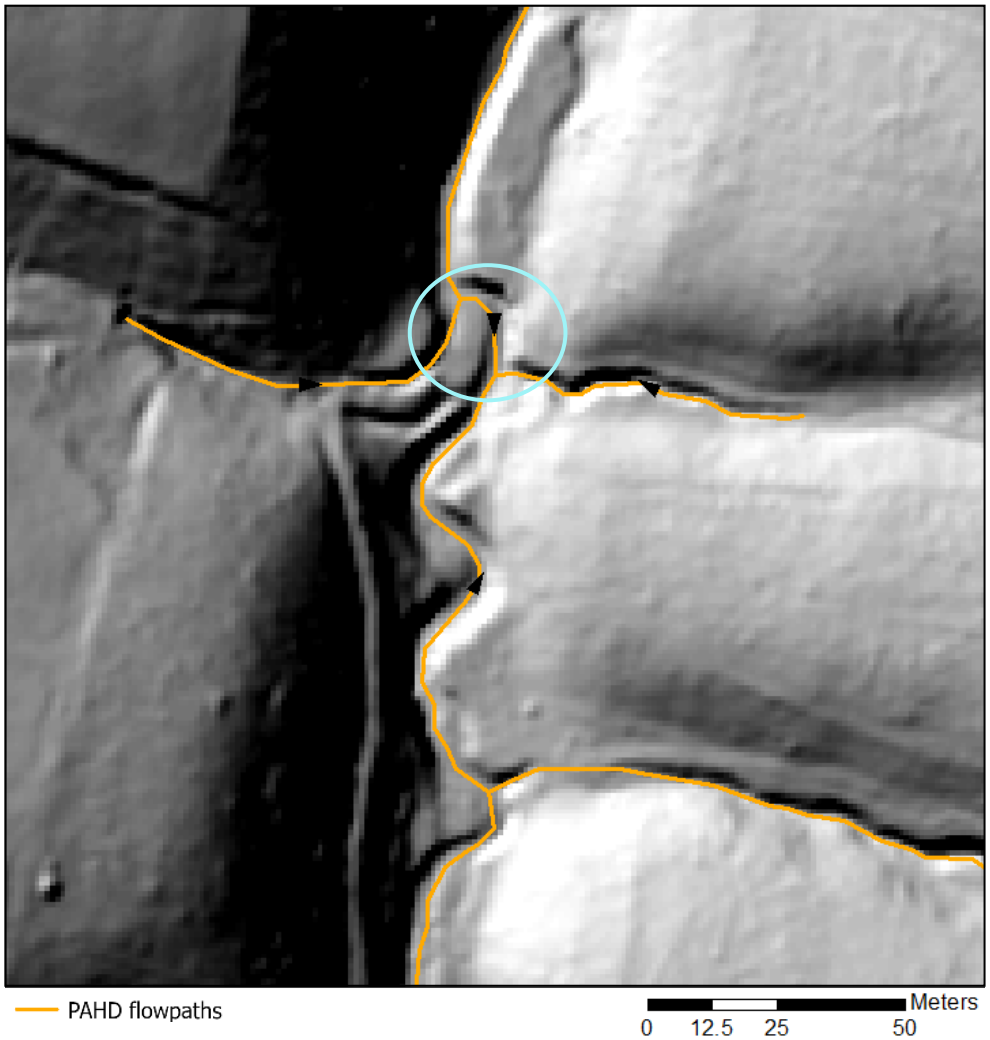
At this point, the results of the Find Next Downstream Line tool (the NextDownID attribute field) can provide a final check on the accuracy of the flow directions associated with the digitization of the flowpath lines. This can be done by opening the flowpaths attribute table and sorting the NextDownID values in descending order. Any flowpath segment that does not have a "next down" line (i.e., any flowpath segment that does not connect to another downstream segment) will have a NextDownID value of -1 (Figure 36).

Line segments with a NextDownID value of -1 are not necessarily digitized in the wrong direction. There are two instances in which flowpath segments should not have a "next down" line: (1) if the flowpath segment is the farthest downstream feature in an HU12 and is not yet connected up with another HU12 (Figure 37), or (2) if a flowpath geometry flows into a sink feature (Figure 38).

Realistically, hydrology features are not naturally isolated from the overall hydrologic system. In some instances, these features are only depicted as being insular because their actual interaction with the overall system is unknown—as is the case with hydrography features connected to the greater network by subsurface flow (Figure 38). In cases where subsurface flow is known, due to supplemental data like MS4, these features can be shown as connected to the overall geometric network and will have an Enabled value of "True."

To perform a final examination of digitized flowpath direction, a temporary geometric network should first be created based on the flowpaths. The flow directions for individual lines can then be displayed (Utility Network Analyst toolbar > Flow > Display Arrows For), and these flow direction arrows can be compared to erroneous NextDownID values (as

shown in Figures 36, 37, and 38). If any flowpath directions are determined to be incorrectly digitized (e.g., Figure 36), this can be corrected in an editing session by double-clicking the segment and right-clicking > Flip. If this is done for any flowpath segments, the Generate From/To Node For Lines and Find Next Downstream Line will have to be reperformed.



	OBJECTID *	Shape *	Enabled	Z_start	Z_end	Z_diff	Shape_Length	HYDROID	NextDownID *
	466	Polyline Z	True	143.9642	143.9368	0.027405	19.214125	466	-1

Figure 36. Franklin County example of how the -1 value of the NextDownID attribute field can be used to identify flowpath segments that have been digitized in the wrong direction. Here the black arrows denote flow direction, and the incorrect flow direction is circled in cyan. Attributes associated with this segment are shown at the bottom of the figure. The Z values show how these errors can slip by prior checks—if the difference between the starting and ending Z values is less than the lidar product's resolution, then these values may not accurately portray upstream versus downstream directions. The segment highlighted here should be flipped.

Finally, Arc Hydro's Assign River Order tool (using the Strahler method) can be utilized to generate and populate the stream order attribute field.

