

National Water Quality Program

An Enhanced Hydrologic Stream Network Based on the NHDPlus Medium Resolution Dataset



Scientific Investigations Report 2019–5127

An Enhanced Hydrologic Stream Network Based on the NHDPlus Medium Resolution Dataset

By John W. Brakebill, Gregory E. Schwarz, and Michael E. Wieczorek

National Water Quality Program

Scientific Investigations Report 2019–5127

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DAVID BERNHARDT, Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2020

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Brakebill, J.W., Schwarz, G.E., and Wiczorek, M.E., 2020, An enhanced hydrologic stream network based on the NHDPlus medium resolution dataset: U.S. Geological Survey Scientific Investigations Report 2019–5127, 49 p., <https://doi.org/10.3133/sir20195127>.

Associated data for this publication:

Schwarz, G.E., 2019. E2NHDPlusV2_us: Database of Ancillary Hydrologic Attributes and Modified Routing for NHDPlus Version 2.1 Flowlines: U.S. Geological Survey data release, <https://doi.org/10.5066/P986KZEM>.

Foreword

Sustaining the quality of the Nation's water resources and the health of our diverse ecosystems depends on the availability of sound water-resources data and information to develop effective, science-based policies. Effective management of water resources also brings more certainty and efficiency to important economic sectors. Taken together, these actions lead to immediate and long-term economic, social, and environmental benefits that make a difference to the lives of the almost 400 million people projected to live in the United States by 2050. (<http://water.usgs.gov/nawqa/applications/>).

In 1991, Congress established the National Water-Quality Assessment (NAWQA) to address where, when, why, and how the Nation's water quality has changed, or is likely to change in the future, in response to human activities and natural factors. Since 1991, NAWQA has been a leading source of scientific data and knowledge used by national, regional, state, and local agencies to develop science-based policies and management strategies to improve and protect water resources used for drinking water, recreation, irrigation, energy development, and ecosystem needs. Plans for the third decade of NAWQA (2013–23) address priority water-quality issues and science needs identified by NAWQA stakeholders (such as the Advisory Committee on Water Information) and the National Research Council. The plans are designed to meet increasing challenges related to population growth, increasing needs for clean water, and changing land use and weather patterns.

Federal, state, and local agencies have invested billions of dollars to reduce the amount of pollution entering rivers and streams that millions of Americans rely on for a variety of water needs and biota rely on for habitat. Understanding the sources and transport of pollution is crucial for designing strategies to improve water quality. Studies have indicated that, for example, contaminants such as excess total nitrogen and total phosphorus from sources such as wastewater treatment plants or commercial fertilizer contribute to poor water-quality conditions in streams. Understanding of the role of nutrients in determining water-quality conditions is important, and models, supported by adequate data, are important tools for developing the information needed to manage water quality.

Digital hydrologic networks are a key component to any spatial analysis or modeling study. This report describes a topologically reconditioned (enhanced) version of the routing and stream connections to the value-added attributes of NHDPlusV2.1 with the intent of improving the networking capabilities and assigned attributes utilized by SPARROW modeling and other hydrologic routing and attribute accumulating applications.

We hope this publication will provide you with insights and information to meet your water resource needs and will foster increased citizen awareness and involvement in the protection and restoration of our Nation's waters. The information in this report is intended primarily for those interested or involved in resource management and protection, conservation, regulation, and policymaking at the regional and national level.

Dr. Donald W. Cline
Associate Director for Water
U.S. Geological Survey

Contents

Foreword	iii
Abstract	1
Introduction.....	1
Purpose and Scope	3
Material and Methods.....	3
NHDPlusV2.1	3
Constructing a Modified Network.....	8
Filling In Missing Values	8
Routing Modifications.....	11
Addition/Removal Events.....	13
New Variables for Interpreting Network Flow Characteristics	15
Flowline Termination	15
Foreign Drainage	16
Diversion-Affected Flowlines	16
Revised Variables	16
Hydro-sequence	16
Streamflow.....	16
Slope	16
Additional Ancillary Information	16
Aggregations	17
Validation.....	18
Summary and Conclusions.....	21
References Cited.....	22
Appendix 1. Description of Addition and Removal Events Spreadsheet	26
Appendix 2. Description of Methods Used to Update Streamflow Estimates.....	29
Appendix 3. Description of Methods Used to Update Slope Estimates	32
Appendix 4. Description of Attributes in E2NHDPlusV2_us	43
Appendix 5. Description of Selected Ancillary Geospatial Dataset Variables Assigned to the Catchments and Flowlines of NHDPlusV2.1	45

Figures

1. Diagram showing example of how routing of NHDPlusV2.1 flowlines can be affected by an upstream diversion	6
2. Map showing NHDPlusV2 production units.....	9
3. Map showing example of routing errors associated with diversions affecting measured streamflow at a U.S. Geological Survey streamgage in Michigan.....	12
4. Schematic diagram of <i>A</i> , A simple routed transfer and <i>B</i> , a routed transfer with a diversion.....	14
5. Map showing location of Addition/Removal events included in the enhanced hydrologic stream network, by event class	15
6. Graph showing NHDPlus upstream watershed areas in relation to total upstream accumulation values	19
7. Graphs showing Pearson's Correlation for basin percent of GAGESII NLCD 2006 land-use classes in relation to values from the total upstream aggregation method	20

Tables

1. Primary spatial features of the NHDPlusV2 network, divided into sub-categories, with number of occurrences of each.....	4
2. Classes of diversions in NHDPlusV2.1 and their defining characteristics	5
3. Revised names given to NHDPlusV2 variables.....	10
4. Summary of modification method codes for NHDPlusV2 variables.....	11
5. Diversion-related modifications to NHDPlusV2.....	13
6. Number of discrepancy values from attempt to reconstruct original NHDPlus estimates	19

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
Length		
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
Area		
square meter (m ²)	10.7639	square foot (ft ²)
square kilometer (km ²)	247.1	acre
Mass		
kilogram per year (kg/yr)	2.205	pound avoirdupois per year (lb/yr)
metric ton (t)	2204.62	pound, avoirdupois (lb)
millimeter per year (mm/yr)	0.0393701	inch per year (in/yr)
Application rate		
metric ton per square kilometer (t/km ²)	8.92179122	pound per acre (lb/acre)

Datum

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Abbreviations

AR	Addition and Removal
DEM	Digital Elevation Model
EPA	U.S. Environmental Protection Agency
NAWQA	National Water-Quality Assessment
NED	National Elevation Dataset
NHDPlus	National Hydrography Dataset Plus
NHDPlusV2	National Hydrography Dataset Plus version 2
NRSA	National River and Stream Assessment
NWIS	National Water Information System
SPARROW	SPAtially Referenced Regressions On Watershed attributes
USGS	U.S. Geological Survey
WSA	Wadeable Stream Assessment

An Enhanced Hydrologic Stream Network Based on the NHDPlus Medium Resolution DataSet

By John W. Brakebill, Gregory E. Schwarz, and Michael E. Wieczorek

Abstract

The National Hydrography Dataset Plus, Version 2.1 (NHDPlusV2.1) is an attribute-rich digital stream network for the conterminous United States, serving as a foundational infrastructure for reporting hydrologic information at both regional and national scales. SPATIally Referenced Regressions On Watershed attributes (SPARROW) is a process-based statistical model that relies on a digital hydrologic network like NHDPlusV2.1 to establish spatial relations between quantities of monitored contaminant loads and contaminant sources, accounting for the physical characteristics along flow paths affecting contaminant transport. The U.S. Geological Survey National Water Quality Assessment project adopted and modified the medium-resolution NHDPlusV2.1 network for use as the primary framework supporting SPARROW modeling. This report describes the enhancements made to improve the routing capabilities and the value-added attributes of NHDPlusV2.1 to support modeling and other hydrologic analyses. These enhancements include corrections to inconsistencies in network/routing information, filling in missing attribute values of associated characteristics, accounting of water use affecting flow, new variables useful for interpreting network data, revised flowline attributes such as slope and flow, and incorporation of ancillary spatial data into the network. The resulting dataset containing the enhancements to the network is named E2NHDPlusV2_US. Although the enhancements described in the report were developed for use in SPARROW modeling, the enhancements are expected to be useful for a wide variety of hydrologic studies within the United States.

Introduction

Digital hydrologic networks are a key component of any hydrologic spatial analysis or modeling study. A network of connected surface-water pathways, and the areas they drain, is often used as a foundation for a standard, spatial framework for the characterization and analysis of aquatic and watershed processes. Explicit, discrete spatial units can help interested parties visualize and locate the assigned physical properties of

a watershed system (Hill and others, 2016; Sprague and Gronberg, 2012; Brakebill and others, 2011). Geospatial data representing landscape, aquatic, and subsurface watershed characteristics, such as contaminant monitoring and potential supply, land cover and use, slope, soil characteristics, and annual streamflow, are often used to help describe and characterize these watershed systems. These geospatial descriptions help formulate a better understanding of the intrinsic connections between land and water, and their interactions (Alexander and others, 2007; Brakebill and others, 2011). Spatial referencing afforded by these methods has demonstrated improvement in the predictive capabilities of hydrologic models (Smith and others, 1997). Properly connected hydrologic networks provide additional ability to simulate and route the movement of water and associated constituents or properties throughout the landscape. Hydrologic connectivity also can help facilitate the computation of total upstream landscape information for any location along a river system, including river systems with braided channels or diversions.

The medium resolution National Hydrography Dataset Plus, Version 2.1 (NHDPlusV2.1) is an attributed geospatial stream network developed by the U.S. Environmental Protection Agency (EPA) and U.S. Geological Survey (USGS) from USGS 1:100,000 scale topographic maps (McKay and others, 2012). The data structure of NHDPlusV2.1 includes upstream and downstream navigational capabilities, feature naming, and “value-added attributes” (VAAs; McKay and others, 2012) of relevant stream and watershed characteristics. Attributed navigational variables relate stream segments to each other according to spatial topology (hydrologic sequencing) and are essential for routing streamflow and performing other aggregations of associated attributed information. Unique identifiers (FeatureID and ComID) are used as primary keys for identification of waterbody features (McKay and others, 2012; Moore and Dewald, 2016) and relate the waterbodies to specific stream segments (flowlines) in the routed network. The NHDPlusV2.1 contains polygons (catchments) derived from the USGS National Elevation Dataset (NED) that delineate areas draining each flowline (McKay and others, 2012). Additional variables derived from the NHDPlusV2.1 data that are useful for modeling include estimates of mean annual and monthly flow, mean annual and monthly stream velocity, and channel slope.

The comprehensive NHDPlusV2.1 hydrologic network with VAAs has wide applications for hydrologic analyses. It provides and presents a profound foundational infrastructure for organized hydrologic information. Previous versions of NHDPlus have served as the hydrologic network for regional studies (Hill and others, 2016; Tsang and others, 2014; Moore and Dewald, 2016), including regional applications of the USGS-developed SPATIALLY Referenced Regressions On Watershed attributes (SPARROW) empirical water-quality model (Moore and Dewald, 2016; Hoos and others, 2013, Ator and others, 2011), and NHDPlus is the spatial network for the National Oceanic and Atmospheric Administration (NOAA) National Water Model (NOAA, 2016). The USGS National Water Quality Assessment (NAWQA) Cycle III project adopted the medium-resolution NHDPlusV2.1 network to serve as the primary framework supporting SPARROW modeling because it provides a structure for the Nation's drainage system, has rapidly become a uniform reporting unit for hydrologic information, and is available at a national scale (Wieczorek and others, 2018; Schwarz and Wieczorek, 2018; Schwarz, 2019; Wieczorek and others, 2019). The NHDPlusV2.1 also provides a unique navigational infrastructure and framework to add additional spatially referenced attributes to suit relative modeling and assessment applications, including elevation, precipitation, mean annual streamflow, and velocity (McKay and others, 2012).

Despite these advantages and the many improvements since the NHDPlus framework inception, the NHDPlusV2.1 network and associated data exhibit some noteworthy limitations that indicate the need for modifications and enhancements. Existing inconsistent VAAs and attributed navigational information, or duplicate flowlines, can have significant local effects on routing and analysis. Values for each flowline or ancillary information associated with the network (inherent or additional) could be missing for some observations, depending on the origin of the variable or the method of association. Some values, like stream name, simply are not known. Other VAAs, like precipitation, may be missing because a catchment does not exist for that flowline, thus rendering null any aerial geospatial overlay operations designed to capture mean values, by catchment.

Divergences in the stream network, where streamflow (a portion or all the streamflow) is diverted from the natural course of a river, are numerous (more than 64,000 are represented). These divergences present the opportunity for topological errors in the data, such as inadvertently routing streamflow in the wrong direction or through an inappropriate pathway and can result in large errors in watershed parameters derived from the network (for example, drainage area and mean annual precipitation of the basin). Accurate routing and basin parameterization are essential to the development of any hydrologic model. Routing errors can potentially induce large discrepancies between the monitored (or calibration) values and the corresponding model estimates for empirical models. Even if monitoring sites are excluded from diversion-affected rivers, routing errors can still magnify errors in the predictions

of the affected flowlines. Additionally, some corrections or modifications in network routing can render network attributes obsolete or insignificant, particularly variables dependent on the assigned primary pathway, such as channel slopes, hydrologic sequencing, and variables pertaining to aggregations or accumulations of local stream segments or watershed values. Any corrections or modifications in routing require redevelopment of these types of attributes.

The NHDPlusV2.1 network includes attributes which define flowlines as primary, secondary, or divergent pathways. Divergent flowlines also can represent pathways that have been artificially diverted to other locations. This distinction provides a framework for attribution of the fraction of flow assigned downstream along the primary and secondary pathways. However, these diversions and pathways could benefit from a more robust assessment of the proportion of flow that is routed through each of the diverting pathways (or flowline if that is used). The assessment has typically been based on circumstantial factors such as stream name and not observed streamflow data (McKay and others, 2012).

Addition and Removal (AR) events are locations along the network where artificial transfers of water that mimic the rerouting of streamflow occur. The AR events are designed to capture and account for irrigation withdrawals and returns, drinking-water withdrawals, permitted discharges, or other streamflow redirections without creating the necessary physical flowlines to support the transfer. The AR events included with NHDPlusV2.1 represent only a small sample of all such events, but generally for a stream network of this scale, a significant volume of transfers is captured. However, the sources for rerouted flow estimates referenced in McKay and others (2012) could be expanded as opportunities arise to incorporate revisions into the network to account for additional AR events or routing modifications. The U.S. Environmental Protection Agency (EPA) Permit Compliance System (EPA, 2018) database, and representations of aqueducts, canals, conduits, pipes, irrigation ditches, and flowlines corresponding to flow paths that lack flow direction, also could be explored to define and refine AR withdrawal and discharge events.

A topologically reconditioned (enhanced) version of the NHDPlusV2.1 routing and stream connections was created with the intention of improving the networking capabilities and assigned attributes utilized by SPARROW modeling and other hydrologic routing and attribute accumulating applications. The enhancements include corrections to recently discovered inconsistencies in network/routing information, filling in of previously missing attribute values of associated characteristics, revisions in flow routing, expanded accounting of water use affecting flow, creation of new variables useful for interpreting network data, revised flowline attributes such as slope and flow, and incorporation of ancillary spatial data into the network. The resulting enhanced network, available in a single data file for the entire conterminous United States, has the designation E2NHDPlusV2_us. Additional assigned ancillary data can be found in Wieczorek and others (2019).

Purpose and Scope

This report describes a topologically reconditioned version of the NHDPlusV2.1 routing network, including enhanced attributes to improve flow routing and the physical description of flowlines. The modifications pertain only to existing NHDPlusV2.1 data-table values; no changes, enhancements, nor modifications were made to the spatial aspects of NHDPlusV2.1 (McKay and others, 2012). Shapefile flowlines depicting stream channel position and shape, and polygon shapefiles that define and delineate catchments, were unaltered. The enhancements described here also do not modify any presumed artificial flowpaths through waterbodies, despite the arbitrary nature of these pathways and any limitations they induce in describing contaminant processing in the waterbodies.

The approaches described in this report assume the pre-existence of the NHDPlusV2.1 network and its associated attributes; however, many of the attributes can be generated without reference to existing values. The enhanced dataset (E2NHDPlusV2_us, superseding ENHDPlus_us; Schwarz and Wieczorek, 2018) is the result of methods described and referenced within this report and represents a national framework useful for the hydrologic indexing of spatial data like contaminant sources, terrestrial and aquatic properties, characteristics, and information (Schwarz and Wieczorek, 2018; Wieczorek and others, 2018). The data are useful to hydrologic modeling, such as the SPARROW water-quality model, and other analyses relating spatial features to hydrologic data. Data described here and associated with the network flowlines and catchments are publicly available through separate data releases by USGS, which are referenced in this report.

Material and Methods

NHDPlusV2.1

The NHDPlusV2.1 data consist of multiple files containing spatial and tabular information, organized by 21 production units within the conterminous United States. Data are included for some regions of Canada and Mexico where the source of a United States stream is outside the United States, or the United States stream flows across a foreign border and returns to the United States downstream.