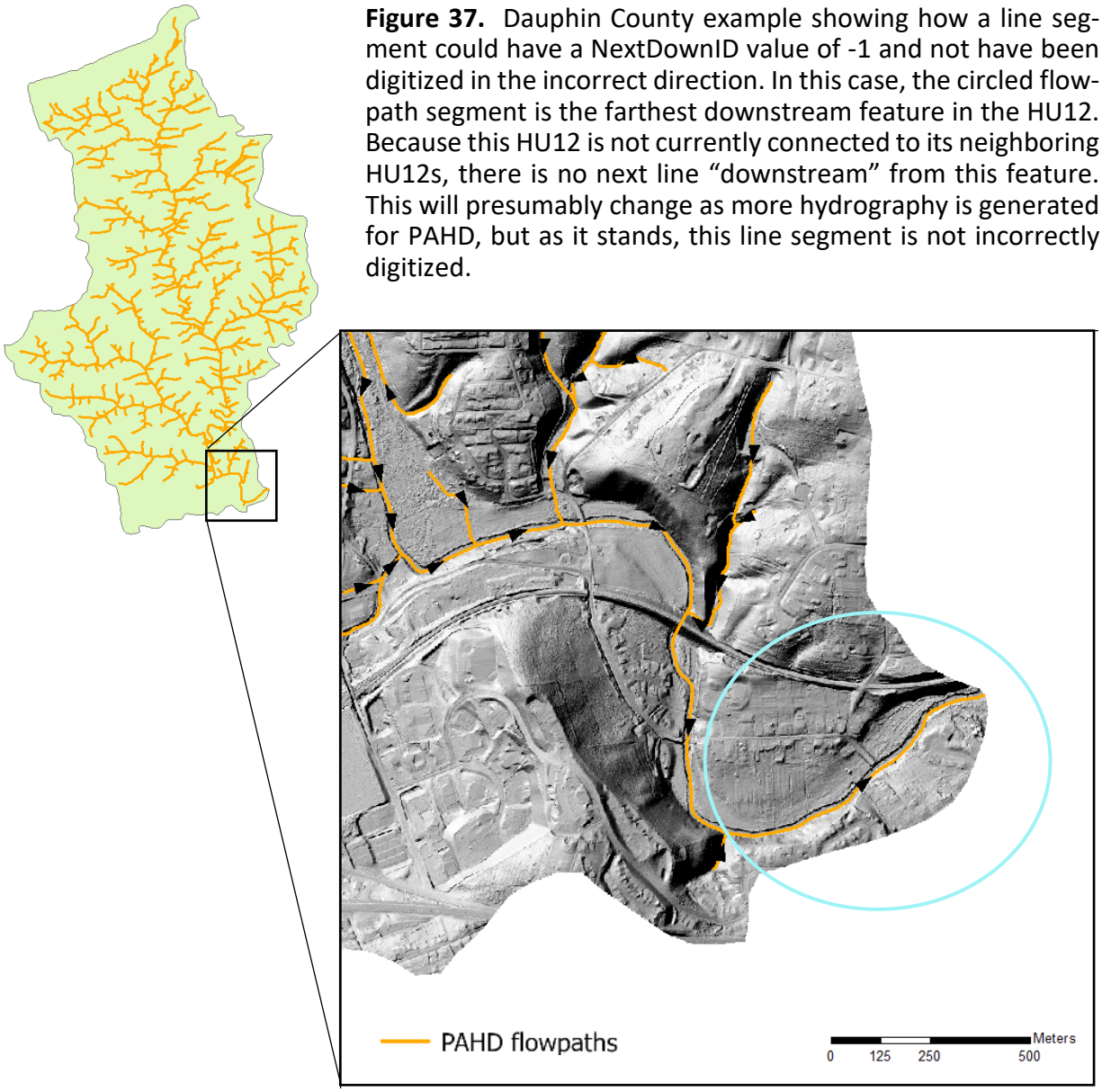


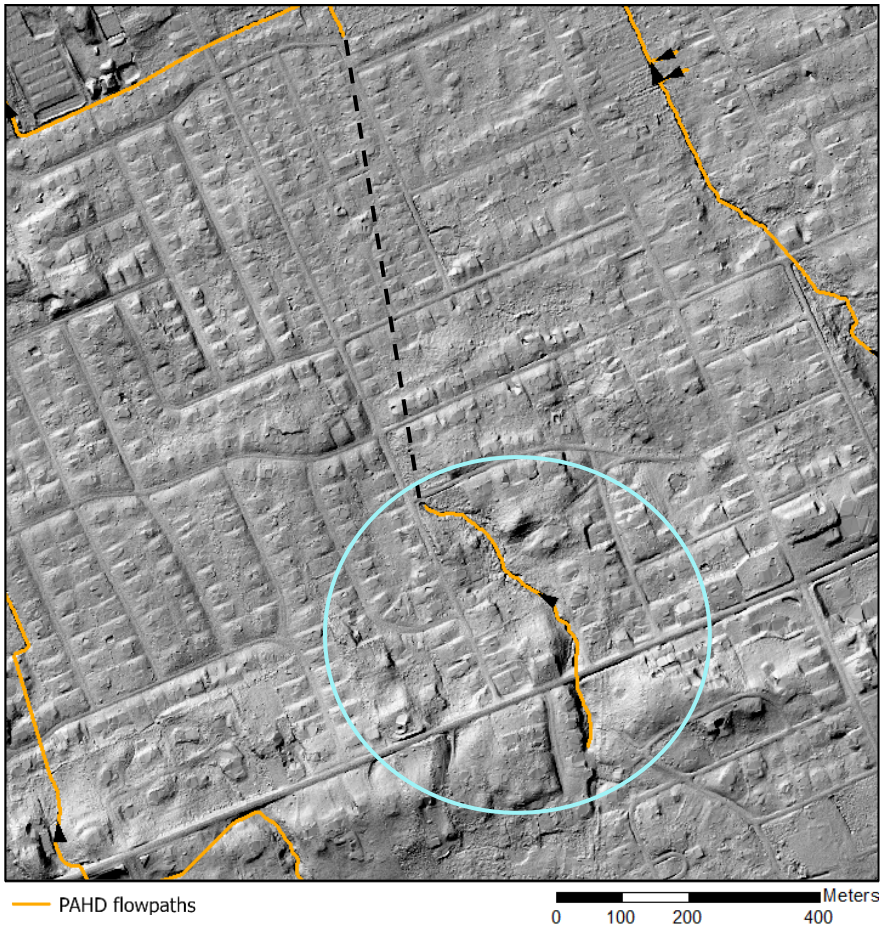
Figure 37. Dauphin County example showing how a line segment could have a NextDownID value of -1 and not have been digitized in the incorrect direction. In this case, the circled flowpath segment is the farthest downstream feature in the HU12. Because this HU12 is not currently connected to its neighboring HU12s, there is no next line "downstream" from this feature. This will presumably change as more hydrography is generated for PAHD, but as it stands, this line segment is not incorrectly digitized.

	OBJECTID *	Shape *	Enabled	HYDROID	NextDownID *
	854	Polyline Z	True	854	-1

3. TRANSFER ATTRIBUTES

A Note About the Transfer of Attribute Data and the NHD

There are many existing hydrography datasets that have competent and comprehensive attribution, but whose geometries, for whatever reason, are not suitable for conflation with PAHD. Perhaps the geometries do not meet PAHD specifications for spatial accuracy, as is the case with the NHD. Another example is the Bureau of Forestry's hydrography dataset, which does not have consistent feature coverage. Whatever the reason a dataset does not qualify for full PAHD conflation, the value of supplemental attribution cannot be overstated.



	OBJECTID *	Shape *	Enabled	HYDROID	NextDownID *
▸	734	Polyline Z	False	734	-1

Figure 38. Erie County example showing how a line segment may have a NextDownID value of -1 and not have been incorrectly digitized. In this case, the line segment circled in cyan is a known flowpath, but its relationship to the overall network is unknown. It likely connects to a downstream geometry via subsurface flow (a hypothetical subsurface flowpath is shown here as a dotted black line), but whether or not this is actually the case is unclear without supplemental data. This feature therefore has an Enabled value of "False," and no additional corrections can be made at this time.

Because of this and in the interest of creating a robust and useful dataset, it is in the best interest of this project to incorporate existing attribute data.

The example cited here to illustrate the transfer of supplemental attributes is from the NHD. Identifying a feature as "equivalent" to an existing NHD feature is not straightforward. The specifications used internally by the USGS to make updates and corrections to the NHD are not published in any publicly accessible document. The only way to learn how the NHD identifies features as "equivalent" is to participate in three training sessions (totaling 12 hours of participation) offered sporadically by the USGS, which focus on the use of USGS-specific toolsets, including the NHD GeoConflation Tool and the NHD Update Tool.

A description of the NHD update and conflation process is provided on the USGS website under the NHD GeoConflation Tool section [boldface added by this author]:

> In this process, the source dataset containing the current geometry and attribution and the target dataset with new geometries and minimal attribution are spatially compared. Where the geometries (e.g. streams, rivers, lakes, etc.) match **within specified tolerances**, the attribution is automatically transferred. **Close matches** are flagged for review by the NHD Editor to determine if there is an actual match or not. If the feature exists in the original, but not in the target, the

> original feature is marked as a "missed feature," which would ultimately lead to the feature **being reviewed** for removal from the NHD Production Database. In the alternative circumstance where the feature exists in the target but not the source, a new ReachCode is applied to the feature and a "create new feature" is automatically generated.
>
> As in any update process the results of the output are only as good as the input. So a majority of the work is actually done during the preparation of the target dataset. The updated geometry in the target dataset must have feature types/codes, connectivity and flow direction for flow lines, and be in the correct NHD schema in order to be used in the GeoConflation process. The end result of the process is a dataset consisting of the new geometry complete with **reach codes and unique identifiers where needed**. All changes to the original data will have been tracked within the model schema. In addition, the target data must pass **required QA/QC checks** before the final data is integrated into the NHD Production Database.

No particulars were found on the USGS website regarding these questions:

- What specified tolerances the USGS uses to compare geometries?
- What constitutes a "close match"?
- Why and how is a feature reviewed for removal when it is not present in the target dataset?
- Why would all new geometries not have reach codes and unique identifiers, as opposed to only having them "where needed"?
- What do the QA/QC checks for final integration entail?

Presumably, these points are all answered in the USGS's training sessions; however, the required training and toolset are neither efficient nor do they promote collaboration—multiple USGS employees have publicly stated the need to simplify the NHD editing process and associated toolset to make it more accessible to the greater scientific and geographic communities. With all this in mind, the BGS approach to determining feature equivalency is not based on the standards and specifications used in-house by the USGS.

PAHD Transfer Steps for NHD Attributes

The BGS starts the transfer of NHD attributes by manually transferring the NHD's Permanent_Identifier field (the only field in the NHD that provides unique identifiers for every flowline segment) to the PAHD flowpaths feature class. To perform the assignment of these values, the Spatial Adjustment toolbar is used to visually match PAHD flowpaths to corresponding NHD flowlines. In this case, an "equivalent feature" refers to the feature in the PAHD that best exemplifies, in generalized order of importance, the NHD feature's

1. Spatial placement
2. Function (FType/FCode)
3. Length/extent
4. Relationship to other features

After the Permanent_Identifier field has been fully transferred to equivalent features in the target dataset, the BGS uses the Join Field tool to automatically add other relevant fields (e.g., FCodes, ReachCodes, GNIS IDs/Names) to the target dataset.

It is important to recognize that the added fields still represent their NHD values and may not be reflective of PAHD or real-world interpretation. For instance, the FCode attribution is

one of the most important fields in the NHD and, presumably, it will be one of the most important in the PAHD. This feature code provides a more specific description of many of the feature types. However, the codes in the NHD have known errors, which the BGS is not attempting to correct at this time (Figure 39).