Three primary spatial hydrologic features are included in the NHDPlusV2.1 data: flowlines, catchments, and waterbodies (described below). A fourth spatial feature, NHDArea, has many characteristics similar to waterbodies; for purposes of this report, NHDAreas were joined with waterbodies. Associated with these spatial features are tabular files representing numeric and textual attributes of each feature. Counts of the three primary features, and subcategories of these features, are given in table 1 (McKay and others, 2012).

Flowlines are linear spatial features representing the paths of flowing water. For most flowlines, the identified path is an actual stream; however, some flowlines represent artificial pathways through a waterbody (further described in this section). Flowlines possess either routed or non-routed navigational properties. Routed flowlines have navigational information that determines the connectivity of each flowline to all the other flowlines within the routed network. Non-routed flowlines, in contrast, have no navigational information (no connection) and without this connection cannot be used in hydrologic models to route flow.

Catchments are polygons that represent the incremental drainage area contributing runoff directly to the flowline it contains. An exception would be in coastal regions where the catchment represents the total area draining each individual coastline segment. Catchments can also represent off-network sink features that capture the total drainage area to a sink because there are no upstream features (McKay and others, 2012). Catchments are derived using a complex process using a flow direction grid derived from 30-meter elevation data (McKay and others, 2012). Some short-length flowlines (or “connector” flowlines) do not have an assigned catchment, mainly because the resolution of the elevation data used to create the catchment was too small to capture an accurate or relevant drainage area (McKay and others, 2012). For these flowlines, the assigned catchment area is zero, no physical catchment exists, and the flowline serves the purpose of a connector. Catchments can, however, be defined without reference to a routed flowline effectively, as residual areas remaining after catchments for routed flowlines have been delineated. Such catchments are referred to as sinks.

Bodies of water (waterbodies) such as lakes, ponds, reservoirs, and wide rivers also are represented as polygons; some are associated with flowlines traversing through the center of the polygon. Most waterbodies associated with flowlines are lakes; however, waterbodies classified as flowing streams also are common and can be interpreted as polygons representing the left and right banks of large rivers. Waterbodies classified as streams typically relate to wide rivers for which the banks of the river have been digitized as polygons. Waterbodies that straddle two NHDPlusV2 production regions in some cases are split into separate polygons along the regional boundaries.

Most routed flowlines are streams with catchments, although many represent artificial pathways through a waterbody. Flowlines routed along coastlines have navigational information and typically have associated catchments. Although coastal flowlines do not possess estimates of stream-flow, the flowline simply represents the shoreline, a destination for runoff collected in the containing catchment. Most non-routed flowlines represent canals or ditches for which flow direction was not determined.

Flowlines and waterbodies are uniquely referenced in NHDPlusV2.1 by an identifying code (ComID). The referencing variable for catchments is called the FeatureID. However, a FeatureID's value also is the same value as the ComID of the flowline it contains. Catchments that are sinks have no

4 An Enhanced Hydrologic Stream Network Based on the NHDPlus Medium Resolution DataSet

Table 1. Primary spatial features of the NHDPlusV2 network, divided into sub-categories, with number of occurrences of each.

[*, There is a discrepancy of one flowline that has a catchment, but is not counted as a separate catchment. This is due to an erroneous duplication of a flowline in two regions.]

	Number of occurrences		Number of occurrences
Routed flowlines	2,691,344	Waterbodies	465,249
With catchments	2,638,935	Attached to routed flowlines	132,900
With headwater area	940,413	Type:	
In waterbody	359,660	Canal	259
Type:		Ice	46
Artificial path	378,207	Lake	122,594
Canal/ditch	25,930	Playa	170
Connector	25,316	Reservoir	396
Coastline	23,590	Stream	6,510
Pipe	183	Swamp	2,481
Stream	2,238,112	Unspecified	444
Unspecified	6	Unattached to routed flowlines	332,349
		Type:	
Non-routed flowlines	302,180	Canal	254
Type:		Ice	1,156
Artificial path	6,959	Lake	251,271
Canal/ditch	177,815	Playa	1,512
Connector	746	Reservoir	4,826
Coastline	3,102	Stream	429
Pipe	2,239	Swamp	63,835
Stream	111,317	Unspecified	9,066
Unspecified	2		
Catchments*	2,647,454		
Sinks	8,520		

contained flowline. For these features, the FeatureID is set to a negative value in the range of -600,000 to -6,000. The ComIDs for flowlines and waterbodies are distinct, even in cases where a flowline is contained within a waterbody; however, in cases where a waterbody is split along regional boundaries, all the separate parts will have the same ComID. Flowlines have an additional referencing scheme called a Reachcode. Reachcodes have the advantage over ComIDs in that they are relatively stable across different versions of NHDPlus (the first eight digits of a Reachcode refer to hydrologic regions); however, a complication of using Reachcodes as a unique identifier is that multiple flowlines can have the same Reachcode, making associations of ancillary data to specific catchments and flowlines using a one-to-one relation difficult.

Flowlines and waterbodies also are referenced by their Geographic Names Information System (GNIS) (USGS, 2017) name. These names are not unique, nor are they available for all features. GNIS names have been used to identify the primary pathway for flowlines with braided channels.

Flowlines and waterbodies have an additional attribute coded that describes the flowline or waterbody physical type. There are many different “type” codes for these features, so they can be collapsed into broader subcategories as described in table 1. The coastline type identifies flowlines that represent flow routing for coastlines in bays, estuaries, the ocean, and the Great Lakes. These flowlines have navigational information and catchments, even though they do not pass flow along the shoreline. A small fraction of the catchments is classified as sinks, meaning they have no flowline contained in them. These are closed basins which contribute no flow to the navigable network. Waterbodies are mostly lakes, reservoirs, and swamps.

The NHDPlusV2.1 data include a set of attributes useful for navigating upstream and downstream in the network. Network topology is defined by flowlines connected by nodes, which are the uniquely integer-numbered endpoints of flowlines that represent the points where flowlines originate and connect. The upstream end of the flowline is termed the

“From-node,” and the downstream end is the “To-node.” All flowlines immediately upstream from a given flowline will have a To-node equal to the given flowline’s From-node. It is useful to consider nodes as distinct features related to, but separate from, flowlines.

The StartFlag and TermFlag network variables identify the most upstream and downstream flowlines in the network. A headwater flowline, which has a StartFlag of one, has a From-node value in which there is no other flowline in the network with a matching To-node value. Similarly, a terminal flowline, having TermFlag of one, has a To-node value that is unmatched by any other flowline’s From-node value.

The HydroSeq variable is a unique sequence number that places each flowline in hydrologic order. It is a compact numbering system based on the node topology numbering system. At any flowline location, all upstream flowlines will have a higher hydrologic sequence number, and all downstream flowlines will have lower hydrologic sequence number. The final HydroSeq values are smallest at the downstream end of the network and largest at the upstream end of the network.

Diversions, locations in the stream network where a portion or all of the flow is diverted from the river’s natural course, are a common feature of an actual stream system and may have many causes and associated classes of features (table 2). The NHDPlusV2.1 network contains more than 64,000 diversions.

Diversions in NHDPlusV2.1 are defined by locations where more than one flowline has the same node as its From-node. Many of these divergences are local with the diverging paths reconnecting with one or two flowlines downstream. The

NHDPlusV2 divergence code (Divergence) identifies which of the multiple downstream pathways is the primary pathway; a divergence code value of 0 indicates the flowline is the only flowline emanating from the upstream node, implying the flowline is on a primary path by default. A value of 1 indicates the flowline is the primary path emanating from diversion at the upstream node. A diversion code of 2 indicates the flowline emanates from a diversion but is on a secondary pathway. Generally, the primary pathway was assigned according to GNIS stream name or by determining the shortest pathway to the point where the diverged paths reconnect. Only one flowline at a diversion is coded as the primary pathway.

The fraction of flow exiting a From-node that is directed down a given flowline is estimated and termed the “diversion fraction.” The sum of diversion fractions across the diverted pathways equals one to account for all flow passing through the diversion. NHDPlusV2.1 estimated diversion fractions (DivFrac) for a handful of diversions; diversion fractions were primarily derived from streamflow monitoring station (streamgage) data. However, most of the diversion fractions were determined simply by setting the diversion fraction equal to 1 for the primary pathway and to 0 for all other pathways (Mckay and others, 2012). If a diversion did not exist, then a diversion fraction of 1 is accurate; however, given inaccuracies discovered in the primary pathway designation and the recognition that some flow will likely pass downstream along a secondary pathway, the potential for errors in apportioned flow through the correct pathway can be significant.

Although some diversions involve secondary pathways that are isolated and appear to terminate in the ground

Table 2. Classes of diversions in NHDPlusV2.1 and their defining characteristics.

Type of diversion	Defining characteristic for routing
Braided channels/islands	Can be important for flow uncertainty depending on monitoring station location; generally affects local routing only.
Natural diversions to ground	Generally not important in affecting flow on main pathways and can be ignored.
Diversions in drainage ditches	Not generally important unless main pathways become incorporated into field drainage pathways; difficult to assess because channels are not naturally formed and flow is intermittent.
Multiple outlets at lakes/reservoirs	Can be important for flow accounting; operational control in reservoirs makes it difficult to assess mean routing of flow.
Irrigation diversions to ground	Important in aggregate in the West; difficult to assess because the channels are not naturally formed and flow is seasonal.
Cross-basin transfers	Important for flow accounting; many have gaging stations assigned.
Diversions in coastal areas	Common but tidal effects make them difficult to assess in non-tidal areas.
Divides	An uncommon feature where NHDPlus shows two flowlines connected across a divide, but the flowlines actually flow in opposite directions, tend to occur in headwater areas of low relief, and are typically associated with a marsh.
Diversions at stream confluences	Common in low relief areas where small tributaries connect with larger streams; affects local routing only; frequent cause of incorrect flow direction and primary path routing.
“Oxbow” diversions	Common in large rivers in flat relief; difficult to assess fraction diverted as width of feature a poor indicator of flow; typically occur in wooded areas making remote visual assessment difficult.

or have no outflow, most diversions reconnect to the main pathway downstream. The simplest such case is a diversion that forms two downstream pathways, only to reconnect at the next downstream node (fig. 1). The two diverted flowlines described here form what is called a diversion group. In this case, although flow in each of the flowlines of the diversion group is sensitive to the apportioning of flow at the diversion, the direction of flow in each flowline and the magnitude of flow leaving the group is unaffected by the diversion.

A more complicated topological situation occurs when an additional tributary merges with the diverted flowlines. Although the direction of flow at the inflows and outflows to the diversion group is known, the locations of the additional diversions and the direction of flow in all the flowlines composing the larger group are ambiguous. For a group with a single outflow, the direction of flow for the group of flowlines

that connect with the outflow must be the same as the outflow. However, the diversion can logically be placed at either inflow, and the flowline that connects the two inflows can flow in either direction. Moreover, the ambiguity associated with these topological groupings implies NHDPlusV2.1 cannot be certain exactly which inflow node of a group is the site of a diversion. The NHDPlusV2.1 algorithm used to define the primary path determined the direction of flow for the group as a whole, but as with the assignment of primary paths, the algorithm also can assign flow to travel in the wrong direction. The diversion-induced complications of flow apportioning and, with multiple inflows to the diversion group, flow direction are referred to collectively as a routing problem.

Diversions where one of the diverted pathways terminates in the ground, implying the diverted paths are not resolved downstream, are not likely to cause major problems

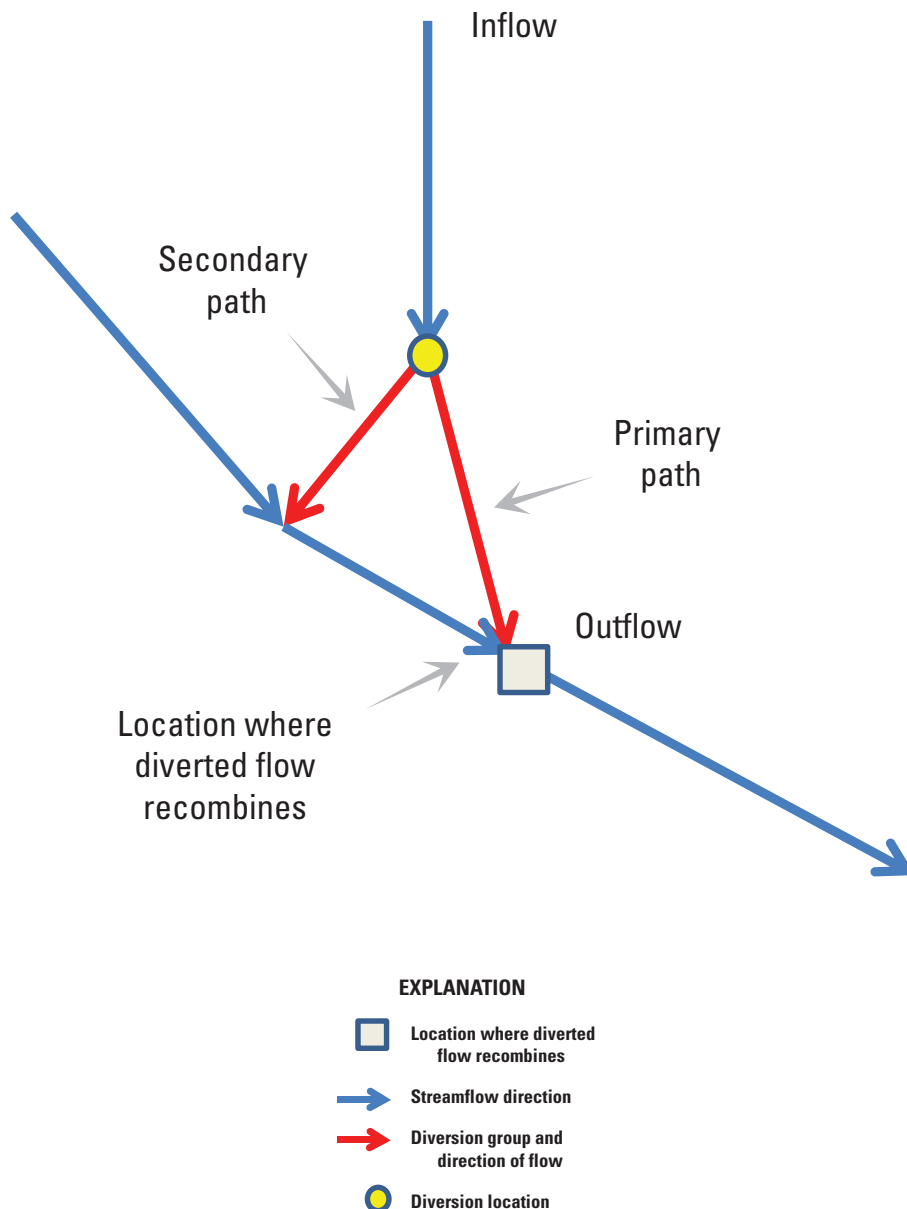


Figure 1. Example of how routing of NHDPlusV2.1 flowlines can be affected by an upstream diversion.

for modeling. This is because the primary path is invariably the pathway that does not terminate, and the pathway that does terminate cannot be too large or a large closed waterbody would have formed. However, the latter feature can occur in areas where no flow escapes. If the primary pathway delivers flow to a coast, the flow that is lost to termination in the ground is likely to be small and can be ignored.

In contrast, diversions that create groups with multiple inflows can raise significant issues for modeling. Depending on the assumed flow direction and apportioning at diversions, the implied flow in any given flowline of these groups can vary greatly. If a streamgaging station is located anywhere along this group of flowlines, the assumptions incorporated into the network can affect the modeling of the flow or contaminant load at the station, resulting in large model residual errors. Accumulations also are sensitive to routing issues, making it difficult to determine the accuracy of stream conditions derived from a processing of upstream characteristics.

NHDPlusV2.1 AR events identify flowlines where water is withdrawn or discharged and include an estimate of flow quantities for the addition (discharge) or removal (diversion). These quantities are used to modify streamflow estimated by routed runoff, or what might be called natural streamflow, thus converting natural streamflow into actual streamflow—the streamflow that is measured at streamgages.

Disconnected and routed transfers are two types of AR events distinguished here. Disconnected withdrawal and discharge events, when paired, represent the NHDPlusV2.1 flowlines that bracket a nonconservative use of water that may also alter the water's quality. Routed transfers are assumed to conserve total flow and are not presumed to alter water quality.

The disconnected withdrawal/discharge events were implemented in NHDPlusV2.1 by including the attribution of such events in the network attribute file—without alteration of the network. The events are further characterized according to their functional purpose, such as supporting public supply or irrigation, and whether the withdrawal or discharge could be associated with a complement discharge or withdrawal AR event. These designations can be useful in modeling to account for changes in water quality along the network and determining whether a withdrawal is a net withdrawal or a total withdrawal. In contrast, routed transfer events involve the creation of additional flowlines to support the routing of flow between two NHDPlusV2.1 flowlines that otherwise would not be connected.

The NHDPlusV2 variable *StreamLevel* generates an ordering of network streams in the inverse, upstream direction. The determination of a flowline's *StreamLevel* requires the definition of two additional variables: cumulative length and principal path. Cumulative length (also referred to as arbolate sum; McKay and others, 2012) is the sum of lengths of all upstream flowline segments, inclusive of the flowline's incremental length. To simplify the concept at diversions, only the primary pathway flowline accumulates area from upstream segments; all secondary flowlines (diversions) have

cumulative length reinitialized at the flowline's incremental length, with no accumulation from upstream segments.

The principal path represents a priority pathway which is distinct from a major path. It is identified based on both upstream and downstream relations. The designation of principal path is assigned to the single flowline, among all conflating flowlines having the same To-node, that either possess the same GNIS stream name as that of the downstream primary path flowline or, if the stream names do not match, the flowline that has the longest cumulative length. If multiple flowlines have a matching GNIS name, then the flowline among these that has the longest cumulative length is designated as being on the principal path.

The NHDPlus data include physical properties of the spatial features, such as flowline length, catchment area, minimum elevation in a catchment, waterbody surface area, and waterbody elevation. Flowlines have additional attributes derived from the accumulation of physical properties: the accumulated length along the primary upstream pathway and the accumulated upstream drainage area (inclusive of the catchment area associated with the flowline). Headwater flowlines have special properties to describe the portion of the catchment that drains directly to the From-node: the area of that headwater node portion and the minimum elevation within that portion. The headwater node area is useful for interpolating area at an arbitrary point along a headwater stream segment. The "minimum area" in the headwater-node portion, referred to in NHDPlusV2 as the maximum raw elevation, is an estimate of the maximum elevation of the delineated headwater stream channel.

There are two concepts of upstream drainage area: diversion-apportioned and total upstream. Diversion-apportioned upstream area is determined by apportioning accumulated drainage area at a node to a downstream flowline, according to the flowline's diversion fraction. This approach is akin to the way streamflow would accumulate. In contrast, total upstream area represents the total area of all catchments upstream from the given flowline, inclusive of the flowline's catchment area. Owing to diversions, this area concept is much more difficult to compute; care is required to avoid the double counting of an area where diverted pathways reconnect.

In addition to navigational properties, many useful watershed (catchment) and stream (flowline) characteristics represented by ancillary geospatial data are referenced to the NHDPlusV2.1 network either by index, or by intersection or overlay with catchment polygons, or by extensive processing of network data. Monitoring locations, monthly and annual mean precipitation, temperature, and runoff are examples of indexing or spatial intersections and overlays, and channel elevation, slope, and monthly and annual mean estimates of flow and stream velocity are representative of extensive processing (McKay and others, 2012).

Constructing a Modified Network

The enhancement and transformation of NHDPlusV2.1 (McKay and others, 2012) into the E2NHDPlusV2_us (Schwarz, 2019) network was undertaken in a series of intermediate steps. A national-level dataset was first created from the regional production units that compose the NHDPlusV2.1 data tables (McKay and others, 2012). Multiple data files from the NHDPlusV2 repository arranged and distributed by 21 production units nationally (McKay and others, 2012) (fig. 2) were downloaded and decompressed into the native file structure. Each record in the newly created database refers to a unique ComID with flowline, catchment, waterbody observations, and VAAs all joined into a single national file. This preliminary national network was then processed to fix discovered errors in routing and to fill in missing values. No cartographic alterations were made to the original NHDPlusV2 flowline line work. The result was the dataset ENHDplusV2_us (Schwarz and Wieczorek, 2018).

The next phase of enhancement used ENHDplusV2_us as a foundation and incorporated procedures that modified the network routing capabilities, including the enhancement of addition/removal events and the assignment of more precise diversion fractions to flowlines, described in the “Routing Modifications” section of this report. New variables were created that describe the extent to which streamflow can be affected by drainage from outside the United States borders (foreign) and by upstream diversions, and maintain an ability to distinguish between flowline, catchment, and waterbody features (table 3). Selected variable names and content also were modified to conform with the revised routing structure and to provide more descriptive names. Node identification numbers used to define the topological relation between flowlines were recomputed into a compact numerical set of numbers in order to expedite and facilitate procedures used to navigate the network. Variables not listed as being renamed (in table 3) were retained with the original NHDPlusV2 names. These enhancements also included the generation of channel slope and flow using modified methods from the original NHDPlusV2.1 methodology. Although not strictly part of the process required to generate E2NHDplusV2_us, a revised hydrologic conditioning of an extensive set of ancillary variables useful in empirical modeling and other hydrologic analyses were additionally generated (Wieczorek and others, 2018, Wieczorek and others, 2019) that can be used in conjunction with this network.

Filling In Missing Values

The SPARROW model infrastructure requires that every observation in the network used to simulate the routing of water, and (or) the parameters (variables) used for the assessment, possess non-missing values (Schwarz and others, 2006). However, not every observation in NHDPlusV2.1 contained a value (inherent or additional) for each variable, depending on the origin of the variable and (or) the method of association.

Some variables for certain observations were missing values simply because the values were not known. For example, not every stream or waterbody had an associated name. For some observations, flowlines with missing values did not have the proper networking connections to logically pass information up or downstream, preventing the ability to make educated assumptions of what the missing value might be. Additionally, the process used to populate original NHDPlusV2.1 flowlines with ancillary attribute data like precipitation intersected data values represented by a geospatial data layer with the catchment polygons. However, as noted earlier, some flowlines did not have an associated catchment because they were too short in length to be estimated. For these flowlines, a missing precipitation value would have resulted because a value could not have been computed from the ancillary spatial dataset. Lastly, flowlines may have been assigned missing value placeholders because the catchment (or part of it) lies outside the spatial extent of the ancillary spatial dataset being referenced or assigned. Depending on the nature of the variable, the lack of information inherent from the original NHDPlusV2 or inadequate ancillary information associated with the network could affect certain modeling applications.

Missing-value placeholders used in NHDPlusV2.1 signify that the attribute value was missing or undetermined (-9998) or that there was no applicable value was assigned (-9999). In cases where replacing a missing-value placeholder would be appropriate, a simple substitution was used to assign values. If the reason for a missing-value code was specific, a condition was deduced and applied for assigning a new value (Schwarz and Wieczorek, 2018). For example, the diversion fraction variable (DivFrac) was unknown for almost all flowlines; however, a simple condition for assigning a new value was to set DivFrac equal to 1 if the divergence code pertained to a primary path (the Divergence variable takes a value of 0 or 1) and to DivFrac equal to 0 for secondary paths (Divergence code of 2).

For most variables, the fill-in process was based on non-missing values from neighboring flowlines. To apply the process, it was necessary to distinguish variables as two types: intensive variables and extensive variables. Intensive variables are values for which the metric does not scale with area. For example, temperature is measured in degrees, and precipitation is in decimeters (millimeter \times 100) per year; neither metric is scaled with area. Extensive variables, in contrast, have units that tend to scale with the area of the catchment that they describe. An example of an extensive variable would be incremental catchment flow, measured in cubic feet per second, which is designated by cfs.

There were two approaches to filling-in missing data values using neighboring non-missing values (Schwarz, 2019). The default method, called the interpolation fill-in method, passed non-missing values from flowlines immediately downstream or upstream from the flowline containing the missing data value. The default method first operated in the upstream direction. If the variable in question was an intensive variable, a catchment-area weighted average of the variable

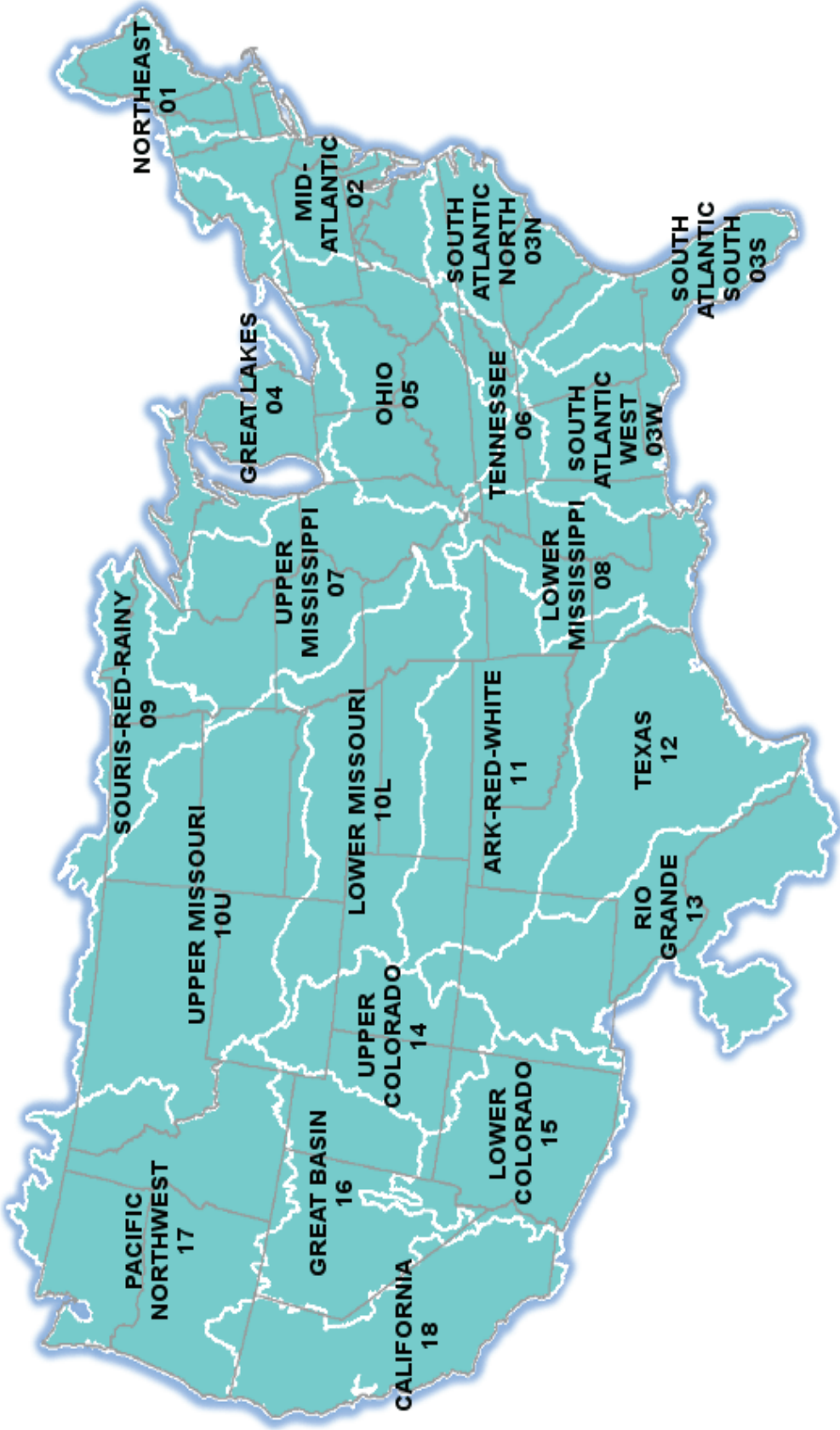


Figure 2. Map showing NHDPlusV2 production units.

Table 3. Revised names given to NHDPlusV2 variables.

[--, denotes name was not revised]

NHDPlus variable name	Revised name for flowline features	Revised name for catchment features	Revised name for waterbody/ nhdarea features
ComID	FL_ComID	--	WB_ComID
WBAreaComI	--	--	WB_ComID
GNIS_ID	FL_GNIS_ID	--	WB_GNIS_ID
GNIS_Name	FL_GNIS_Name	--	WB_GNIS_Name
Ftype	FL_FTypeAB	--	WB_FType
Fcode	FL_Fcode	--	WB_FCode
TerminalFl	TermFlag	--	--
ArbolateSu	TotLngthKm	--	--
AreaSqKm	IncAreaKm2	CatAreaKm2	WB_AreaKm2
DivDASqKm	CumAreaKm2	--	--
TotDASqKM	TotAreaKm2	--	--
MaxElevRaw	MaxElRawCm	--	--
MinElevRaw	MinElRawCm	--	--
MaxElevSmo	MaxElSmoCm	--	--
MinElevSmo	MinElSmoCm	--	--
HwNodeSqKm	HWNodeAKm2	--	--
FeatureID	--	FL_ComID	--
Q0001E	MAFlowUcfs	--	--
V0001E	MAVelUfps	--	--
Qincr0001E	IncFlwUcfs	--	--
Elevation	--	--	WB_ElevaFt

for all flowlines having the same From-node as the upstream flowline's downstream To-node was passed to the downstream flowline and became the fill-in value for that flowline. In computing the area-weighted average, if a flowline had a value but no catchment area, then the weight for that value was determined assuming the flowline had an area of 0.00001 square kilometers, designated by km2, (approximately one-hundredth the area of the smallest catchment with non-zero area in NHDPlusV2.1). For extensive variables, the sum of the values across the downstream flowlines was first normalized by the sum of the downstream flowlines' catchment areas. The normalized value was then passed upstream, where it was extrapolated by multiplying by the catchment area of the flowline with the missing value. If the missing value was in a flowline with no catchment, in which case catchment area was zero, the missing value was set to zero, regardless of the downstream value. If there were successive flowlines with missing values, the default process passed the initiating downstream flowline's non-missing value, or normalized value, to each of the successive flowlines, with subsequent expansion according to catchment area if the variable is extensive.

The default process was continued throughout the network to fill in as many missing values as possible. Thereafter,

if missing values remained, the default process was applied going in the downstream direction, passing to the downstream flowline possessing a missing value the catchment-area weighted average or the normalized sum determined for all flowlines having the same To-node as the downstream flowline's From-node. Performing the default process in the downstream direction could be effective in cases where the missing value was caused by alignment issues, most likely in coastal areas where the variable's spatial coverage did not fully extend over the catchment coverage.

If missing values remained after application of the interpolation fill-in process, then the method switched to obtaining neighboring values from flowlines using an extrapolation method in the same spatial region as the missing-value flowline. In the case of intensive variables, the fill-in value was the area-weighted average of all non-missing values in the region; in the case of the extensive variables, the fill-in value was the product of the missing-value flowline's area and the sum of the non-missing values in the same region, normalized by the sum of the catchment areas for the associated non-missing value catchments. The method initially attempted to fill in values using the region defined by the 8-digit hydrologic cataloging unit (HUC), which was defined for each flowline

Table 4. Summary of modification method codes for NHDPlusV2 variables.

[Method code of 1 = simple substitution rule; 0 = flowline upstream or downstream fill-in; 8 = fill-in via area-weighted average of 8-digit cataloging unit]

Variable	Method code	Count
CumAreaKm2	1	9,243
TotAreaKm2	1	9,243
CatAreaKm2	1	52,410
HWNodeAKm2	1	1,750,931
Divfrac	1	2,691,310
Slope	1	47,812
IfSink	1	684,080
MAVelUfps	1	23,574
WB	1	188,446
StartFlag	1	3,197
TermFlag	1	3,197
StreamOrde	1	1
StreamLeve	1	1
LevelPathI	1	1
MinElRawCm	0	73,611
MinElRawCm	8	1,963
MaxElSmoCm	0	3,412
MaxElSmoCm	8	757
MinElSmoCm	0	3,412
MinElSmoCm	8	757
PrecipV	0	50,963
PrecipV	8	1,447
TempV	0	50,963
TempV	8	1,447
RunOffV	0	50,963
RunOffV	8	1,447

by the first eight characters of the assigned ReachCode. If the 8-digit region contained only missing values, then the method switched to the 6-digit HUC region. Sinks missing a ReachCode assignment necessitated the determination of these flowlines by use of geographic information system (GIS) software to intersect catchments with the 8-digit HUCs (Steves and Nebert, 1994) and setting the first eight characters of Reachcode to the HUC that includes the largest share of the catchment (table 4).

Routing Modifications

It was not feasible to assess all possible routing assumptions (corrections, substitutions, and modifications) or errors

that may occur that could affect proper routing in the assigned primary path and (or) in the assumed direction of flow. A priority was established to focus on flowlines identified in NHDPlusV2.1 (McKay and others, 2012) that contained streamgages where the diversion-apportioned drainage area of the flowline differed greatly from the streamgage drainage area, as reported in the USGS National Water Information System. Additionally, the correlation between drainage areas of all flowlines affected by an upstream diversion (diversion group) with flowlines having a streamgage (Stewart and others, 2006) helped to identify other diversion groups warranting investigation for routing accuracy, regardless of the drainage area inconsistencies identified. Flowlines of lesser priority for investigation were those that were sensitive to diversion assumptions defined by the primary path identifier that carried a significant amount of flow through a complex set of alternative flowpaths. Figure 3 depicts a case in Michigan where routing assumptions affected the assumed flow. The original NHDPlusV2 routed the main source of flow from the Farmers Creek area (A) south and entering Lake Lapeer (B). The problem was a diversion assumed to be at a point along Farmers Creek (A), identifying a minor tributary as the only source of flow to streamgage 04146000. To rectify the routing for this example, the direction of flow was reversed (switched From-node and To-node values of flowlines) between locations (A) and (B). With this change, flow was routed north through Farmers Creek (A) and through the flowline, which included the streamgage.

Edits to the network, specifically edits to the flow direction, resulted in inconsistencies in the hydrologic sequence number which need to be recomputed to maintain hydrologic order. Although changes in hydrologic order may not have been necessary in simple edits, for situations involving the change of direction of multiple flowlines, a change in hydrologic ordering was always required.