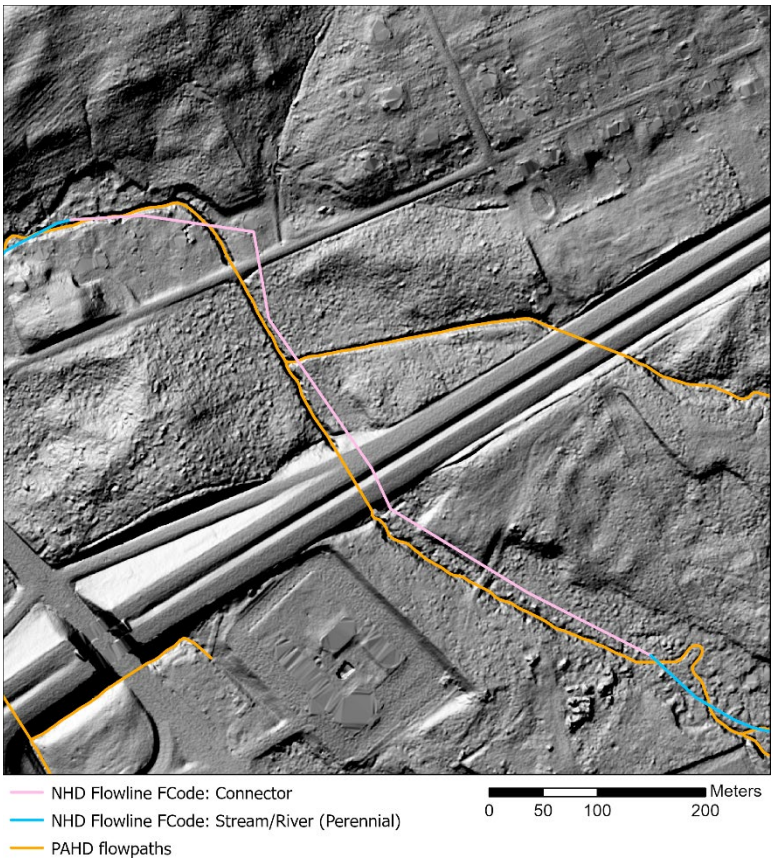


Figure 39. Example from Erie County of an inaccurate NHD attribute that was transferred to the PAHD. Where the PAHD flowpath is correlative to the NHD line segment with the FCode indicative of a connector, it is obvious that the PAHD flowpath is mostly a natural channel feature with two culverts.

Strategies and Examples for Relating NHD Flowlines to PAHD Flowpaths

Example 1: NHD Flowlines With Multiple Permanent Identifiers

The NHD flowline shown in Figure 40A–B is labeled with each feature's Permanent_Identifier value, but there appears to be no reason for the segments with the permanent identifiers 66921509, 66907227, and 66907671 to exist as separate features. Why is this flowline broken up into these segments? There are no convergences or divergences. The attribution of these features (Figure 40) shows that they all have the same FType, FCode, ReachCode, GNIS_ID, GNIS_Name, and FlowDir. It is impossible to fully understand the NHD attribution for this section of the stream without supplemental layers.

In this case, bringing in the NHD Waterbody layer (Figure 40C) makes it possible to infer that the features with the permanent identifiers mentioned above denote the NHD flowline's overlap with the NHDWaterbody SwampMarsh polygon. This example shows how it may occasionally be necessary to bring in additional hydrography layers to provide context for NHDFlowline attribution, as the NHDFlowline attribution was not designed to function as a stand-alone set of data.

OBJECTID *	Shape *	Permanent_Identifier *	FDate	Resolution	GNIS_ID *	GNIS_Name	LengthKm	ReachCode *	FlowDir	WBArea_Permanent_Identifier *	FType	FCode	MainPath	InNetwork	VisibilityFilter	Shape_Length
63395	Polyline ZM	66907227	3/9/2012 8:24:49 AM	High	01198434	Big Run	0.357	02050206001065	WithDigitized	<Null>	StreamRiver	StreamRiver: Hydrographic Category = Perennial	Unspecified	Yes	Approximately 1:250,000 or Larger Scale	355.326458
76752	Polyline ZM	66921509	8/24/2015	High	01198434	Big Run	0.256237	02050206001065	WithDigitized	<Null>	StreamRiver	StreamRiver: Hydrographic Category = Perennial	Unspecified	Yes	Approximately 1:250,000 or Larger Scale	255.217996
330660	Polyline ZM	66907671	3/9/2012 8:24:49 AM	High	01198434	Big Run	1.664	02050206001065	WithDigitized	<Null>	StreamRiver	StreamRiver: Hydrographic Category = Perennial	Unspecified	Yes	Approximately 1:250,000 or Larger Scale	1655.047071

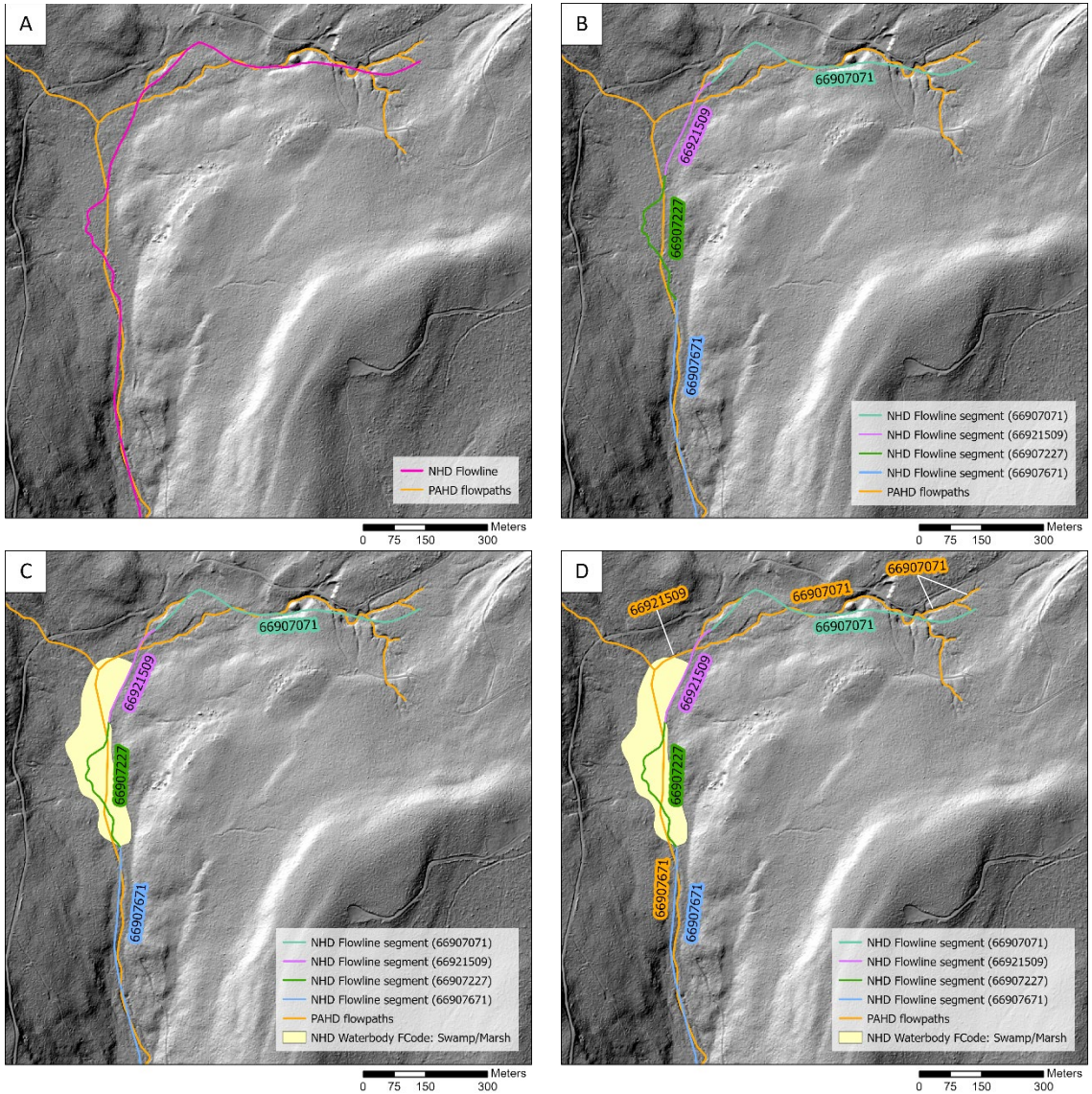


Figure 40. Example from Sullivan County of how NHDFlowline attribution does not always make sense without the context of additional feature layers. **A.** An NHD flowline and PAHD flowpaths shown on a hillshade. **B.** The NHD flowline is shown to have four segments with permanent identifiers of 66907071, 66921509, 66907227, and 66907671, respectively. The first of these segments is an intermittent stream/river, while the other three are perennial. The attribution of this layer (table above) does not show any reason for the 66921509, 66907227, and 66907671 features to exist as separate segments: the only differences between them are their permanent identifier and their FDates. **C.** The same features as in 40B shown with the NHDWaterbody layer. It is now obvious that the reason the perennial flowlines exist as separate segments is to denote where they enter and exit a waterbody feature. **D.** The transfer of permanent identifiers from NHD flowlines to equivalent PAHD flowpaths is illustrated here.