A multi-step manual process was used to validate flowline routing and diversion fractions. The process began by identifying the latitude and longitude of a target diversion and using this location to acquire remote imagery through Google Earth™. The map imagery was enhanced with an overlay of USGS topographic maps, which provided feature names and topography helpful in assessing flow direction. In cases where the imagery allowed the identification of stream ripples, flow direction was verified visually. An attempt was made to apportion flow at diversions. Stewart and others (2006) include estimates of mean flow and the period for which that mean is determined. Upon inspection, if all paths emanating from a target diversion had a streamgage, and if mean flow was available for a period of no less than 2,000 days, then the diversion fraction for the diverting flowlines was based on these flows.

Streamgages were rarely located on each of the diverted paths; therefore, visual information from Google Earth™ was then used to obtain a more comprehensive assessment of flow apportionment. By comparing channel width at the narrowest points downstream from the diversions, it was then possible to obtain a rough estimate of the fraction of flow diverted to

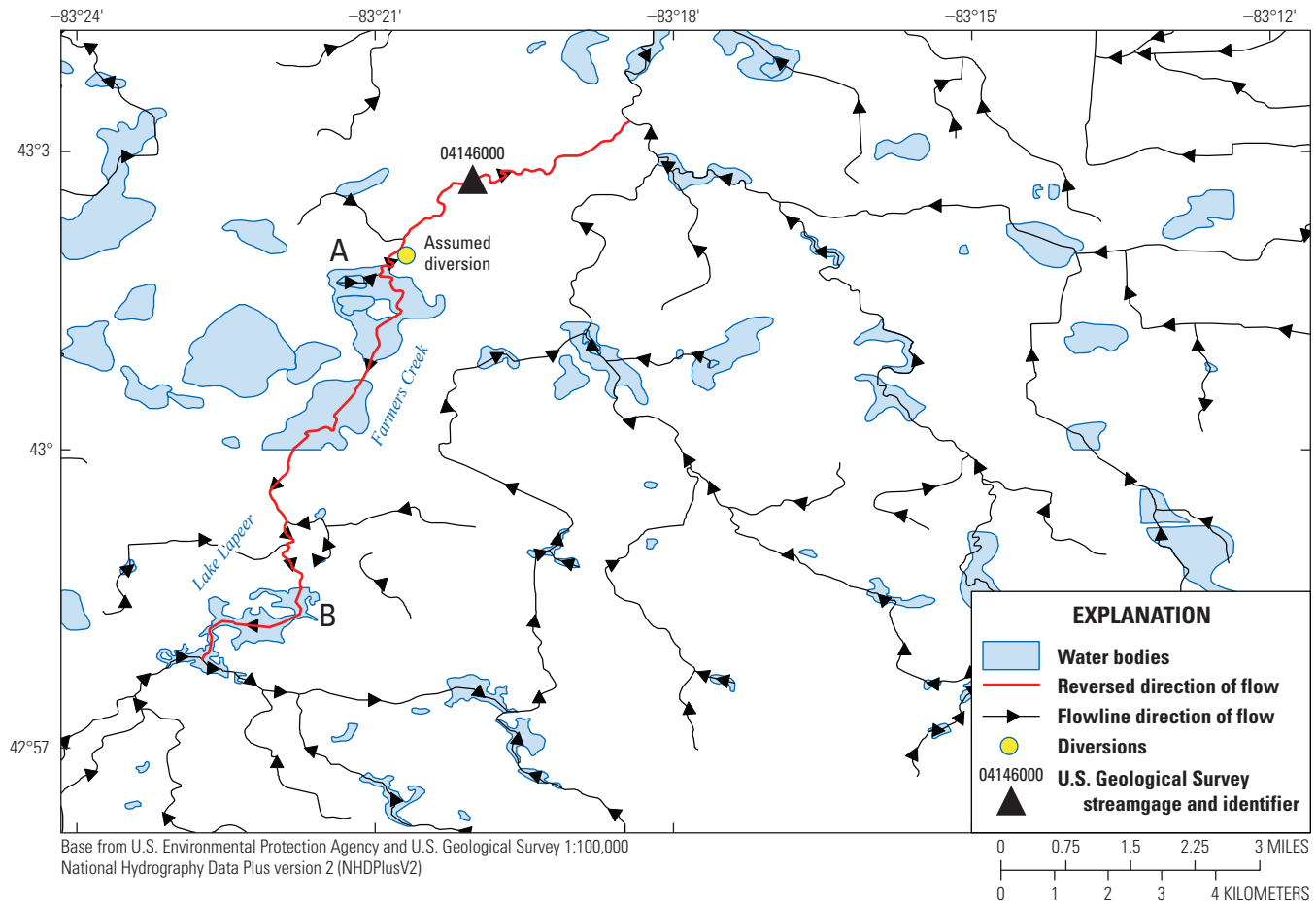


Figure 3. Example of routing errors associated with diversions affecting measured streamflow at a U.S. Geological Survey streamgage in Michigan. (A, Farmers Creek area; B, Lake Lapeer area)

each pathway, the fraction being computed by the ratio of the pathway width to the sum of pathway widths across all diverted pathways. Comparisons of diversion fractions based on this method with mean flows at streamgage locations showed reasonable correlation; however, no comparisons were performed on flowlines classified as canals. In some cases, it was concluded that the primary pathway needed to be altered, but it was not possible to assess the amount of flow allotted to each diversion pathway. In these cases, the explicit coding of the diversion fraction equal to one implies the primary pathway was confirmed and that all upstream flow was routed through the primary pathway.

The methods for estimating diversion fractions using imagery was sensitive to the flow conditions at the time the imagery was generated. To reduce uncertainty, historical imagery was scanned to ensure conditions were not overly dry or wet. Historical imagery also was useful for obtaining “leaf off” scenes to assess channel width at locations that were not visible in warmer months owing to the presence of thick forest canopy.

The results of diversion assessments were recorded with the following items: the ComID, modified value of the diversion code (0 for no diversion, 1 for primary pathway, and 2 for secondary pathway); the estimated diversion fraction for each pathway, a 0/1 code to identify flowlines requiring a reversal of flow direction; the revised value of the hydro-sequence number; and a comment (Schwarz, 2019). The comment column describes noted problems in the routing that may not be possible to fix, like an extensive modification of hydro-sequence numbering, or the existing line work in NHDPlusV2.1 that did not properly describe the actual channels. If streamgages were used to determine divergence fraction, the station identifiers were recorded in the comments column.

The number of diversion-related modifications applied by NHDPlus region and the totals for the Nation are listed in table 5. The values reported for modifications to diversion codes and fractions pertain to distinct diversions, which generally consist of multiple flowlines. The values reported for modified flow direction pertain to the number of individual flowlines with a reversed direction, multiples of which could be associated with a contiguous set of reversed flowlines.

Table 5. Diversion-related modifications to NHDPlusV2.

[Numbers for modified diversion code and fraction pertain to distinct diversions consisting of multiple flowlines; values for modified flow direction pertain to individual flowlines and do not include flowlines requiring a modification of the hydro-sequence attribute]

NHDPlus Region	Modified diversion code	Modified diversion fraction	Modified flow direction
01	41	37	23
02	14	41	2
03N	47	85	16
03S	20	43	1
03W	14	23	8
04	171	252	82
05	77	136	22
06	62	106	23
07	736	1,309	267
08	19	12	5
09	270	378	93
10L	43	57	2
10U	18	42	0
11	39	38	12
12	16	17	3
13	3	4	0
14	10	22	0
15	7	17	0
16	13	34	1
17	22	61	3
18	47	95	4
Total	1,689	2,809	567

Addition/Removal Events

Addition/Removal (AR) events are identified at locations along the E2NHDPlusV2_us network where flow is gained or lost due to anthropogenic changes to natural streams, specifically the addition or removal/diversion of water from the streams. These alterations can have multiple purposes: move water across basin divides to improve its economic or beneficial value; support specific economic activity, such as public supply or irrigation; or support navigable waterways like canals. Modifications to routing and (or) the attribution of network variables support the navigation through the network and account for the AR events; however, the AR events do not represent modifications of water quality nor quantity between the withdrawal and discharge locations.

AR events were incorporated by adding virtual flowlines to the network, along with appropriate modifications of the navigation variables (the From- and To-nodes, hydrologic

sequence numbers, and diversion fractions), to support the additional flowlines. AR routed transfers were assumed to attach to the network at the upstream node of the flowline to which they are referenced (fig. 4A). A simple routed transfer consists of at least two “pseudo” flowlines: a flowline that connects to the network at the withdrawal node and a flowline that connects to the network at the discharge node. This framework allows for additional pseudo flowlines to be included to account for diversions (fig. 4B), whereby a given withdrawal was transferred to multiple locations on the network. Pseudo flowlines were assigned COMIDs in a special range (less than -90000000) to distinguish them in the network from standard flowlines. Pseudo flowlines also were assigned a length of zero, the implication being that no instream losses could occur on the pseudo portion of the revised network, which would be consistent with the assumed properties of routed transfers. Multiple pseudo flowlines were permitted to merge either at a common node of the pseudo network or as discharge to a common node on the NHDPlusV2 network. Disconnected AR withdrawal/discharge events (including non-routed transfers) were assumed to connect to the specific flowline they reference.

Only one AR withdrawal/discharge event was assigned to a given flowline. If there were multiple withdrawals or discharges associated with the same flowline, they were incorporated only as a single combined event. This convention was not greatly restrictive if the multiple withdrawals/discharges were similar in nature, for example, summing the withdrawals from multiple public-supply facilities. The convention was more restrictive if the multiple withdrawals/discharges were dissimilar. In the latter case, it was necessary to spread the dissimilar withdrawals/discharges over multiple neighboring flowlines in the network, sacrificing the strict locational accuracy of the events.

Multiple sources of information were used to define and characterize AR events classes. The principal requirements for inclusion were that the event had a measure of mean flow withdrawn or discharged and that the event was indexed to NHDPlusV2.1. A principal source of flow information was streamgage records; many irrigation districts and transfers in the U.S. West have monitoring at the inflows and outflows to facilitate water-rights accounting (Idaho Department of Water Resources, <http://idwr.idaho.gov/water-rights/overview.html>). Additionally, public-supply facilities typically disseminate estimates of water withdrawals and discharges online as part of their efforts to inform the public of their activities.

Location information for AR events was available from numerous sources. A component of the regional files composing the NHDPlusV2 data consists of listings of AR events deemed important for flow accounting in the network; these data were directly incorporated into the determination of flow estimates provided with the original NHDPlusV2 data. The AR event classes included routed interbasin transfers, prominent withdrawals and discharges associated with public supply in the U.S. East and several events in the U.S. West associated with irrigation (McKay and others, 2012) (fig. 5).

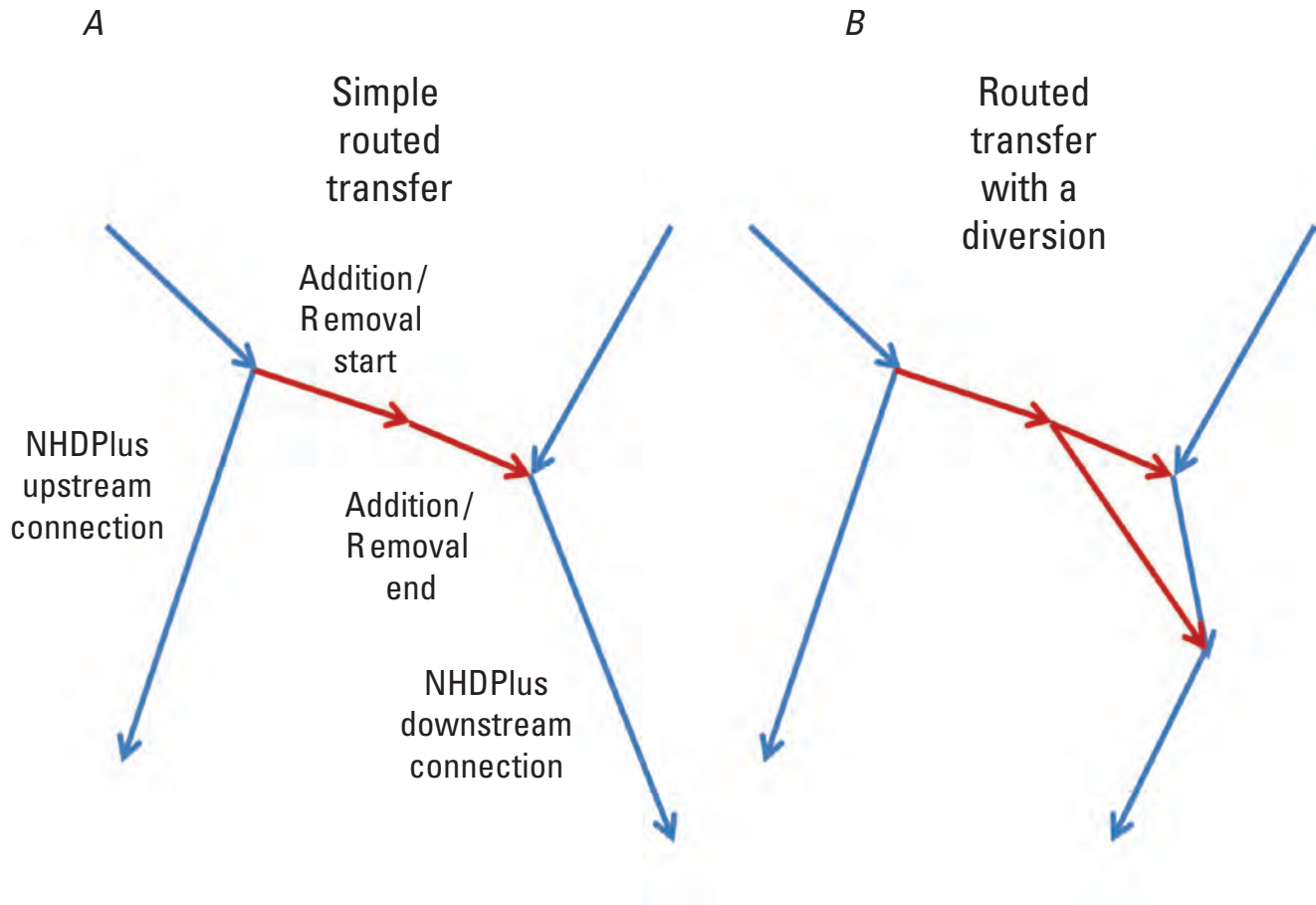


Figure 4. A, A simple routed transfer and B, a routed transfer with a diversion.

An additional source of AR events is the pair of USGS publications documenting major interbasin transfers in the conterminous United States, one for the West (Petsch, 1985) and one for the East (Mooty and Jeffcoat, 1986). These publications identify prominent routed transfers for various purposes, some of which duplicate those in the NHDPlusV2 AR event tables. The events from publications do not link directly to the NHDPlusV2 network and require additional locational research either by finding them visually in Google Earth™ and translating the location to NHDPlusV2 or by finding an undirected flowline in NHDPlusV2 that corresponds to the transfer. A useful feature of these events is that they include estimates of mean flow.

AR events modified in E2NHDPlusV2_us were specified using a spreadsheet (appendix 1; Schwarz, 2019), which was processed to incorporate the events into the enhanced network. The incorporation of AR events into the enhanced network led to the creation of pseudo flowlines and associated navigational information—nodes, diversion fractions, for example—to affect the routed transfer. Also created were numerous

additional variables describing the events, including the source region of the event; the quantity of flow associated with the withdrawal, transfer, or discharge event; the purpose or type of event in terms of public supply, irrigation, or transfer; a code that signifies the event as either a disconnected withdrawal, a disconnected discharge, or routed transfer; and the diversion fraction associated with a withdrawal or routed transfer. The incorporated routed-transfer AR events were tested, accumulating flow and area to insure the topological integrity of the modified network. Should the opportunity arise to incorporate revisions into the network to account for additional AR events or other routing modifications, previous AR routed-transfer events should not be reapplied; doing so would duplicate the virtual flowlines created to implement these transfers. If the existing routed transfers need modification such as for additional diversions or confluences among the existing virtual flowlines or to change the location of withdrawal or discharge, then it will be necessary to rebuild the network in its entirety, beginning with the compilation of the NHDPlusV2 data.

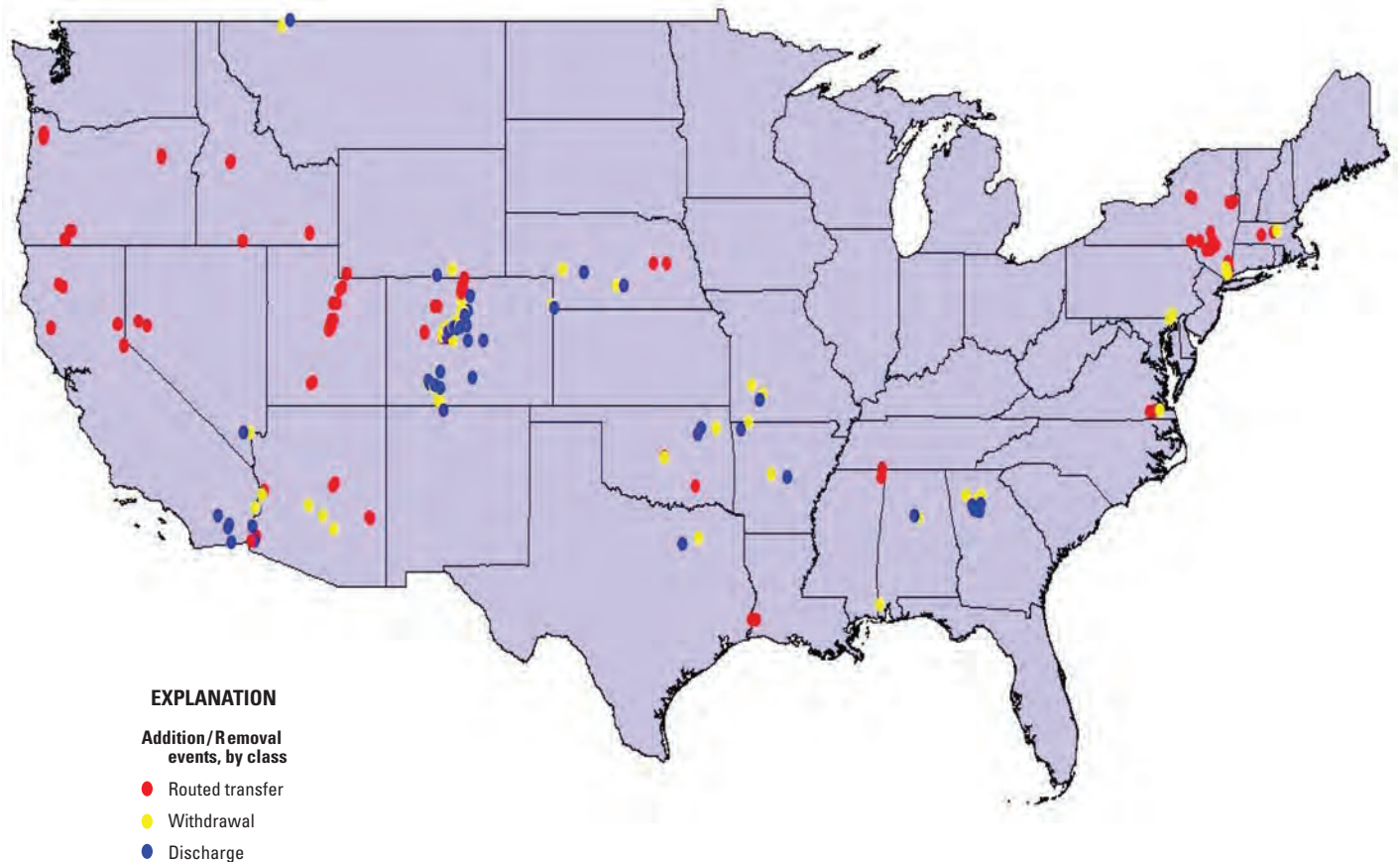


Figure 5. Location of Addition/Removal events included in the enhanced hydrologic stream network, by event class.

New Variables for Interpreting Network Flow Characteristics

Three groups of additional variables useful for interpreting results were derived from network applications related to flow characteristics. The first group (Flowline Termination) identifies flowlines terminating at foreign borders, shorelines, or in the ground. The second group of variables (Foreign Drainage) describes the area of contributing foreign drainage in each catchment and the cumulative upstream basin areas of each flowline. The last group of variables (Diversion-Affected Flowlines) identifies flowlines having basin aggregates that can be sensitive to the assumed flow apportionment made to upstream diversions.

Flowline Termination

The variable *StreamLevel* can help to identify and distinguish terminating flowlines at the coast from other terminations. However, *StreamLevel* did not differentiate between termination to ground, so-called internal termination, or termination at a foreign border, called external termination.

It was not possible to visually identify external terminating flowlines from the geospatial dataset because several apparent terminating flowlines at U.S. borders were topologically connected via common nodes to other flowlines that reenter the United States at different locations. However, the identification of exterior terminating flowlines was straightforward given the recognition that all such flowlines must be flowlines with To-nodes that have no matching From-node, and the To-node must be located near a linear U.S. external border, of which there were only 10—5 near Canada and 5 near Mexico. If the latitude and longitude of the To-node for a terminating flowline was within a 0.011-degree buffer of one of these linear foreign borders, the borders being described by simple linear equations of latitude and longitude coordinates, then the flowline was classified as having an external termination, and the variable *IfExtTerm* was set to a value of 1. A companion variable, *IfIntTerm*, identifies flowlines that have a termination in the interior of the network, and *IfIntTerm* was assigned similar to *IfExtTerm*, except that it takes a value of 1 if the non-coastline terminating flowline also is not an external terminating flowline.

Foreign Drainage

Three derivative attributes were created to describe the effect of foreign drainage on the conditions of a given flowline. They include (1) the amount of foreign area in the incremental catchment (IncUSArKm2), (2) the cumulative amount of foreign drainage at and upstream from the flowline (CumUSArKm2), and (3) the percent of upstream drainage that is foreign (PctForgnAr). This information is useful if the network is used with models that are based on input data that are only available domestically, in which case the characteristics for flowlines having substantial amounts of foreign drainage need to be closely inspected.

The incremental foreign drainage was determined by intersecting the NHDPlusV2.1 catchment layer, which includes foreign drainage, with the 2011 National Land Classification Dataset (NLCD; Homer and others, 2015), which is confined to U.S. borders. Areas within both the NHDPlusV2.1 incremental catchment and NLCD were defined as incremental domestic drainage. The cumulative upstream domestic drainage area was determined by accumulating the incremental domestic area downstream using the 0/1 diversion fraction implied by the primary path diversion classification (Divergence code) to apportion accumulated area to the diverting downstream flowlines. The variable Percent foreign area was determined by differencing the cumulative upstream domestic area from the diversion-apportioned cumulative drainage area (the CumAreaKm2 variable), normalizing by the cumulative drainage area and converting to percent.

Diversion-Affected Flowlines

Another group of new variables was created to assist in identifying flowlines affected by upstream diversions. These variables were useful for indicating which flowlines have greater uncertainty in cumulative properties owing to the potential error associated with assigning an upstream diversion fraction. This uncertainty also could be elevated to an extreme level if there are multiple upstream diversions that remain unresolved, perhaps as a result of a braiding pattern. These added variables provide information on the presence of a diversion effect and the level of uncertainty upstream diversions could induce in accumulative properties. This is critical information, given that the routing of approximately 2.5 percent of the flowlines in the dataset can be affected by diversions. As more information on diversion fractions is included in the network, the uncertainty caused by diversions should attenuate. Schwarz (2019) contains variable names and descriptions identifying diversion-affected flowlines.

Revised Variables

The routing changes induced by diversion and flow direction modifications, and the inclusion of AR routed-transfer events, necessitate revisions to several E2NHDPlusV2_us variables in order to maintain consistency with the modified routing.

Hydro-sequence

As part of the network revision, particularly the operations of reversing flow direction of selected flowlines and the addition of virtual flowlines to effect AR routed-transfer events, the hydro-sequence variable (HYDSEQ) used to order the network was revised. Because routing changes may affect hydrologic sequencing across multiple regions, new sequence numbers were computed for the entire network simultaneously, rather than regionally as was done for the original NHDPlusV2 network. The result was unique sequence numbers in hydrologic order for the entire dataset.

Streamflow

The regeneration of streamflow follows much of the Enhanced Unit Runoff Method (EROM) described in the NHDPlusV2 documentation (McKay and others, 2012) but was modified to correct some logical inconsistencies in the conditioning of flow from streamgage measurements (appendix 2). In the application of the EROM method to the revised network, the entire network was processed simultaneously, making it necessary to remove the boundary condition flow component from the incremental catchment flows at the boundary locations. The EROM method also added or subtracted flow quantities based on AR events. The revised network incorporated a number of AR events that were not included in the original NHDPlusV2, in addition to revised quantities of many events that were common to both datasets.

Slope

A method used to compute channel slope for each flowline is described in appendix 3. This method is similar to the one used to generate slope estimates for NHDPlusV2.1 (McKay and others, 2012). The revised method expanded the search for set points (the smallest elevation among all pathways converging at one node) beyond a given path to include all nodes upstream from the current set point. The revised method includes evaluating for potential diversions.

Additional Ancillary Information

The NHDPlusV2.1 flowlines and catchments, and subsequently the enhanced network, (E2NHDPlusV2_us) provide unique frameworks for additional spatially referenced attributes suited for relative modeling and assessment applications. By associating additional characteristics, models like SPARROW can be used to empirically assess statistical correlations between stream-monitoring and specific contaminant sources, watershed characteristics, and transport processes at regional and national scales (Brakebill and others, 2011; Robertson and Saad, 2011; Preston and others, 2009; Rowe and others, 2010).

Many additional ancillary datasets have been acquired or developed as separate geospatial data layers, either as points, polygons, or a cell-based rasters representing a wide variety of

landscape, subsurface, or hydrologic information and characteristics (Wieczorek and others, 2018; Wieczorek and others, 2019; Schwarz, 2019; McKay and others, 2012; Saad and others, 2018; Skinner and Wise, 2019). Values representing these characteristics were assigned to the catchments or the flowlines of NHDPlusV2.1 networks using standard GIS intersection and overlay techniques. Examples include streamflow and water-quality monitoring locations, municipal point sources, dam locations, land use, atmospheric deposition, natural features like lithology, and commercial fertilizer and manure applications. Examples of land surface and subsurface transport characteristics include soil properties, slope, precipitation, and temperature. Instream transport processes include streamflow and water-chemistry monitoring locations, estimates of mean-annual streamflow and water velocity, stream length, mean-annual water travel time, reservoir surface area, and stream-reach type. In addition to flowline networking features and some instream processes, ancillary streamflow estimates and some landscape and climate information native to the NHDPlusV2 suite of data were updated or used (McKay and others, 2012; Schwarz and Schwarz, 2019). These data were applied at various stages of SPARROW model development and exploration. Appendix 4 of this report contains a comprehensive list of variables developed for E2NHDPlusV2_us. Appendix 5 lists selected ancillary geospatial dataset variables assigned to the catchments of and flowlines of NHDPlusV2.1 (Wieczorek and others, 2019).

Automated methods of transference or association of these additional ancillary geospatial data to the network by either GIS intersection or overlay operation created challenges in obtaining a complete set of ancillary values for every flowline in the network, a requirement for the SPARROW infrastructure. In some cases, the spatial extent of many ancillary geospatial datasets represented only areas within the U.S. border, creating areas of missing information within some catchments. Differences in spatial resolutions between the network and ancillary data also created inconsistencies between the coastline boundaries of catchments and the coastline extent of the ancillary variable when overlaying or intersecting polygons or rasters. In cases where the coverage of ancillary information did not fully overlay the NHDPlus catchments, the transfer of values to the network was incomplete, creating areas of missing data. Missing values also arose when a flowline was too small to allow for the generation of a catchment, making it impossible to transfer area-based ancillary data. Some geospatial data were missing data qualifiers for open-water areas. Methods for estimating missing values are explained in the “Fill-in Missing Values” section of this report.

Aggregations

An aggregation of ancillary variables associated with the network allows for assessments, including total upstream characteristics and contributions, to occur at any location along the network. A secondary purpose of the automated procedures

was to navigate the network and perform aggregations of the ancillary variables. Aggregation always operated in a downstream direction, so that the aggregated values pertain to an aggregation of all flowlines upstream and inclusive of any given flowline.

Four types of aggregations can be performed on sets of associated ancillary data (Schwarz and Wieczorek, 2018; Wieczorek and others, 2019): accumulation, area-average, minimum, and maximum. Variables classified as accumulation variables can be aggregated by accumulating the values; variables classified as averages can be aggregated by simple area-average values. Variables for which an upstream minimum or maximum could be computed can be treated as average variables for purposes of the fill-in procedures described in section “Filling In Missing Data Values.”

Two methods were used to determine accumulation and area average. The default method used the NHDPlus diversion fractions (either the coded values or the default 0/1 values assigned according to the primary path at a diversion) to allocate a cumulation along diverted paths at a diversion. For accumulation variables, this is a simple apportion of the cumulated values at a diversion. For average variables, the incremental value of the variable would first be multiplied by incremental catchment area. This product would then be accumulated and apportioned at diversions. To recover an area-weighted average, the variable’s accumulated product would be divided by its diversion-apportioned accumulated incremental catchment area.

The second method of accumulation/area-averaging computed a total aggregation, representing an aggregation of all values of the variable upstream and inclusive of the flowline. This method did not apportion at diversions but was implemented in such a way that values were never double counted because they appear in a diverted path. As with the standard, diversion-apportioned aggregation, an area average would be undertaken by forming the product of the variable and its incremental area, performing separate total accumulations of the product and the incremental catchment area, and recovering the area average by dividing the total accumulated value of the product by the total accumulated catchment area.

The estimation of minimum and maximum aggregations was more straightforward. The algorithm descended the network while recording the maximum (and minimum) of the values stored in the flowline’s upstream node and the incremental flowline value. If the upstream node had no value, the flowline’s value would then become the default maximum (and minimum). The flowline’s maximum (and minimum) value would be passed to the downstream node only if the value was greater (or less) than the value recorded in the node. If the downstream node was not assigned a value, the flowline value of the maximum (and minimum) would then be automatically passed to the node.

Last, results were visually inspected and verified. The focus was on areas of the network with braided streams, as well as complex areas at various positions within the stream network to ensure accuracy.

Validation

A series of checks was performed to ensure the logical consistency of the modifications such as ensuring only one flowline at a diversion was coded as the primary pathway, diversion fractions across the pathways summed to one, and the flow connectivity of the network was not compromised by the altering flow direction (the process of flipping the From- and To-nodes) after reordering the network with the revised hydro-sequence numbers.

Changes and modifications to the routing network rendered some attribute values obsolete, particularly variables dependent on the assigned primary pathway, such as channel slopes, hydrologic sequencing, and variables pertaining to aggregations or accumulations of incremental (local) segment or catchment values.

Most modifications to the routing network were performed in region 07 specifically and the Midwest generally (table 5). Few modifications were applied in the more semi-arid regions of the United States. The numbers reported for modified diversion fraction include cases where there is confirmation of the original NHDPlusV2 assessment of the primary path. Most of the modifications reported in the table, particularly for regions 04, 05, 06, 07 and 09, were not predicated on an observed problem at streamgages. Thus, the ratio of the number of modified diversion codes to the number of modified diversion fraction, 60 percent, is an approximate estimate of the error rate for NHDPlusV2 diversion assignment. This ratio, estimated nationally at 61 percent, indicates that the algorithms used to assign diversion codes in the original NHDPlusV2 are incorrect and revisions to the method should be considered.

The modification process performed revisions on several NHDPlus variables so that they remain consistent with the revised routing. Changes in routing imply diversion-allotted cumulative variables, such as drainage area (CumAreaKm2) and streamflow (MAFlowUcfs or Q0001E from the EROM table), must be recomputed from incremental flowline values. Additionally, the computationally challenging cumulative variables expressed as upstream totals (TotAreaKm2 and TotLnghKm [Arbolatesu]) must be recomputed. Other NHDPlus attributes that potentially affected by network modification are StartFlag, TermFlag, TerminalPa, StreamCalc, StreamOrde, StreamLeve, and LevelPathI (See “Methods and Material” and appendix 4 for descriptions). These variables were regenerated to be in accordance with the revised network. The methods used to revise the derived NHDPlus variables have been evaluated to determine that the results reproduce the existing attributes for non-coastline

flowlines using the existing network topology (hydroseq, From-/To-nodes, divergence), excluding the alterations described in table 5 and a few isolated incidences. The results of this evaluation are reported in table 6. An investigation of each attribute discrepancy, excluding flow and velocity, showed the discrepancy to be the result of an error in the original NHDPlusV2 dataset.

Comparing reported drainage areas from USGS streamgages with accumulated areas at the same location has proven to be an effective way to evaluate and verify proper stream connectivity and routing. These steps were completed using the best available data and quality assurance measures throughout the process. Computed watershed areas (divergence routed and total upstream accumulation; fig. 6) were compared with those distributed with the NHDPlusV2 data suite (McKay and others, 2012). This comparison demonstrated that, by comparing the accumulated drainage areas with those reported in NHDPlusV2, catchment, divergence routed, and total upstream watershed-level metrics were calculated correctly. The outliers in the plots (fig. 6) could be explained by the modifications done to the original routing of the NHDPlus dataset where flow directions were reversed and diversions were edited. Overall, the results show an overwhelmingly positive correlation with R-squared value of more than 0.99.

Routing and accumulating upstream watershed area also were compared with streamgage drainage areas published in the GAGESII dataset, which used digitized watershed boundaries defining the drainage outlet as the point locations of selected streamgages, combined with ancillary geospatial data (Falcone, 2011; Falcone and others, 2010). GAGESII basin outlets can occur anywhere along the flowline, which can result in varying metrics, depending on the source data being compared. In some cases, NHDPlusV2_us flowlines are 95 kilometers in length. In addition, some source data can be widely variable within an NHDPlusV2 flowline catchment. More than 4,700 GAGESII basins were selected for comparison. By matching GAGESII outlet points with NHDPlusV2 COMIDs, only those GAGESII basins that had a basin area within 1 percent of the NHDPlusV2 total accumulated basin area were selected for comparison. Results demonstrate a significant positive Pearson’s R-squared value (fig. 7). Outliers from the comparisons likely can be attributed either (1) to the difference in the basin delineation for GAGESII compared to that of the network accumulation catchment representations in NHDPlusV2 or (2) to the fact that the NHDPlus network reports data from the downstream end of flowline, whereas GAGESII reports information at the exact locations of the streamgages.