Example 2: Spatial Placement

The correct transfer of permanent identifier values from the NHDFlowline source layer to the preliminary PAHD flowpaths layer shown in Figure 41 can be interpreted by examining spatial placement. The single NH D flowline with permanent identifier 25989620 has an ArtificialPath FCode, which correlates spatially with the PAHD artificial paths representing flow through a wide flow feature. It does not share a similar spatial placement with the PAHD flowpath that offshoots to the right (east) of the main channel (red arrow, Figure 41D). The permanent identifiers from the flowline are therefore transferred to the flowpaths in the wide flow feature and are not transferred to the offshoot feature.

Example 3: Function

Like the example illustrated in Figure 41, the example of attribute transfer shown in Figure 42 shows an NHD Flowline ArtificialPath FCode against more complicated PAHD flowpaths. The difference here is that the "island" feature is not surrounded by PAHD wide flow features (Figures 42C and 42D). Because the PAHD flowpath bordering the southeastern part of the "island" (red arrow, Figure 42C) does not represent a wide flow feature, functionally it is not equivalent to an artificial path. As a result, this flowpath is not considered equivalent to the source NHD ArtificialPath FCode. Instead, this PAHD flowpath segment is more analogous to the NHD Flowline StreamRiver (Perennial) FCode and so receives the permanent identification 25989472 to reflect this functionality and preserve networking relationships.

Example 4: Length/Extent

Consideration of feature length/extent becomes important in determining equivalency when the functionality (FType/FCode) of the source feature cannot be immediately determined for a target feature. For example, in Figure 43, the NHD flowline consists of two segments with two individual permanent identification values. This separation exists because the two segments have different FCodes: the feature identified as 66908825 is a perennial stream/river feature, and the feature identified as 66909099 is an intermittent stream/river feature. Whether or not the PAHD flowpath in Figure 43 could be considered perennial or intermittent cannot be determined from the available elevation data or orthoimagery. Other factors such as spatial placement and relationship to other features are not helpful in this interpretation, and so the permanent identifier assigned to the NHD flowline segment with the most similar length/extent to the source PAHD flowpath should be transferred (Figure 43B).

A Note on Intermittent/Perennial/Ephemeral FCodes: At this time, the BGS does not have a set method for determining a flow feature's status as either intermittent, perennial, or ephemeral based on elevation data; however, work has been done by the Allegheny National Forest using flow accumulation thresholds to quantitatively differentiate these types of features (Charles Keeports, Allegheny National Forest, written communs., 2021). Moving forward, the BGS hopes to incorporate similar methods into this attribution workflow.

Example 5: Relationship to Other Features

Figure 44 demonstrates how network relationships can be used to determine feature equivalency. In terms of branching and overall connectivity, the southerly flowing NHD tributary (permanent identifier 25990418) in these images is equivalent to the southerly

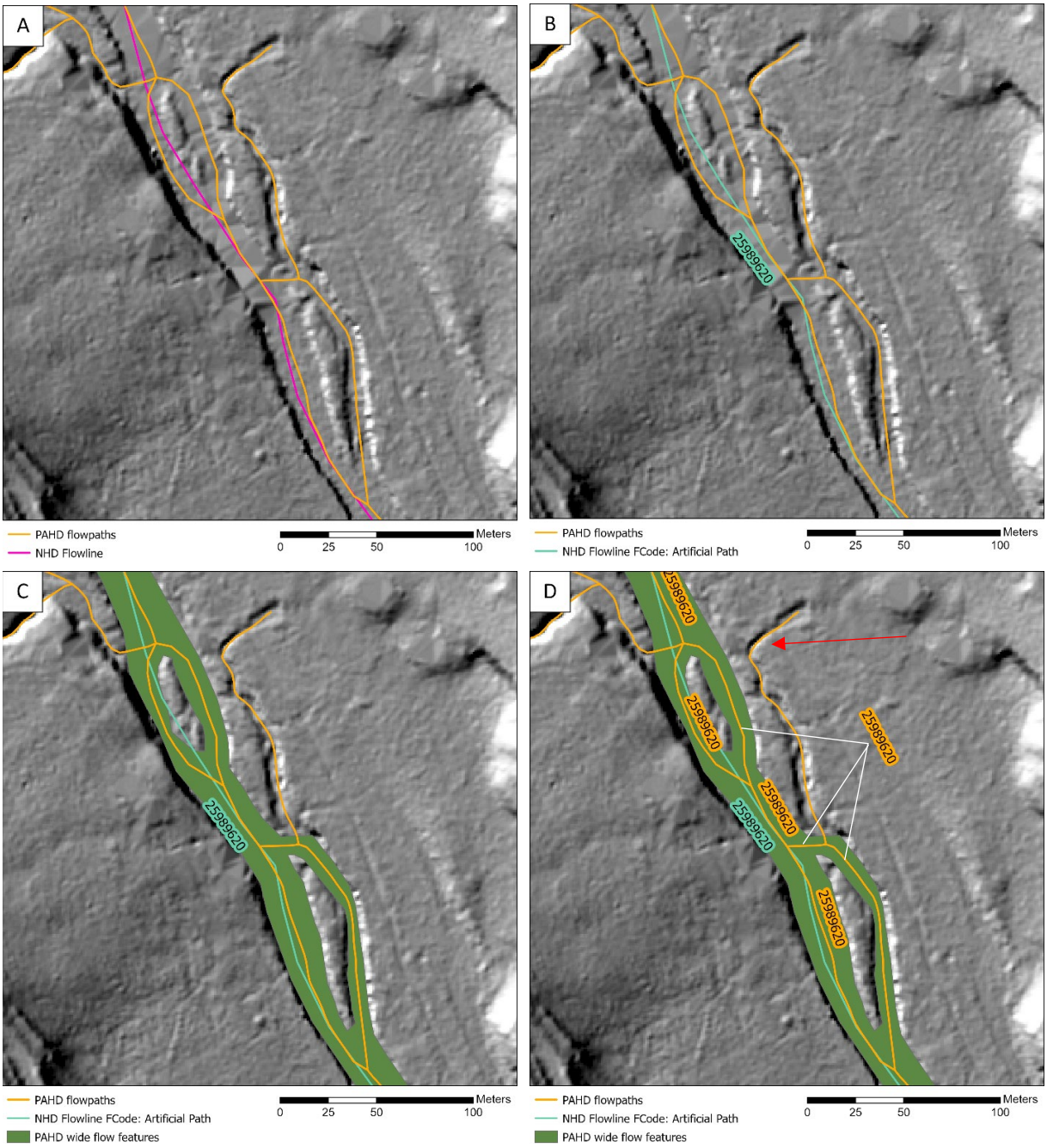


Figure 41. Example from Chester County of how to relate an NHD flowline to PAHD flowpaths using spatial placement. **A.** Islands noticeable in the hillshade were not recognized in the NHD. **B.** The entire flowline is one segment with a permanent identifier of 25989620. Its FCode indicates it is an artificial path (i.e., it is contained in an NHDWaterBody or NHDArea). **C.** The islands are surrounded by PAHD wide flow features. **D.** Spatial placement determines which PAHD flowpaths are most like the NHD flowlines. In this case, the flowpaths within the wide flow features are related to the NHD artificial path designation and have been assigned the identification number of the flowline. They will inherit the FCode attribute of ArtificialPath when the attributes are joined. The flowpath marked by the red arrow is not contained in a PAHD wide flow feature and does not have any NHD flowline equivalent. Therefore, it will not receive any attribute information during this transfer.