Table 6. Number of discrepancy values from attempt to reconstruct original NHDPlus estimates.

[Results exclude coastline flowlines and do not include modifications of NHDPlus described in table 5]

Variable	Number of flowline discrepancies
StartFlag	0
TermFlag	0
TerminalPa	15
StreamCalc	77
StreamOrde	81
StreamLeve	200
LevelPathI	115
TotAreaKm2	1
CumAreaKm2	0
TotLnghKm	0
MAFlowUcfs	157,066
MAVelUfps	14,702

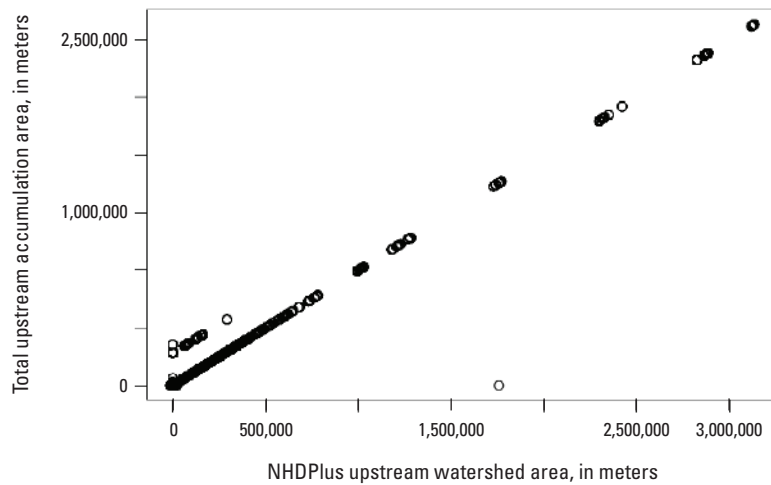


Figure 6. NHDPlus upstream watershed areas in relation to total upstream accumulation values.

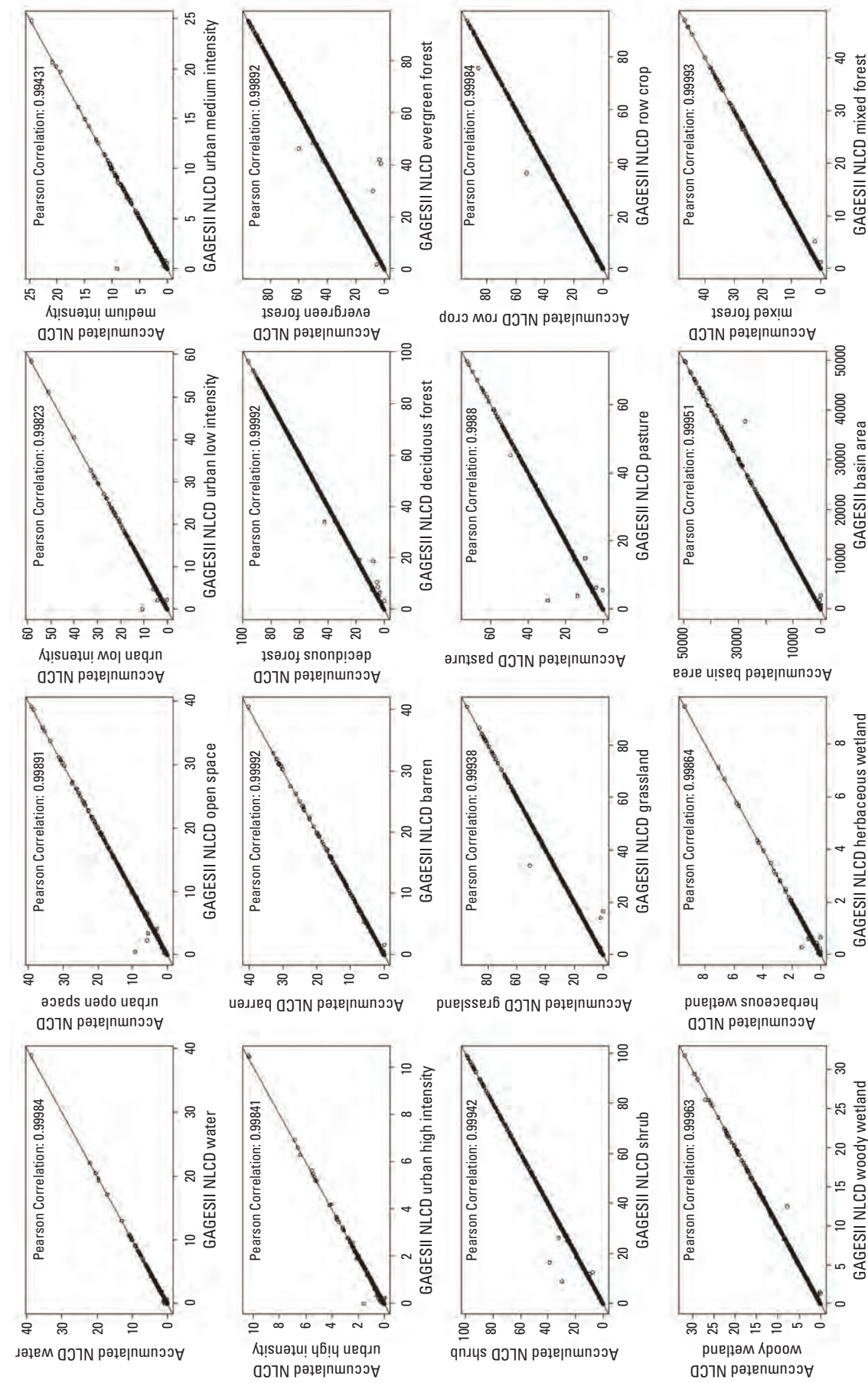


Figure 7. Pearson's Correlation for basin percent of GAGESII NLCD 2006 land-use classes in relation to values from the total upstream aggregation method.

Summary and Conclusions

Digital hydrologic networks provide common spatial units that can be used to frame the descriptions of aquatic and watershed processes and properties. These descriptions have been applied to hydrologic analysis and modeling. The National Hydrography Dataset Plus (NHDPlus), Version 2.1 (NHDPlusV2.1) represents an attribute-rich geographic information system (GIS) stream network developed at the 1:100,000 scale. The structure includes networking, feature naming, and “value-added attributes” of relevant characteristics that make the dataset useful for hydrologic modeling and interpretations. The U.S. Geological Survey National Water-Quality Assessment project adopted the medium-resolution NHDPlusV2 network as the primary framework supporting SPAtially Referenced Regression On Watershed (SPARROW) regional and national modeling. The modeling technique relies on accurate navigational attributes and structures to establish relations between associated quantities of monitored contaminant loads, contaminant sources, and physical characteristics affecting contaminant transport. Versions of the NHDPlus network have proven useful for assessing and applying stream and watershed characteristics for hydrologic applications at regional scales; limitations in NHDPlusV2 networking capabilities do exist. Modifications to the NHDPlusV2 routing and stream connections were performed with the intention of providing important networking improvements and enhancements utilized by SPARROW modeling and other hydrologic routing and accumulating applications.

The steps required to create and update the enhanced network (E2NHDPlusV2_us; Schwarz, 2019) were undertaken in a precise order, ensuring that correct values of routing attributes were updated before they could be used as input affecting other attributes. It was first necessary to compile the NHDPlusV2.1 files, compactify the nodes of the network, apply the fixes that correct known errors, and fill in missing values. The next step differentiated non-coastal terminal flowlines that terminate in the U.S. interior from those that terminate at a U.S. border. Following this, the Addition/Removal (AR) events were created—a step that inserts virtual flowlines into the network to account for routed transfers. The critical step in generating a new set of hydro-sequence variables to account for changes in flow direction and the inclusion of the AR virtual flowlines was then performed. This step also produced revised estimates of the diversion-fraction apportioned drainage area.

Once modification of the network was completed, the additional variables related to foreign drainage and diversion-affected flow were created and made consistent with the revised routing. The value-added attributes that identify flow-path relations were reevaluated, and then smoothed elevation, channel slope, and mean streamflow and velocity were generated using the revised algorithms. At each step, the entire network was processed concurrently, avoiding the complications

from making changes consistent across multiple production regions, which occurs if the regions are processed separately. The last step generated a copy of the enhanced national network as a set of regional files, one file for each production region, which was then merged with regional spatial data for incorporation into a GIS.

Given the existence of the enhanced network, there may arise the opportunity to incorporate further revisions into the network to account for additional AR events or routing modifications. In implementing these changes, it is important that previous AR routed-transfer events are not reapplied; doing so would duplicate the virtual flowlines created to implement these transfers. If the existing routed transfers need modification, for example, to account for additional diversions or confluences among the existing virtual flowlines or to change the location of withdrawal or discharge, then it will be necessary to rebuild the network in its entirety, beginning with the compilation of the NHDPlusV2 data. A reapplication of the other AR events, the withdrawal and discharge events, would not affect the integrity of the enhanced network. Similarly, care is required when performing modified routing to the enhanced network; a reapplication of diversion fraction and divergence code modifications is innocuous, but reapplying the changes in flow direction would negate the previous modifications.

Assuming the AR events and routing modifications pertain only to new instances of routed transfers and flow direction reversals, then the revision process could start with these steps. All the steps that follow AR and routing modification processing as part of the creation of the enhanced network must be repeated. The creation of new variables and the network revision process are sensitive to these changes and must be regenerated to restore the enhanced network to a valid state.

An approach was developed to efficiently and accurately associate and aggregate landscape information for all flowline catchments and their upstream areas in the E2NHDPlusV2_us dataset. Because the modifications to the routing involved a quality-control inspection of routing assumptions incorporated into the original NHDPlusV2.1 dataset, based on published flow data and local knowledge from hydrologists, a more accurate accounting of basin characteristics for every flowline, waterbody, coastal catchment, and sink from headwater streams to downstream units could be computed. The described methods produced accurately acquired summaries through braided streams without double counting of values of streams or their catchments.

The sources of AR events listed here are not exhaustive, and numerous additional events could be used to fully describe the artificial withdrawal and discharge of water to the stream network. Examples of sources include the U.S. Environmental Protection Agency (EPA) Permit Compliance System (EPA, 2018) database, which reports on outfall locations and flow discharge for thousands of permitted facilities, and numerous irrigation districts in the U.S. West that have non-directed drainage ditch flowlines in NHDPlusV2.1 and may have pertinent streamgaging information. Although not incorporated in the network, representations of aqueducts, canals, conduits,

pipes, irrigation ditches, and flowlines corresponding to flow paths that lack flow direction could be used to define and refine AR withdrawal and discharge events.

Results from an evaluation of the modified routing network database demonstrate a positive correlation with published datasets containing similar information, such as accumulated land use and drainage area. These and other associated ancillary data provide an important asset for hydrologists and other researchers to understand and characterize the Nation's rivers and streams. Updates, additions, and modifications to ancillary data assigned to the NHDPlus network likely will continue as needed to support regional and national modeling and assessment activities.

References Cited

- Alexander, R.B., Boyer, E., Smith, R.A., Schwarz, G.E., and Moore, R.B., 2007, The Role of Headwater Streams in Downstream Water Quality1: JAWRA Journal of the American Water Resources Association (JAWRA), v. 43, p. 41–59. DOI 10.1111/j.1752-1688.2007.00005.x. [Also available at <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1752-1688.2007.00005.x>.]
- Ator, S.W., Brakebill, J.W., and Blomquist, J.D., 2011, Sources, fate, and transport of nitrogen and phosphorus in the Chesapeake Bay watershed—An empirical model: U.S. Geological Survey Scientific Investigations Report 2011–5167, 27 p., accessed March 16, 2016, at <https://pubs.usgs.gov/sir/2011/5167/>.
- Brakebill, J.W., Wolock, D.M., and Terziotti, S.E., 2011, Digital hydrologic networks supporting applications related to Spatially Referenced Regression Modeling: Journal of the American Water Resources Association (JAWRA), v. 47, no. 5, p. 916–932, accessed April 20, 2015, at <https://doi.org/10.1111/j.1752-1688.2011.00578.x>.
- Falcone, J.A., 2011, Geospatial Attributes of Gages for Evaluating Streamflow (Gages II): U.S. Geological Survey, data release, accessed April 21, 2015, at https://water.usgs.gov/lookup/getspatial?gagesII_Sept2011.
- Falcone, J.A., Carlisle, D.M., Wolock, D.M., and Meador, M.R., 2010, GAGES—a flowline gage database for evaluating natural and altered flow conditions in the conterminous United States: Ecology, v. 91, no. 2, [variously paged], Data Paper in Ecological Archives E091-045-D1, accessed March 16, 2016, at <http://esapubs.org/Archive/ecol/E091/045/metadata.htm>.
- Hill, R.A., Weber, M.H., Leibowitz, S.G., Olsen, A.R., and Thornbrugh, D.J., 2016, The Stream-Catchment (Stream-Cat) Dataset—A database of watershed metrics for the conterminous United States: Journal of the American Water Resources Association, v. 52, no. 1, p. 120–128, DOI: 10.1111/1752-1688.12372. [Also available at <https://doi.org/10.1111/1752-1688.12372>].
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database (NLCD) for the conterminous United States—Representing a decade of land cover change information: Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p. 345–354, accessed November 6, 2017, at http://www.asprs.org/a/publications/pers/2015journals/PERS_May_2015/HTML/files/assets/basic-html/index.html#345/z#noFlash.
- Hoos, A.B., Moore, R.B., Garcia, A.M., Noe, G.B., Terziotti, S.E., Johnston, C.M., and Dennis, R.L., 2013, Simulating stream transport of nutrients in the eastern United States, 2002, using a spatially-referenced regression model and 1:100,000-scale hydrography: U.S. Geological Survey Scientific Investigations Report 2013–5102, 33 p.
- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., 2012, NHDPlus Version 2: User Guide: U.S. Environmental Protection Agency, 182 p., accessed May 1, 2018, at https://s3.amazonaws.com/nhdplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf.
- Moore, R.B., and Dewald, T.G., 2016, The road to NHDPlus—Advancements in digital stream networks and associated catchments: Journal of the American Water Resources Association (JAWRA), v. 52, iss. 4, p. 890–900, accessed March 12, 2017, at DOI: 10.1111/1752-1688.12389.
- Mooty, W.S., and Jeffcoat, H.H., 1986, Inventory of interbasin transfers of water in the eastern United States: U.S. Geological Survey Open-File Report 86–148, 38 p.
- National Oceanic and Atmospheric Administration (NOAA), 2016, The National Water Model, [not paged], accessed January 26, 2018, at <http://water.noaa.gov/about/nwm>.
- Petsch, H.E., 1985, Inventory of interbasin transfers of water in the western conterminous United States: U.S. Geological Survey Open-File Report 85–166, 34 p.
- Preston, S.D., Alexander, R.B., Woodside, M.D., and Hamilton, P.A., 2009, SPARROW MODELING—Enhancing understanding of the Nation's water quality: U.S. Geological Survey Fact Sheet 2009–3019, 6 p. [Also available at <http://pubs.usgs.gov/fs/2009/3019/>.]