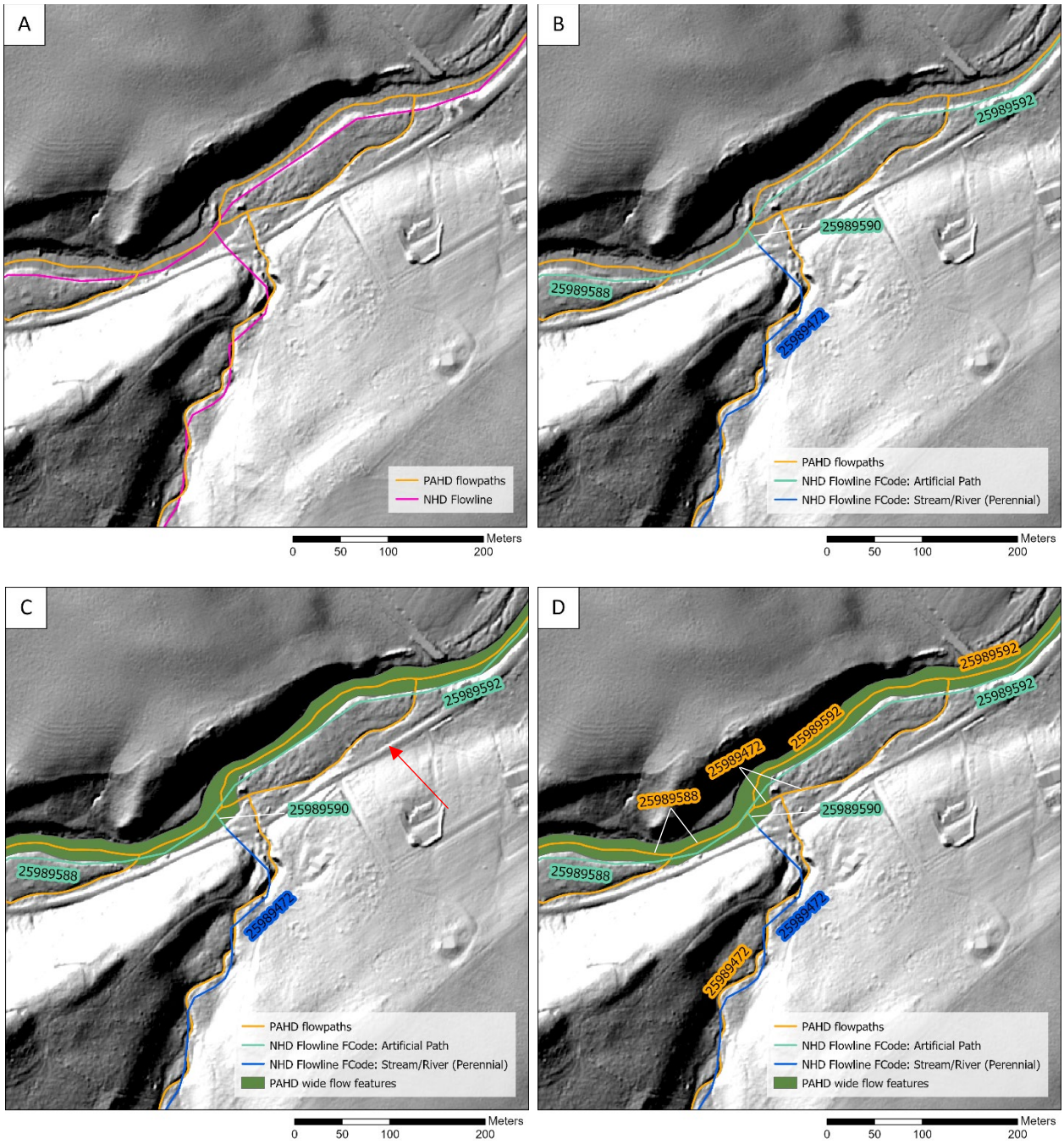


Figure 42. Transfer of permanent identification values from NHD flowlines in Chester County to PAHD flowpaths based on function. **A.** Like the example shown in Figure 41, the NHD flowline shown here does not recognize the existence of an island feature. **B.** The NHD flowline has been color coded to highlight associated FCodes. **C.** The "island" in this example is not surrounded by PAHD wide flow features. While the spatial placement of the PAHD flowpaths surrounding the island is like the NHD flowlines features, the PAHD flowpaths do not have the same functionality. One PAHD flowpath (denoted with a red arrow) is not within the PAHD wide flow feature and is therefore not considered an artificial path. It should not share the FCode with the spatially similar NHD flowline, and so the identification number of that flowline was not transferred to this feature. **D.** Permanent identification values shown on the NHD flowlines and the PAHD flowpaths.

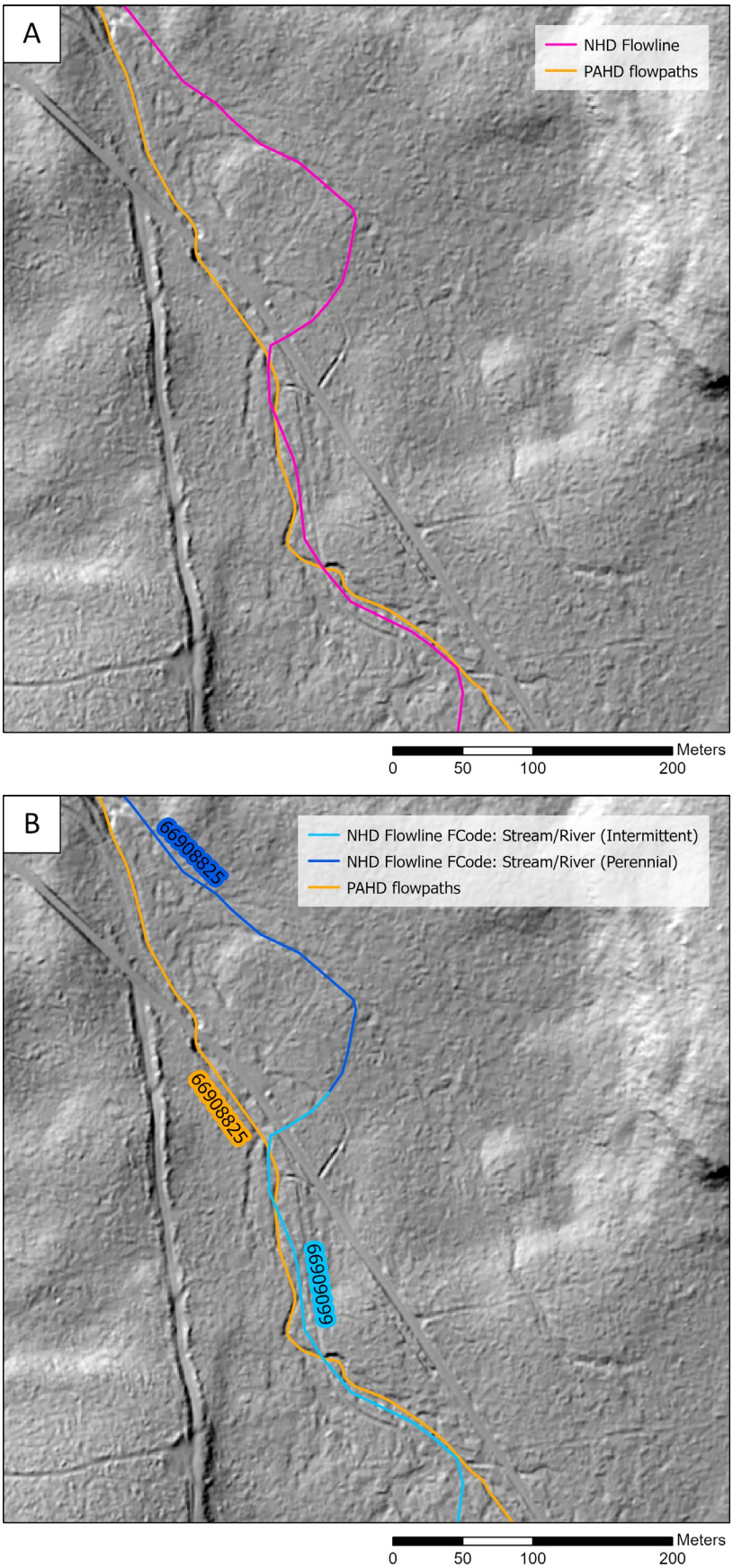
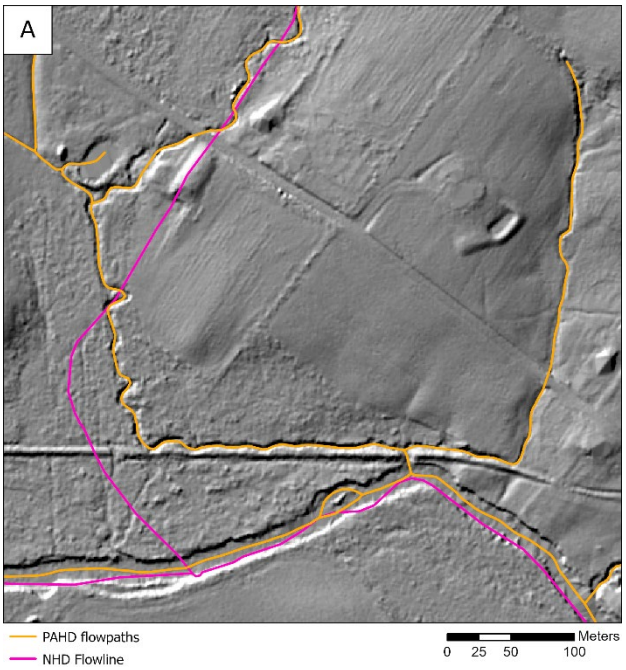
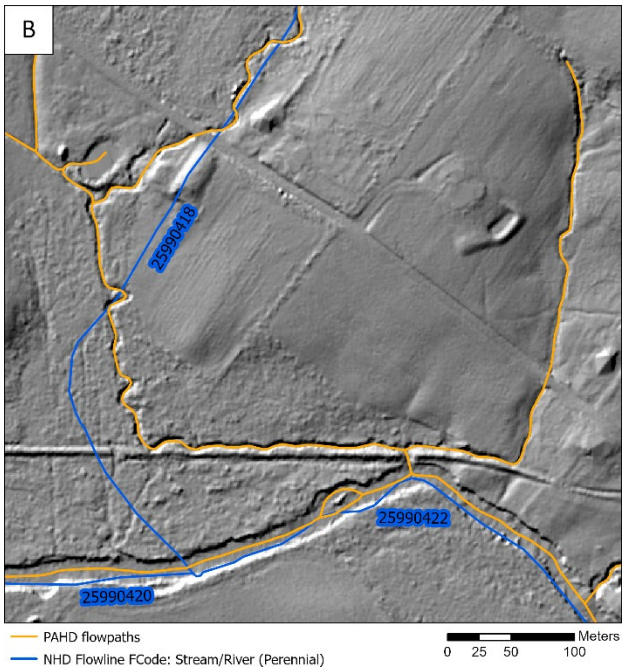


Figure 43. Determining equivalency between an NHD flowline in Sullivan County and a PAHD flowpath by comparing length/extent. **A.** An NHD flowline and a PAHD flowpath shown on a hillshade. **B.** Permanent identification values are shown for the NHD and PAHD features. The NHD line colors makes it apparent that the differing NHD identifier values are the result of different FCodes: one segment is perennial, and the other is intermittent. Currently, it is not possible to assign intermittent, perennial, or ephemeral designations to PAHD flowpaths using only elevation data. Unless there is relevant supplemental data, the PAHD flowpath is assigned the identifier value of the more prevalent (longer) NHD feature.





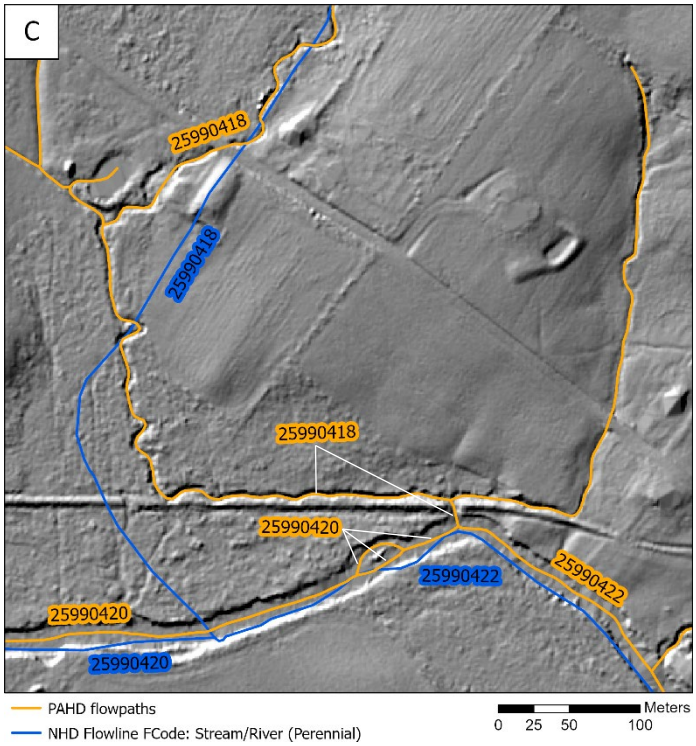


Figure 44. Example from Chester County highlighting networking relationships between features. **A.** NHD flowlines and PAHD flowpaths shown on a hillshade. **B.** Three NHD StreamRiver (Perennial) FCode features shown in relationship to PAHD flowpaths. Flowlines identified as 25990420 and 25990422 might be considered equivalent to the similar trending, adjacent PAHD flowpath. With similar spatial placement, function, and length/extent, the final equivalency of these features is determined using network relationships. The southerly flowing NHD stream/river (number 25990418) is equated to the southerly flowing PAHD flowpath segment it crosses despite having dissimilar spatial placements and extents. This is because the transfer of attributes between these two features will preserve the network relationships of the NHD flowlines. **C.** PAHD flowpaths are assigned NHD identifiers that will preserve the network relationships. Note how the identifying values of the east-west-trending flowpath change on either side of the tributary connection.

flowing PAHD feature it crosses despite an offset of about 200 m between them in the lower reaches. This equivalency preserves the relationship between the three features.

Example 6: When NHD Feature Relationships are Incorrect

The NHD does not always exhibit network relationships that match PAHD flowpaths (Figure 45). In these situations, it is important to keep in mind that, ultimately, spatial placement is the most important factor in determining feature equivalency.

Example 7: Difficult Equivalency

Sometimes it is especially difficult to determine feature equivalency, as is the case with the features shown in Figure 46. Looking at these lines, the main point of contention is where to transfer the attribution of the NHD flowline feature identifed as 66908719. Of the two potential target features in the PAHD flowpaths feature class (red/purple arrows, Figure 46B), which one has the most similar spatial placement, function, length/extent, or relationship to other features?

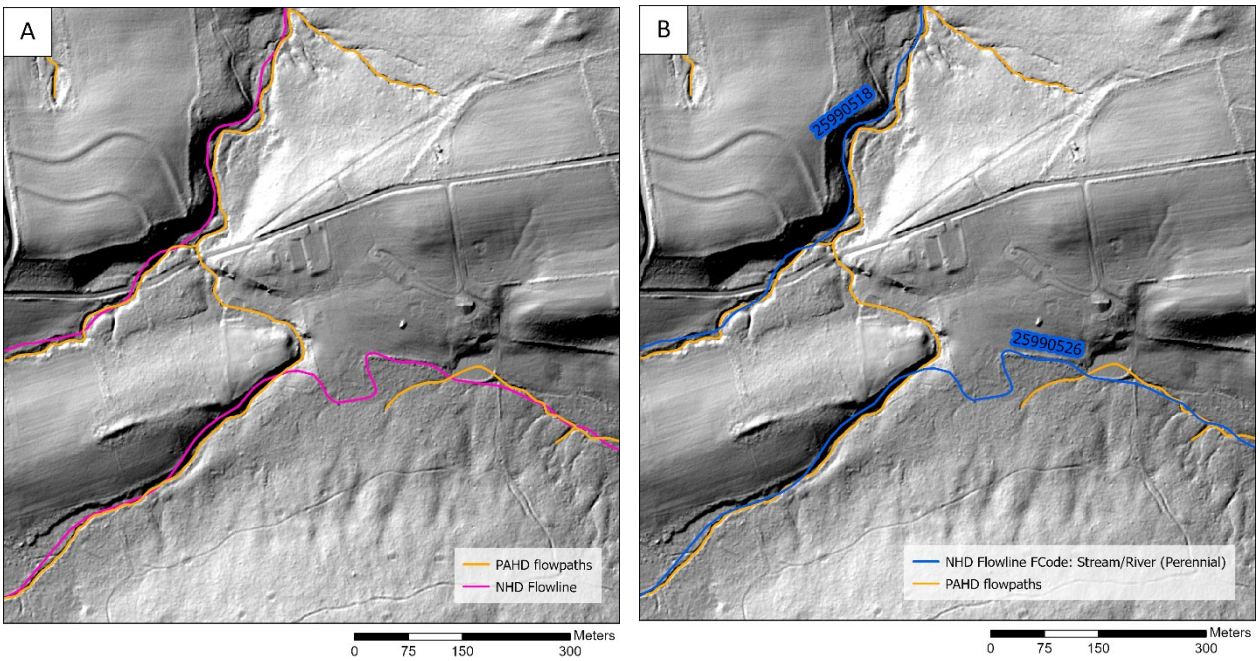
- Spatial extent: Following the NHD flowline 66908719 from upstream to downstream (southeast to northwest), it initially appears that the NHD flowline feature is more similarly placed to the shorter southern PAHD flowpath segment (purple arrow, Figure 46B). About halfway down the shorter southern PAHD flowpath segment, however, the NHD flowline migrates to roughly match the delineation of the longer northern PAHD flowpath segment (red arrow, Figure 46B).
- Function: Both PAHD flowpaths will conceivably have the same functionality, so this does not help with differentiating one from the other in terms of equivalency.
- Length/extent: The length of the NHD flowline is between that of the shorter PAHD flowpath and the longer PAHD flowpath, though it is more similar in length to the longer PAHD flowpath.
- Relationship to other features: The permanent identification value from the NHD flowline could be applied to either PAHD flowpath segments and still maintain the same feature relationships.

In this situation, the interpretation of an "equivalent feature" is somewhat subjective. It is therefore recommended to transfer the original NHD flowline's permanent identification value to both PAHD flowpath features: this will help fill the attribution of both these segments during the later Join Field. The attribution of these features can always be edited later, as may be desired if the USGS publishes guidelines for feature equivalency.