- Robertson, D.M., and Saad, D.A., 2011, Nutrient inputs to the Laurentian Great Lakes by source and watershed estimated using SPARROW watershed models: *Journal of the American Water Resources Association*, v. 47, no. 5, p. 1011–1033, accessed April 15, 2017, at <https://doi.org/10.1111/j.1752-1688.2011.00574.x>.
- Rowe, G.L., Belitz, K., Essaid, H.I., Gilliom, R.J., Hamilton, P.A., Hoos, A.B., Lynch, D.D., Munn, M.D., and Wolock, D.W., 2010, Design of Cycle 3 of the National Water-Quality Assessment Program, 2013–2022: Part 1: Framework of water-quality issues and potential approaches: U.S. Geological Survey Open-File Report 2009–1296, 54 p. [Also available at <https://pubs.usgs.gov/of/2009/1296/>.]
- Saad, D.A., Benoy, G.A., and Robertson, D.M., 2018, Estimates of long-term mean-annual nutrient loads considered for use in SPARROW models of the Midcontinental region of Canada and the United States, 2002 base year: U.S. Geological Survey Scientific Investigations Report 2018–5051, 14 p., accessed February 17, 2019, at <https://doi.org/10.3133/sir20185051>.
- Schwarz, G.E., 2019, E2NHDPlusV2_us: Database of Ancillary Hydrologic Attributes and Modified Routing for NHDPlus Version 2.1 Flowlines: U.S. Geological Survey data release, accessed October 10, 2019, at <https://doi.org/10.5066/P986KZEM>.
- Schwarz, G.E., and Wiczorek, M.E., 2018, Database of modified routing for NHDPlus Version 2.1 flowlines: ENHDPlusV2, 2018: U.S. Geological Survey data release, accessed October 3, 2018, at <https://doi.org/10.5066/P9PA63SM>.
- Schwarz, G.E., Hoos, A.B., Alexander, R.B., and Smith, R.A., 2006, The SPARROW surface water-quality model: Theory, application, and user documentation: U.S. Geological Survey Techniques and Methods, book 6, chap. B3, available online only at <http://pubs.usgs.gov/tm/2006/tm6b3/>.
- Skinner, K.D., and Wise, D., 2019, Point-source nutrient loads to streams of the conterminous United States, 2012: U.S. Geological Survey data release, accessed June 12, 2019, at <https://doi.org/10.5066/P9PYVPFT>.
- Smith, R.A., Schwarz, G.E., and Alexander, R.B., 1997, Regional interpretation of water-quality monitoring data: *Water Resources Research*, v. 33, no. 12, p. 2781–2798.
- Sprague, L.A., and Gronberg, J.M., 2013, Estimated anthropogenic nitrogen and phosphorus inputs to the land surface of the conterminous United States—1992, 1997, and 2002: U.S. Geological Survey Scientific Investigations Report 2012–5241, 14 p., accessed June 23, 2018, at <https://pubs.usgs.gov/sir/2011/5167/>.
- Steves, P., and Nebert, D., 1994, 1:250,000-scale Hydrologic Units of the United States: U.S. Geological Survey Open-File Report 94–0236, accessed September 20, 2018, at <https://water.usgs.gov/lookup/getspatial?huc250k>.
- Stewart, D.W., Rea, A., and Wolock, D.M., 2006, USGS streamgages linked to the medium Resolution NHD: U.S. Geological Survey Data Series, DS–195, accessed September 15, 2018, at <https://water.usgs.gov/GIS/metadata/usgswrd/XML/streamgages.xml>.
- Tsang, Y.P., Wieferich, D., Fung, K., Infante, D.M., and Cooper, A.R., 2014, 3:589 An approach for aggregating upstream catchment information to support research and management of fluvial systems across large landscapes: *SpringerPlus*, v. 3, accessed May 12, 2015, at <http://www.springerplus.com/content/3/1/589>.
- U.S. Environmental Protection Agency, 2018, PCS-ICIS Search: U.S. Environmental Protection Agency Web site, accessed February 22, 2017, at <https://www.epa.gov/enviro/pcs-icis-search>.
- U.S. Geological Survey (USGS), 2017, Geographic Names Information System (GNIS)—USGS National Map Downloadable Data Collection: U.S. Geological Survey, accessed September 15, 2018, at <https://catalog.data.gov/harvest/object/a88f4496-63ed-43cb-a352-c48e89efbc24/html>.
- Wiczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2019, Select Attributes for NHDPlus Version 2.1 Reach Catchments and Modified Network Routed Upstream Watersheds for the Conterminous United States (ver 2.0, October 2019): U.S. Geological Survey data release, <https://doi.org/10.5066/F7765D7V>.
- Wiczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2018, Select attributes for NHDPlus Version 2.1, Reach catchments and modified network routed upstream watersheds for the conterminous United States: U.S. Geological Survey data release, <https://doi.org/10.5066/F7765D7V>.

Appendixes 1–5

Appendix 1. Description of Addition and Removal Events Spreadsheet

The specification of Addition and Removal (AR) events is through the AR events spreadsheet (Schwarz, 2019). The spreadsheet contains information on the network locations of withdrawal and discharge components of the events, connections between pseudo flowlines, flows associated with each event, characteristics of the event needed for incorporation into models, and reference information that documents information sources. For disconnected withdrawal or discharge events (and non-routed transfers), each observation in the spreadsheet represents an individual withdrawal or discharge. For routed transfers, an observation can be either the withdrawal or discharge point of the transfer, where a pseudo flowline is created, or the observation can represent connections of pseudo flowlines that occur between transfer withdrawals and discharges. Pseudo flowlines have no nodal structure to describe the connections between them. Consequently, the observations in the spreadsheet pertaining to routed transfers represent connections in terms of FromComid and ToComid (table 1.1). Every routed transfer will consist of at least three entries in the spreadsheet: two observations corresponding to the withdrawal and discharge events and an observation that shows the connection between the two pseudo flowlines created from these events.

The spreadsheet includes several variables useful for incorporating the AR events in models. A list of these variables with a description of each one is presented in table 1.1. The AREventID is used to link multiple observations pertaining to a common event, for example, the disconnected withdrawal and discharge events associated with a common use, such as single municipal public-supply or irrigation district. The ARSrcReg can be used to identify AR events that initiate outside the region in which the NHDComid flowline resides (given by the variable Region). This could be useful in modeling for determining where boundary conditions may be needed to replace a routed transfer because the transfer is outside the model study area. The AR events spreadsheet currently consists of 294 observations.

FromComid and ToComid identify connections between NHDPlusV2 flowlines and AR events, and between pseudo flowlines associated with routed transfers. For a disconnected withdrawal, FromComid is the NHDPlusV2 flowline Comid where the withdrawal is located, and ToComid is blank. For a disconnected discharge, ToComid is the NHDPlusV2 Comid where the discharge is located, and FromComid is blank. For routed transfers, FromComid and ToComid are generally valued, with FromComid corresponding to an NHDPlusV2 Comid for the initial routed-transfer withdrawal and ToComid corresponding to an NHDPlusV2 Comid for the terminating routed-transfer discharge. In these cases, the complement ToComid or FromComid relates to a pseudo flowline with a unique identification number having a value less than -90000000. The ARQFrom variable represents the amount

of flow, expressed in cubic feet per second, in the flowline designated by the FromComid. If FromComid has a positive value, meaning the observation pertains to a withdrawal event (either disconnected or routed transfer), this is the flow in the NHDPlusV2 flowline from which the withdrawal is taken before the withdrawal is made. For disconnected discharge events, there is no FromComid, and the ARQFrom is blank. ARQuantity is the flow, expressed in cubic feet per second, routed to the ToComid flowline. If the event is a disconnected withdrawal event, in which case the ToComid value is blank, ARQuantity represents the amount of flow withdrawn from the FromComid. In contrast, if the event pertains to a discharge event, either for a disconnected discharge or to complete a routed transfer (in which case ToComid is a positive number), the stated flow is the amount discharged and added to the ToComid flowline. For discharges of routed transfers, ARQuantity is not generally relevant because the flow discharged will be determined by the amount routed to the ToComid; however, the amount can be relevant as a boundary condition if the source region for the transfer (ARSrcReg) is outside the model study area. If ARQuantity and ARQFrom are valued, then ARQuantity cannot exceed ARQFrom.

The ARQFrom and ARQuantity variables can be used to define the fraction of flow diverted either from the network or to a pseudo-flowline diversion path. If the FromComid is positive, meaning the observation is either a disconnected withdrawal or the withdrawal point for a routed transfer, then the share of flow diverted from the network is given by the ratio of ARQuantity to ARQFrom. If the observation represents a connection between two pseudo flowlines as part of a routed transfer, and ARQuantity is less than ARQFrom, then the flow in the FromComid is diverted, and the fraction diverted to the ToComid flowline is the ratio of ARQuantity to ARQFrom. These ratios can be used in modeling to remove flow from the network without requiring the absolute diverted flow be less than the flow in the NHDPlusV2 flowline.

The variables IfPublic and IfIrr are 0/1 dichotomous variables used to identify flowlines as either public supply (IfPublic = 1 and IfIrr = 0) or irrigation (IfPublic = 0 and IfIrr = 1). Both variables are zero if the observation pertains to a transfer, either routed or non-routed, or another unspecified activity. ARDWType describes the type of withdrawal/discharge event. ARDWType is 0 if the observation pertains to a routed transfer (both FromComid and ToComid are coded with values), 1 if the observation pertains to public-supply withdrawal or discharge (IfPublic = 1), 2 if the observation pertains an irrigation withdrawal or discharge (IfIrr = 1), and 3 if the event is a non-routed transfer withdrawal or discharge (FromComid or ToComid is blank, and IfPublic and IfIrr are 0). The ARDWType variable can be used in modeling to differentiate the quality of discharge between the two principal withdrawal uses, public supply and irrigation. The AREventCd

Table 1.1. Description of variables in the Addition/Removal spreadsheet found in Schwarz and Wieczorek, 2018.

Variable	Description
ARObs	Sequential number for each observation.
AREventID	Sequential number for groupings of observations signifying a single event (a single row for non-transfer withdrawal or discharge events; at least observations for non-routed transfer events (the withdrawal observation and the discharge observation; and at least three observations for a routed transfer (adding -90000000's pseudo flowlines to the network; for routed transfers at least one -90000000 flowline must be specified).
Region	The up-to-3-character NHDPlusV2 production region code for the end point of the event.
ARSrcReg	The up-to-3-character NHDPlusV2 production region code for the initial point of the event.
FromComid	The NHDPlusV2 comid corresponding to where a withdrawal initiates, or the upstream comid of a pseudo-flowline created as part of a routed transfer. Leave blank for discharge events. For virtual flowlines created to support a transfer, designate a unique number in the -90000000's; check that there are no conflicts.
ToComid	The NHDPlusV2 comid corresponding to where a discharge occurs, or the downstream comid of a pseudo-flowline created as part of a routed transfer. Blank for withdrawal events. See numbering convention described for FromComid.
ARQFrom	Flow (cubic feet per second) in FromComid, prior to withdrawal; blank for discharge events (ToComid is blank).
ARQuantity	For withdrawal events (ToComid is blank), this is flow (cfs) removed from the FromComid; for transfer events with both FromComid and ToComid specified, this is the flow originally in FromComid that is routed through ToComid (note, it does not necessarily equal the flow in ToComid as there may be other upstream FromComid's that route to the same ToComid; for discharge to the non-virtual network (ToComid is > 0), this is the flow added to ToComid (in addition to flow that is already in the ToComid).
Units	The units for flow (always use cfs).
IfPublic	0/1 variable equals 1 if the row corresponds to either a withdrawal or discharge (either FromComid or ToComid is blank) and the use or source of the water is a public supply facility.
IfIrr	0/1 variable equals 1 if the row corresponds to either a withdrawal or discharge (either FromComid or ToComid is blank) and the use or source of the water is irrigation.
ARDWType	Withdrawal/discharge type. Takes the value: 0 if a routed transfer, 1 if withdrawal/discharge is associated with public supply (IfPublic = 1), 2 if withdrawal/discharge pertains to irrigation (IfIrr = 1), and 3 if a non-routed transfer (both IfPublic and IfIrr are 0).
AREventCd	Value is 1 if a routed transfer (transfer with either FromComid or ToComid in the -90000000s), 2 if withdrawal (public supply, irrigation, or non-routed transfer), and 3 if discharge (public supply, irrigation, or non-routed transfer).
ARName	Descriptive name for the AR event component.
Notes	Text explaining the event context, along with information sources for the ARQFrom and ARQuantity (specify station ids with range of years on which flow is based, if available).
OFRID_E	Reference item number if event refers to, Mooty, W.S., and Jeffcoat, H.H., 1986, Inventory of interbasin transfer of water in the Eastern United States, U.S. Geological Survey Open-File Report 86-148, 47 p.
OFRID_W	Reference item number if event refers to, Petsch Jr., H.E., 1989, Inventory of interbasin transfers of water in the western conterminous United States, U.S. Geological Survey Open-File Report 85-166, 45 p.
NHDComid	NHDPlusV2 comid equal to either FromComid or ToComid, if either is coded with a comid from NHDPlusV2, and blank otherwise.
Lat	The latitude of the midpoint of NHDComid.
Lon	The longitude of the midpoint of NHDComid.

describes the type of observation and takes a value of 1 if the observation is part of a routed transfer, 2 if the event is a discharge (including non-routed transfers), and 3 if the event is a withdrawal (including non-routed transfers).

The Notes variable contains information describing the context of the observation and documents the source of the flow information. The flow documentation is typically given in parentheses, with the first part representing the source of the withdrawal or discharge flow and the second part corresponding to the NHDPlusV2 flow. The form of the flow documentation may be a URL (Universal Resource Locator) linking to a web page that provides flow data (typically for public-supply facilities), a streamgage number and period used to form an average flow, or other reference.

It is not necessary for an AR routed transfer to reconnect to the original network. It is possible that a routed transfer ends prior to returning water to the network because the entire consumption of the routed water is associated with some use. In this case, FromComid will have a negative value corresponding to the terminal pseudo flowline, and ToComid is blank. To be consistent with other coding conventions, the terminal pseudo flowline is classified as a disconnected withdrawal event (AREventCd = 2) and has a designated use given by the ARDWType code, and possibly one of the IfPublic or

IfIrr variables is non-zero. The terminal pseudo flowline is identified as a terminal flowline in the network.

All disconnected withdrawal observations in the AR events spreadsheet are either paired with a disconnected discharge observation or the discharge occurs off the network and terminates in the ocean. If future observations are added to the spreadsheet that are not paired, it may be necessary to create another variable that distinguishes them from paired observations. For paired observations, the withdrawal of water from the network can be interpreted as a total withdrawal; any water that is not consumed by this withdrawal is returned to the network by another AR event. However, if a withdrawal observation is not paired with discharge observation, and discharge does in fact occur from this withdrawal, then the observation must be adjusted to reflect the net consumption of water.

Reference Cited

Schwarz, G.E., 2019, E2NHDPlusV2_us: Database of Ancillary Hydrologic Attributes and Modified Routing for NHDPlus Version 2.1 Flowlines: U.S. Geological Survey data release, accessed October 10, 2019, at <https://doi.org/10.5066/P986KZEM>.

Appendix 2. Description of Methods Used to Update Streamflow Estimates

The method to update streamflow utilized the intermediate incremental catchment flow in NHDPlusV2.1 that was dependent on unit runoff (Qincr0001A), added an adjustment for stream evaporation (Qincr0001B), and produced selective substitutions of regression estimates based on reference streamgage unit runoff (Qincr0001C). The original EROM method (McKay and others, 2012) was applied to the network (NHDPlusV2) in individual production units, approximately equivalent to two-digit hydrologic regions. Because some of these regions are hydrologically upstream from other regions, the original EROM method incorporated boundary conditions into the incremental catchment flow where inflows from an upstream region enter a given region.

In the application of the EROM method to the revised network (E2NHDPlusV2_us), the entire network was processed simultaneously, making it necessary to remove the boundary condition flow component from the incremental catchment flows at the boundary locations. This adjustment required accumulating the values of Qincr0001C, starting from the headwaters of the upstream region (or regions, in the case of multiple upstream regions terminating at the same point) through to the most downstream node of flowlines leaving each region. In the downstream regions, for flowlines representing the continuation of flow from the upstream region, the accumulated flow in the upstream node was subtracted from the flowline's incremental catchment flow Qincr0001C that included a boundary condition. In performing this operation, it was discovered that the original EROM method failed to incorporate the flow from the Ohio River into the Mississippi River, an oversight that was corrected in the revised flows. It was also determined that the boundary condition for the boundary between regions 14 and 15 was too small to account for the incoming flow from region 14 but that the difference was incorporated into the incremental flow of the next flowline downstream from the region 15 boundary flowline.

Given a consistent set of incremental level-C catchment flows (in this application, with boundary conditions removed), the EROM method added or subtracted flow quantities known as Addition/Removal (AR) events. These events represent withdrawals or discharges of flow owing to water use or inter-basin transfers. The revised network (E2NHDPlusV2_us) incorporated a number of AR events not included in the original NHDPlusV2 and revised quantities of many events that were common to both. The resulting adjustments to incremental catchment flow Qincr0001C produced the revised incremental values Qincr0001D.

In assessing the veracity of the original EROM method it was determined that incremental flows were incorrectly computed for locations where the AR adjustment caused cumulative flow to go negative. The incremental flow Qincr0001D reported at these locations equaled the amount by which the AR adjustment exceeded level-C cumulative flow (Q0001C,

inclusive of the flowline's value of Qincr0001C) rather than the amount that the level-D flow (computed as the analog of the NHDPlusV2 EROM variable, Qincr0001D; McKay and others, 2012) delivered from upstream flowlines, which must be reduced to drive level-D flow to zero. This error indicated the accumulation of level-D incremental flows was inconsistent with the reported cumulative flow. This error did not affect the derivation of revised flow undertaken in this study because the revised flow was based on the level-C incremental flows.

The last step of the EROM method used flow from streamgage measurements to adjust the AR-adjusted incremental flows. Two adjustments were applied, one downstream and one upstream. [It was discovered that the documented EROM method for this step was not the one used to derive the final EROM estimates contained in the existing NHDPlusV2.1 (the Q0001E flow estimates). Through detailed analysis, it was possible to determine the actual method used to produce the EROM estimates, although this method contained many logical flaws that were corrected in the revision.]

The downstream adjustment descended the network to accumulate level-D incremental flows. If a monitored flowline was encountered, the measured flow was substituted for the accumulated flow in proceeding the accumulation past the monitored flowline. The accumulation of incremental level-D flows continued downstream using the adjusted flow as the base for accumulation. An upstream adjustment was applied to any flowline within a monitored nested basin having cumulative area at least one-half as large as the cumulative area at the nearest downstream monitored flowline. For these qualifying flowlines, the downstream-adjusted level-D flows were adjusted on the basis of two factors: a base adjustment factor, common to all qualifying flowlines, and a flowline-specific apportionment ratio that apportions the base adjustment to an individual qualifying flowline. To obtain the adjustment, the two factors were multiplied, and the product was added to the downstream-adjusted, cumulative, level-D flow.