4. EDIT ATTRIBUTES

After any transfer of attributes, there will need to be some amount of QA/QC. This is particularly true when using the NHD as a source. Reasons that these transferred attributes will require careful manual editing include the following:

- NHDFlowline FCodes are inaccurate, and this inaccuracy is maintained during automated attribute transfer (i.e., Join Field).
- Aside from the previously mentioned USGS workshops, there is no published documentation from the USGS explaining how or when new features should be considered "equivalent" to current NHD features.



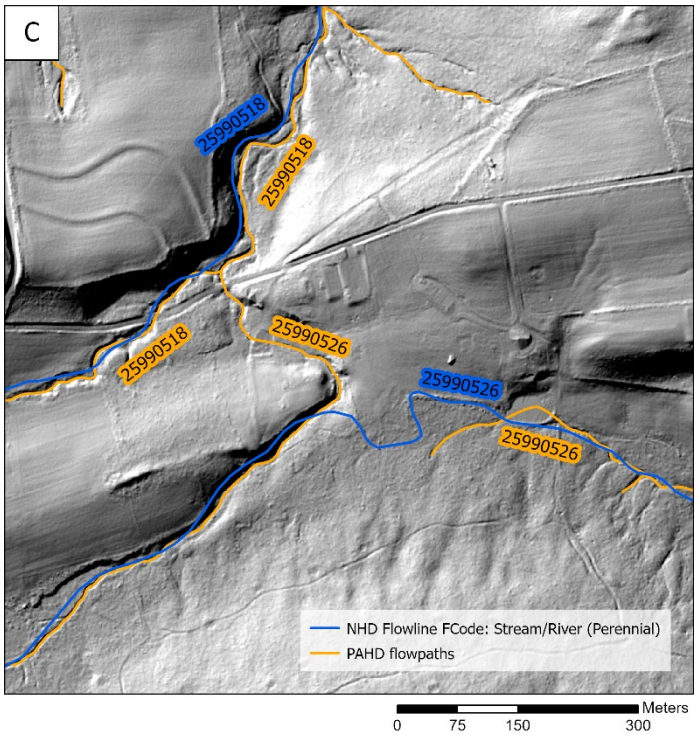


Figure 45. Example from Chester County that shows when to ignore NHD flowline networking relationships. **A.** NHD flowlines and PAHD flowpaths shown on a hillshade. **B.** Permanent identification values labeled on NHD flowline features. Notice the similar spatial placement, function, and length/extent as their PAHD flowpath counterparts. The attribution of NHD feature 25990526 could be transferred to two different PAHD flowpaths based on these factors, but this would destroy the original NHD Flowline network relationships. **C.** The original NHD flowline network relationships shown here are likely incorrect, so the attribution of the NHD flowline 25990526 should be transferred to the two PAHD flowpaths that exhibit similar spatial placement, function, and length/extent. This results in network relations that differ from the original NHD flowlines, but which are ultimately more accurate.

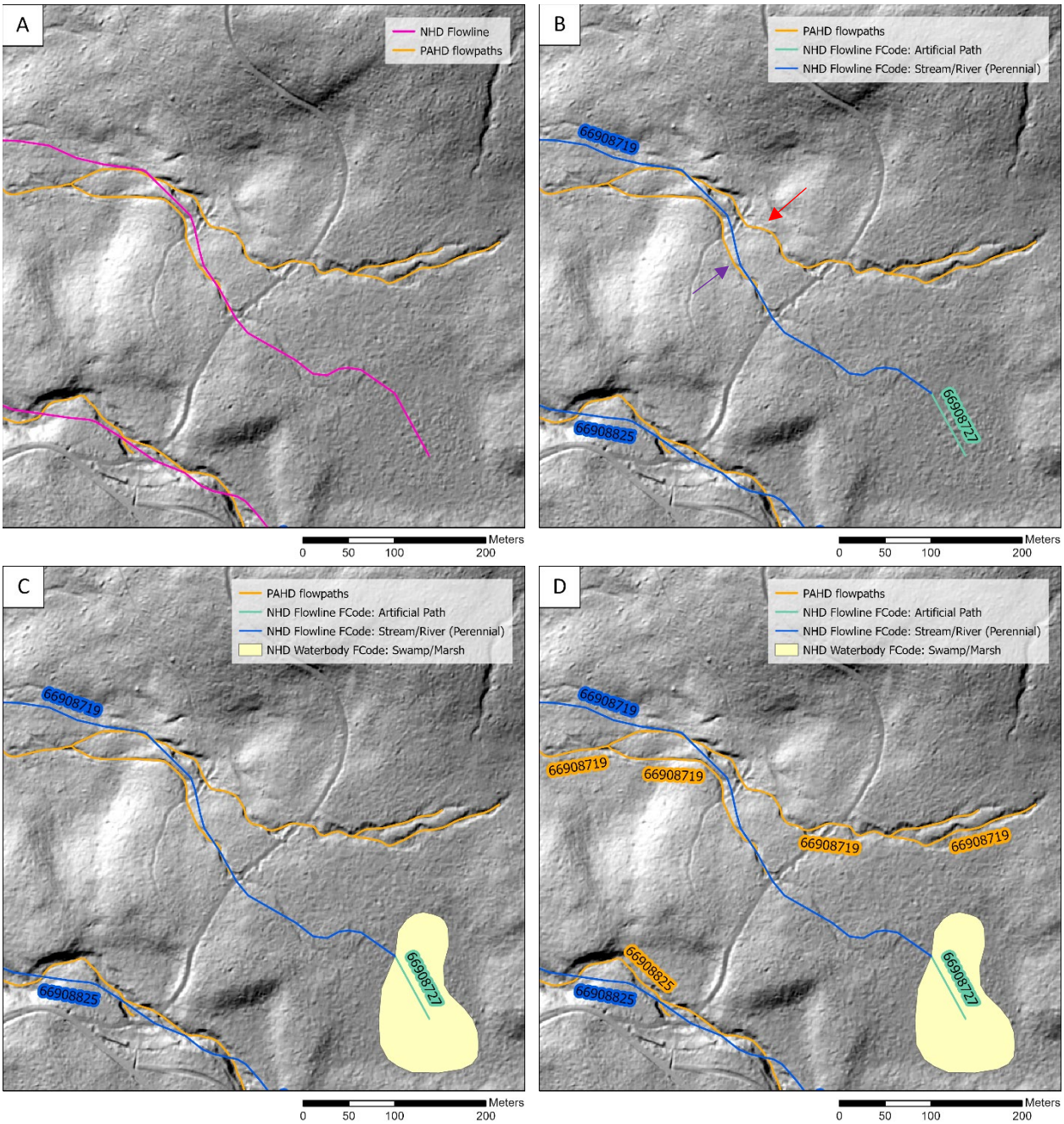


Figure 46. Example from Sullivan County showing a difficult interpretation of flowline-flowpath equivalency. **A.** NHD flowlines and PAHD flowpaths shown on a hillshade. **B.** Flowlines showing permanent identifications and symbolized by FCode. It is difficult to determine equivalency here. The StreamRiver (Perennial) FCode feature with identifier 66908719 could be considered equivalent to two different PAHD flowpath features (purple and red arrows). **C.** The introduction of the NHD Waterbody layer does not clarify which PAHD flowpath should receive attribution. **D.** Because spatial extent, function, length/extent, and network relationships are all similar, the permanent identification number 66908719 was transferred from the NHD feature to both PAHD flowpaths.

5. ASSIGN ADDITIONAL ATTRIBUTES

Certain fields specific to the methodology outlined in this document will need to be manually added. Luckily, this is neither particularly difficult nor time-consuming. Examples of these fields include things like DataSource (i.e., the source of the data from which PAHD flowpaths were generated; e.g., Quantum Spatial QL2 Lidar derivatives and PEMA leaf-off orthoimagery from 2018) and Method (i.e., the method or methods used to delineate PAHD flowpaths; e.g., geomorphon analysis, least-cost analysis). These attributes can commonly be applied en masse to the flowpath segments within a single HU12.

6. CREATE A NETWORK

A Note About Networks

In ArcGIS products, two types of networks are recognized: (1) geometric networks and (2) network datasets. Geometric networks model the movement of a feature or features that act under natural forces (e.g., river networks, electrical networks, and sewer lines), whereas network datasets model the movement of a feature or features that can "decide" a direction (e.g., transportation networks like streets, pedestrian walkways, and railroads). A geometric network therefore best serves the purpose of this project. Since the Esri Utility Network extension only operates on and with network datasets, this license is not necessary for constructing PAHD flowpath networks.

Network Inclusivity

PAHD flowpaths have been organized by HU12 geometric networks. At some point, these networks will be coalesced into larger and larger regions with the end goal of a single network with statewide coverage.

The EDH Acquisition Specifications (Terziotti and Archuleta, 2020) outline a number of limitations to avoid creating a confusing network [boldface added by author]:

> Canal and ditch features **that are too detailed for a hydrologic network** (see "Canals and Ditches" subsection of "Special Cases" in the "Delineation of Hydrographic Features" section) should not be coded as Fcode=33600. These features should be coded as described previously for "additional user-defined features," and a UserCode should be added to the table to distinguish these features. Text that describes the feature, such as "canal or ditch for project use" may be added to the Desc field.

This implies that when canal and ditch features are "too detailed" they should not be included in the final network—it does *not* indicate that these features should not be collected at all.

In the "Canals and Ditches" subsection referenced in the above quote, it is explained that

> The problem with many interconnected canals as a part of a hydrologic network is that it becomes difficult to navigate the stream network. The direction of flow is not always apparent and may vary based on weather or season. Creating a network that flows in a downstream direction, with z-values descending from upstream to downstream, is a challenge when many segments with low slope are present. Culverts are also difficult to identify for smaller canal systems, leading to isolated networks.

The same subsection includes READ Rules (listed below) that help a creator "avoid over-collection" of canal/ditch features. Presumably, these rules are also intended to avoid the inclusion of overly complicated canal/ditch features in the final network; however, this is never explicitly stated.

- If a canal/ditch is named, or if a canal/ditch is greater than or equal to 984 ft (300 m) along the longest axis, then it should be captured.
- If a canal/ditch is needed to provide network connectivity, it should be captured.
- If a canal/ditch feature is within agricultural fields, is less than 300 m, and drains to another canal/ditch or other hydrologic feature, it should not be captured.
- Isolated canal/ditch features or networks should not be collected unless they are of particular importance to the regional hydrology.
- If a project has a special need for canal/ditch features not covered by these capture conditions, a separate UserCode attribute should be added to allow those features to be removed from an NHD conflation process.

The problem with the above statements is that, initially, the EDH Acquisition Specifications imply that *all* canal/ditch features should be collected regardless of detail (see first quote and discussion in this section above). But the above list of READ Rules appear to prohibit the collection of these overly detailed features, unless a project has a special need for these features.

This contradiction raises two issues: (1) restricting canal/ditch feature collection, excepting in the case of a specific project's needs, will result in inconsistent feature collection, and (2) these rules preclude the collection of many new canal/ditch features. Furthermore, the justification of excluding features from collection because they are "too detailed" for a network is not sound. If the issue is whether a flow feature's directionality is indeterminant, then that feature (or even collection of features) can be disabled while still being a part of a greater network.

Any feature in a geometric network can be enabled or disabled. An enabled feature allows flow to pass through it, while a disabled feature does not. Disabling features allows the user to treat these features as if they were disconnected from the network without removing the topological connections that they have to other features in the network. By default, all features in a geometric network are enabled when the network is created. To enable or disable a feature, the feature's Enabled attribute must be edited in the Attributes dialog box. This can be done by opening the attribute table of the relevant feature class, finding the feature in the table, and editing the value (True to enable, False to disable) for this attribute.

DISCUSSION AND COLLABORATION

The BGS is open to and enthusiastic about the prospect of collaborative work. The author encourages any interested parties to review this workflow and to submit

- Hydrography data for areas of interest based on this methodology
- Existing hydrography data that may be validated based on this methodology
- Critique of this methodology and suggestions for improvement
- Suggested means of automation, such as scripts

AUTOMATION

One of the major goals of the BGS is to automate as many of the workflow steps as possible. Automation could be established through coding, new tools, or any number of other methods. Sections of this workflow that are most time-consuming, such as the manual delineation of culvert/bridge features and the manual correction of gaps in flowpath geometries, would also derive the most benefit from automation.

FEATURE IDENTIFICATION

The Permanent_Identifier field is the NHD's primary means of differentiating one feature from another. The PAHD is very likely to adopt a similar system: a GUID datatype that uniquely identifies a feature or table row within a geodatabase and across geodatabases. Having said this, the BGS would very much appreciate suggestions on identification structure. They wish to consider this and to wait for the initial datasets to be vetted before adopting an identifier system.

DATA ACCEPTANCE

PAHD inclusion is not based on methodology; rather, it is based on a set of standards to assess data quality. Standards for acceptance of user-submitted geometries and geometry corrections include the following:

- Data must be derived from QL2 lidar or better
- Attribution should include specifics on source data and derivation methods
- Data should have a minimum of 1-m horizontal accuracy and 30-cm vertical accuracy (with reference to the most current elevation data)

ACKNOWLEDGMENTS

The author greatly appreciates the project guidance and technical mentorship of Mike Moore (BGS, retired), Eric Jespersen (JTM), and Matt Mercurio (CivicMapper); the advice and constructive criticism provided by David Saavedra (Chesapeake Conservancy), Lisa Mirth (Apartments.com), Matthew Baker (University of Maryland, Baltimore Campus), Eliza Gross (USGS), Steve Aichele (USGS), Roger Barlow (USGS), and David Anderson (USGS); the extensive review performed by Caron Pawlicki (BGS); the final review done by Gale Blackmer (BGS); and the input and support of Craig Ebersole (BGS), members of the Lidar Working Group Hydrography Applications Network, and attendees of the 2020 and 2021 Hydrography Workshops.

REFERENCES

Archuleta, C. M., and Terziotti, Silvia, 2020, Elevation-derived hydrography—representation, extraction, attribution, and delineation rules: U.S. Geological Survey Techniques and Methods, book 11, chap. B12, 60 p., https://doi.org/10.3133/tm11B12. [Accessed July 1, 2020.]

Baker, Matthew, Saavedra, David, and Norton, Michael, 2018, Scope#10—Methodology for developing high-resolution stream and waterbody datasets for the Chesapeake Bay watershed: Report to Chesapeake Bay Trust, 59 p., https://cbtrust.org/wp-content/uploads/2018_Scope10_FinalReport_wAppendix.pdf. [Accessed July 10, 2019.]

Esri, 2016, Creating sources and sinks: Esri ArcMap web page, https://desktop.arcgis.com/en/arcmap/10.3/manage-data/geometric-networks/setting-flow-direction-creating-sources-and-sinks.htm. [Accessed September 2020.]

Heidemann, H. K., 2018, Lidar base specification (ver. 1.3, February 2018): U.S. Geological Survey Techniques and Methods, book 11, chap. B4, 101 p., https://doi.org/10.3133/tm11b4. [Accessed July 8, 2019.]

Jasiewicz, Jarosław, and Stepinski, T. F., 2013, Geomorphons—a pattern recognition approach to classification and mapping of landforms: Geomorphology, v. 182, p. 147–156, https://doi.org/10.1016/j.geomorph.2012.11.005. [Accessed July 8, 2019.]

Kovesi, Peter, 2010, Fast almost-Gaussian filtering, *in* DICTA '10—Proceedings of the 2010 International Conference on Digital Image Computing—Techniques and Applications: Washington, D.C., IEEE Computer Society, p. 121–125, https://doi.org/10.1109/DICTA.2010.30. [Accessed December 5, 2019.]

Lukin, Alexey, 2007, Tips & tricks—Fast image filtering algorithms: GraphiCon'2007, Moscow, Russia, 2007, [4] p., https://www.researchgate.net/publication/228758794_Tips_tricks_Fast_image_filtering_algorithms. [Accessed April 13, 2020.]

Pennsylvania Environmental Council, 2016, An introduction to MS4 stormwater permitting: Pennsylvania Environmental Council web page, https://pecpa.org/program-update/introductiontostormwaterpermitting. [Accessed September 2020.]

Smith, S. W., 1999, The scientist and engineer's guide to digital signal processing (2nd ed.): San Diego, Calif., California Technical Publishing, 650 p., https://www.analog.com/en/education/education-library/scientist_engineers_guide.html#. [Accessed April 13, 2020.]

Stepinski, T. F., and Jasiewicz's, Jarosław, 2011, Geomorphons—a new approach to classification of landforms: Geomorphometry, Redlands, Calif., 2011, p. 109–112, http://www.geomorphometry.org/system/files/StepinskiJasiewicz2011geomorphometry.pdf. [Accessed July 8, 2019.]

Terziotti, Silvia, and Archuleta, C. M., 2020, Elevation-derived hydrography acquisition specifications: U.S. Geological Survey Techniques and Methods, book 11, chap. B11, 74 p., https://doi.org/10.3133/tm11B11. [Accessed July 1, 2020.]

U.S. Geological Survey, [2010], NHD user guide: U.S. Geological Survey Hydrography web page, https://nhd.usgs.gov/userguide.html. [Accessed July 2019.]

_____, 2018, Lidar base specification (version 2.0): U.S. Geological Survey, National Geospatial Program, [40] p., https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/DRAFT%20Lidar_Base_Spec_v_2-0.pdf. [Accessed July 2019.]

_____, 2020, Lidar base specification (version 2020, rev. A): U.S. Geological Survey, National Geospatial Program, [47] p., https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/Lidar-Base-Specification-2020-rev-A.pdf. [Accessed August 2020.]

ADDITIONAL READING

Newcomb, Doug, 2015, Introduction to GRASS GIS with terrain analysis examples [revised, March 2016]: GRASS-Wiki web page, https://grasswiki.osgeo.org/wiki/Introduction_to_GRASS_GIS_with_terrain_analysis_examples. [Accessed August 2019.]