The original EROM method used as the base adjustment factor the difference between measured flow at streamgages and the monitored flowline's value of non-adjusted, cumulative, level-D flow. The apportionment factor for this method used the ratio of the qualifying flowline's cumulative area to the sum of cumulative area over all qualifying flowlines in the monitored nested basin. The ratio is generally a small number that grows in a downstream direction but remains small, even at the monitored flowline.

A problem with the EROM methodology just described is that the streamgage adjustment formula, if applied to the monitored reach, would not result in a final flow estimate that equaled measured flow at that flowline. This inconsistency arose for two reasons. First, the base adjustment for the monitored nested basin is given by the difference between streamgaged flow and cumulative, level-D flow at

the monitored location, whereas the actual discrepancy in flow should be the difference between measured flow and the downstream-adjusted, cumulative, level-D flow, prior to applying the downstream adjustment at the monitored flowline. Second, the apportionment factor for the monitored flowline is given by the ratio of its cumulative area to the sum of the cumulative area of all qualifying flowlines in the nested basin, a ratio that is considerably less than one. Because the downstream adjustment set flow at a monitored flowline is equal to the streamgaged flow, the adjustments tended to show a discontinuous jump between the monitored flowline and the flowlines on the main flow path immediately upstream.

The revised EROM method used in this analysis corrected the original EROM methodology by smoothing the upstream adjustment. First, the base adjustment for the monitored nested basin was determined by the difference between measured flow at the streamgaged and the downstream-adjusted, cumulative, level-D flow at the monitored flowline, prior to applying the adjustment at the monitored flowline. Second, the apportionment factor was given by the ratio of the qualifying flowline's sum of cumulative area, as determined from a downstream accumulation of the qualifying flowlines' cumulative area, to the sum of cumulative area across all qualifying flowlines in the monitored nested basin. As a result, the ratio was the same as the original EROM method's ratio for the most upstream, adjustment-qualifying flowlines; the numerator of the ratio was the drainage area of the flowline under both methods. However, the revised ratio increased more rapidly in the downstream direction than the original ratio, with the ratio equaling one at the monitored flowline. Thus, the upstream adjustment if applied at the monitored flowline would exactly equal the difference between measured flow at the streamgage and the downstream-adjusted, cumulative, level-D flow, yielding an adjusted flow that exactly equaled measured flow. The result of the modified adjustments was a smoother transition of flow in flowlines approaching the monitored flowline.

Table 2.1 of this appendix shows the relative performance of flow estimation across different estimation methodologies and networks. The performance measures were determined using a sequestration method, whereby the flow-estimation algorithm was applied using a smaller set of streamgages than was used to derive the results. Flow estimates at the locations of the excluded streamgages were then compared to measured flows at streamgages to determine a percent error (100 times the ratio of the difference between measured and estimated flow to estimated flow). The first set of columns in table 2.1 pertains to the original EROM methodology applied to the original NHDPlusV2 network. The middle set of columns pertains to the revised EROM methodology described above, also applied to the original NHDPlusV2 network. The last set of columns pertains to the revised EROM methodology

applied to the revised network described in this present analysis. Although not uniform across all regions, the new EROM methodology tended to outperform the original methodology, holding the network fixed. Significant improvements in root mean-squared percent error were observed using the revised EROM methodology for regions 03S, 11, 13 and 15 (fig. 2 in the body of the report), with the overall root mean-squared percent error falling from 58.1 to 56.5 percent. Additional improvement in estimates from using the revised network was more modest, with only region 15 showing a significant reduction in root mean-squared percent error. Switching to the revised network lowered the overall root mean-squared percent error by only 0.1 percent.

The method for regenerating stream velocity differs slightly from that used in the original production of NHD-PlusV2. In the original NHDPlusV2 the determination of velocity differs depending on the availability of channel slope. Equation 12 of Jobson (1996) is used if slope is known, and equation 14 is used if slope is not known. This distinction is made even though mean annual flow is zero, a questionable approach given that both Jobson equations show that, if flow is zero, there is no effect of slope on the determination of velocity. Consequently, in regenerating velocity it was assumed that if mean annual flow is zero, then velocity is set to the small value 0.06561, which is consistent with the unknown-slope equation evaluated at zero mean annual flow. The estimation of seasonal velocity based on the Jobson equations depends on mean annual flow and seasonal flow. The equations have the property that there is no effect of slope on velocity if either mean annual or seasonal flow is zero. Consequently, if either mean annual or seasonal flow was zero, velocity was set to the value 0.06561.

The original NHDPlusV2.1 flow estimates accounted for losing streams by including an adjustment that interpolated flow between measured flows at monitored flowlines, an adjustment that implied incremental catchment flow was negative for losing streams (McKay and others, 2012). Because streamflow was established for the existing network routing, it was possible that flowlines with negative incremental flow made flow accumulated for the revised network routing negative. The problem of negative flow is exacerbated by frequent inconsistencies in the original NHDPlus data, as described in table 6 of this report, where flow cannot be reproduced by accumulating incremental flow. Consequently, negative accumulated flow for the revised network was not uncommon when such flows were not constrained to be non-negative during the process. Improvements could be made if an algorithm for the estimation of cumulative and incremental flow that is not as sensitive to network routing were developed. The smoothed elevations (MinElSmoCm and MaxElSmoCm) used to estimate channel slope also are sensitive to network routing and could benefit from a revised algorithm.

Table 2.1. Relative performance of the EROM and revised EROM flow estimation methods, applied to the original NHDPlusV2 and revised E2NHDPlusV2_us networks.

Region	Number of sequestered gages	Original EROM method applied to NHDPlusV2 network			New EROM method applied to NHDPlusV2 network			New EROM method applied to E2NHDPlusV2_us network		
		Mean percent error	Standard deviation percent error	Root mean square percent error	Mean percent error	Standard deviation percent error	Root mean square percent error	Mean percent error	Standard deviation percent error	Root mean square percent error
1	52	3.1	12.1	12.3	3.3	12.2	12.5	3.3	12.2	12.5
2	129	-0.7	34.9	34.8	-0.4	34.4	34.3	-0.2	34.6	34.5
03N	54	3.8	15.8	16.1	3.7	15.6	15.9	3.8	15.6	15.9
03S	32	12.6	51.9	52.6	7.1	30.5	30.8	7.1	30.5	30.8
03W	44	6.5	26.2	26.7	6.8	26.1	26.7	6.6	26.1	26.7
4	69	-4.9	17.1	17.6	-5.6	16.9	17.6	-5.6	16.9	17.6
5	115	3.2	12.7	13	3.1	12.6	12.9	3.1	12.6	12.9
6	25	-0.2	9.5	9.3	0.1	9.4	9.2	0.1	9.4	9.2
7	82	6	25	25.6	5.7	25	25.5	4.7	24.4	24.7
8	21	3.4	14	14	3.5	13.9	14	3.6	13.9	14
9	13	14	22.6	25.8	12.1	25	26.9	11.5	23.8	25.6
10L	71	8.2	66.3	66.3	8.5	63.2	63.4	8.2	63.1	63.2
10U	70	-0.1	49.9	49.6	0.9	48.6	48.3	0.9	48.6	48.3
11	76	-7.9	56.5	56.7	-13.8	50.4	52	-15.6	49.2	51.3
12	64	13.7	79.6	80.2	10	78.5	78.5	10	78.5	78.5
13	21	0.7	111.9	109.2	-2.5	107.2	104.8	-2.5	107.2	104.8
14	63	-21	32.9	38.9	-20.1	32.5	38	-18.6	32.6	37.3
15	27	-19.5	48.5	51.5	-21.1	41.7	46	-22.9	37.8	43.6
16	37	-1.9	42.9	42.3	0.2	43.2	42.6	0.4	43.1	42.6
17	125	-4.1	34.6	34.7	-4.3	33.8	33.9	-5.6	34.1	34.4
18	107	15.8	143.1	143.3	14.5	141.6	141.7	14.4	141.8	141.9
All	1,297	1.4	58.1	58.1	0.7	56.5	56.5	0.4	56.5	56.4

References Cited

- Jobson, H.E., 1996, Prediction of traveltime and longitudinal dispersion in rivers and streams: U.S. Geological Survey Water-Resources Investigations Report 96-4013, 69 p.
- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., 2012, NHDPlus Version 2: User Guide: U.S. Environmental Protection Agency, 182 p., accessed May 1, 2018, at https://s3.amazonaws.com/nhdplus/NHDPlusV2/Documentation/NHDPlusV2_User_Guide.pdf.

Appendix 3. Description of Methods Used to Update Slope Estimates

The NHDPlusV2.1 method used to compute channel slope established an elevation for each flowline node based on the largest value of the minimum catchment elevations among all flowlines conflating at the node. For headwater nodes, the elevation was simply established by the minimum elevation within the sub-catchment area that flowed to the headwater node. The method then started from the most downstream node and navigated the network upstream along a path having the greatest flow to find the next upstream node along a path with an elevation value that was higher than that at the downstream node.

Nodes so identified are called set points of the slope determination method. The difference in elevations between these two successive set-point nodes, divided by the sum of the flowline lengths between the two nodes determined the slope for all flowlines along the connected pathway. The elevations of all nodes along the connecting path were then re-established by linear interpolation using the estimated slope.

Slopes for tributaries of the level path were determined using the interpolated or established elevation at the node at which the original level path and the tributary level path intersected and climbed the tributary level path to find the next upstream node having an elevation value greater than the intersecting node. In this way, by climbing the network along level paths to find the next local maximum elevation, slopes were estimated for all flowlines in the network. If no upstream node with an elevation value greater than the downstream node's established elevation was observed, a default slope of 0.00001 was used for all flowlines upstream from the established node. Flowlines classified as being in a waterbody also were assigned the default slope (McKay and others, 2012).

A revised method used here initialized the flowline nodes with the same initial elevations as the NHDPlusV2.1 method; these elevations were the maximum of the minimum initial catchment elevations (MinElRawCm), corresponding to all catchments directly upstream from a given node, rounded to the nearest centimeter. For headwater flowlines, the upstream node was initialized with the flowline's reported maximum initial catchment elevation (MaxElRawCm); this elevation is the minimum raw elevation corresponding to the headwater node drainage area—the area determined by the digital elevation model to flow through the headwater flowline's upstream node.

The revised method then expanded the search for set points beyond a given path to include all nodes upstream from the current set point. Given the initial set-point elevations at the terminus nodes of the network, determined according to the initial elevations for these nodes, the algorithm climbed the network in reverse hydrologic order (as determined by the NHDPlus hydrosequence variable) to identify a node having an initial elevation that is greater than all initial elevations downstream from it. The nodes so identified represent

a collection of set points having monotonically increasing elevations as the network is traversed in the upstream direction. Like the NHDPlusV2.1 method, the determination of set-point nodes excluded nodes having initial elevations that represented local minima.

The determination of set points did not guarantee that headwater nodes would be classified as set points. Such nodes were classified as set points only if they had an initial elevation that exceeded all downstream initial elevations. An alternative, regression-based method was used to determine elevation in the case of headwater nodes and any subsequent nodes immediately downstream that did not qualify as set points nor had an upstream set point on which to interpolate an elevation.

In the next step, the interpolation algorithm descended the network in hydrologic order to determine for each node the upstream set point having the minimum elevation and the maximum flowline distance between the node and this set point. If the node was classified as a set point, then the node referenced its own initial elevation as the minimum upstream set-point elevation, and the distance was set to zero. If there were multiple upstream set points having the same minimum elevation, then the set point with the greatest stream distance was referenced. This convention indicates that if there were diversions creating multiple paths connecting the node to an upstream set point, the effective path forming the connection was the one having the greatest distance.

In the last step, the interpolation climbed the network to determine an elevation for each node that is not a set point but referenced a set point somewhere upstream. Starting from the most downstream flowline in the network, which had an initialized value for the downstream node elevation, the algorithm climbed the network to determine the elevation for each upstream node, E_u (eq. 3.1). If the node was a set point, then the elevation of the node was set equal to the node's initial elevation, E_0 . If the node was not a set point, then the elevation was determined by linearly interpolating an elevation based on the determined downstream node elevation, E_d ; the elevation of the minimum-elevation upstream set point, E_s ; the distance from that set point to the upstream node, D_s ; and the length of the flowline corresponding to the upstream node, L , according to the equation

$$E_u = E_d + (E_s - E_d) \cdot L / (L + D_s) \quad (3.1)$$

If the node corresponded to a diversion, and the node was not a set point, then there were multiple downstream node elevations, E_d , and applying the equation to each resulted in multiple alternative values for the node elevation, E_u . In such cases, E_u was assigned according to the maximum of the alternative values. The selection of the maximum elevation

assured that the elevation at the diversion node exceeded the elevations for all other nodes downstream.

Because each non-terminus set point represented the smallest elevation upstream from the downstream set-point elevation, a linear interpolation of elevation between that point and the downstream set point, following the connected path, resulted in elevations at all nodes along the path that were strictly below the smallest elevation set points in each tributary basin, if such a set point existed. Thus, if set points existed, a positive slope between them was guaranteed. Only if a tributary had no elevation above the downstream set-point elevation, meaning the tributary had no upstream set point, the regression-based slope would have been required.

An example network that demonstrates the relation between set points and interpolated elevation is shown in figure 3.1. In the example, an examination of the initial node elevations (shown only for set points) determines the basin has two basin-determined set points—one with an initial nodal elevation of 20 and the other with 25 (units are arbitrary). From this, we can assume that all other nodes have elevations less than or equal to the elevation of the terminus, which is 10. All flowlines along the red path reference the upstream set point that has an elevation of 20, and all the green flowlines reference the upstream set-point elevation of 25. Note, also, that the green path contains a diversion, which complicates the determination of elevation for the node coinciding with the diversion. Blue flowlines correspond to tributaries that have a headwater node elevation that is less than or equal to the elevation of the terminus node (10), indicating they have no upstream set point and therefore must have slope determined using the regression method.

The equations in figure 3.1 show the application of equation 3.1 to determine elevation, E_u , for each node in the downstream path of a set point. In applying these equations, the length of each flowline is given by the value equilibrated with the variable L , and D_s in equation 3.1 is evaluated by summing the lengths for all flowlines in the path between the downstream node of the flowline and the upstream set-point node. This approach leads to an ambiguity at the diversion node, which is resolved by taking the maximum of the alternative estimated elevations.

The determination of slope and elevation for flowlines having no upstream set point was based on regression analysis. Separate regressions relating the logarithm of slope to a set of explanatory variables were estimated for each of the 21 NHDPlus production regions. The regressions were based on flowline observations for which the downstream node is a set point and the upstream node elevation is interpolated (not based on an initial elevation). The specific subset of flowlines used as observations in the regression indicates that the input data have the same characteristics as the set of flowlines to be filled in; therefore, both sets pertain to flowlines having no set point immediately upstream.

The explanatory variables for the regression include the logarithm of the distance to the upstream set point, the logarithm of the minimum slope pertaining to the flowline

immediately downstream (use the logarithm of 1E-7 if the flowline is terminal), the logarithm of total upstream drainage area (use the logarithm of 0.5 square kilometer if the flowline has no upstream drainage area), and the logarithm of elevation of the downstream node (use the logarithm of 1 centimeter if the downstream node has a non-positive elevation). The coefficients for the logarithms of distance to the upstream node and upstream drainage area are constrained in the regression to be non-positive; the coefficients for the logarithms of downstream slope and elevation are constrained to be non-negative.

The regression results for the slope equation are shown in appendix table 3.1. The results show that the estimated coefficients are generally highly significant with the expected sign. The elevation coefficient is constrained to equal zero for regions 04, 07, 09 and 14; the distance coefficient is constrained to equal zero for regions 13 and 16. The R^2 of the regressions ranges from 0.24 to 0.53, with an average of 0.41, and the root mean squared error (in logarithm space) ranges from 1.36 to 1.83 with an average of 1.56. These fits, while significant, indicate a considerable amount of residual uncertainty. The adjustment factor in the rightmost column corresponds to the retransformation-bias correction factor used to convert slope predictions from log to real space and is computed as the average of the exponentiated residuals from the log-space regression model.

The regression estimates from table 3.1 were used to estimate slope for flowlines with no upstream set point. In forming these estimates, the method begins with the most downstream flowlines in network without an upstream set point—the flowlines that have a set point for a downstream node but no upstream set point. However, to use the regression equation for prediction, an estimate of the distance to an upstream set point is required, a distance that is not known because there is no upstream set point. To determine the distance to be incorporated into the slope equations, a second set of regional regression equations are estimated, each having as the dependent variable the logarithm of distance to the upstream set point—the same variable used in the corresponding region's slope equation. The explanatory variables included in the second set of regressions are the same as those used in the slope regressions, except for the logarithm of distance to the upstream set point. The estimation results for the distance equations are given in appendix table 3.2.

The relevant relations for a given regional pair of models can be described as follows. Let \tilde{s} denote the logarithm of slope; \tilde{d} , the logarithm of distance to the upstream set point; X_0 , a row vector representing the set of explanatory variables excluding the logarithm of distance to the upstream set point; β_0 and θ_0 , vectors of parameters associated with X_0 in the slope and distance equations; β_d , the coefficient associated with the distance variable in the slope equation; and e and u , residuals in the slope and distance equations, assumed to be independent with mean 0 and having standard errors σ_e and σ_u . Independence of the residuals follows from the assumption that the slope equation is a valid regression model,

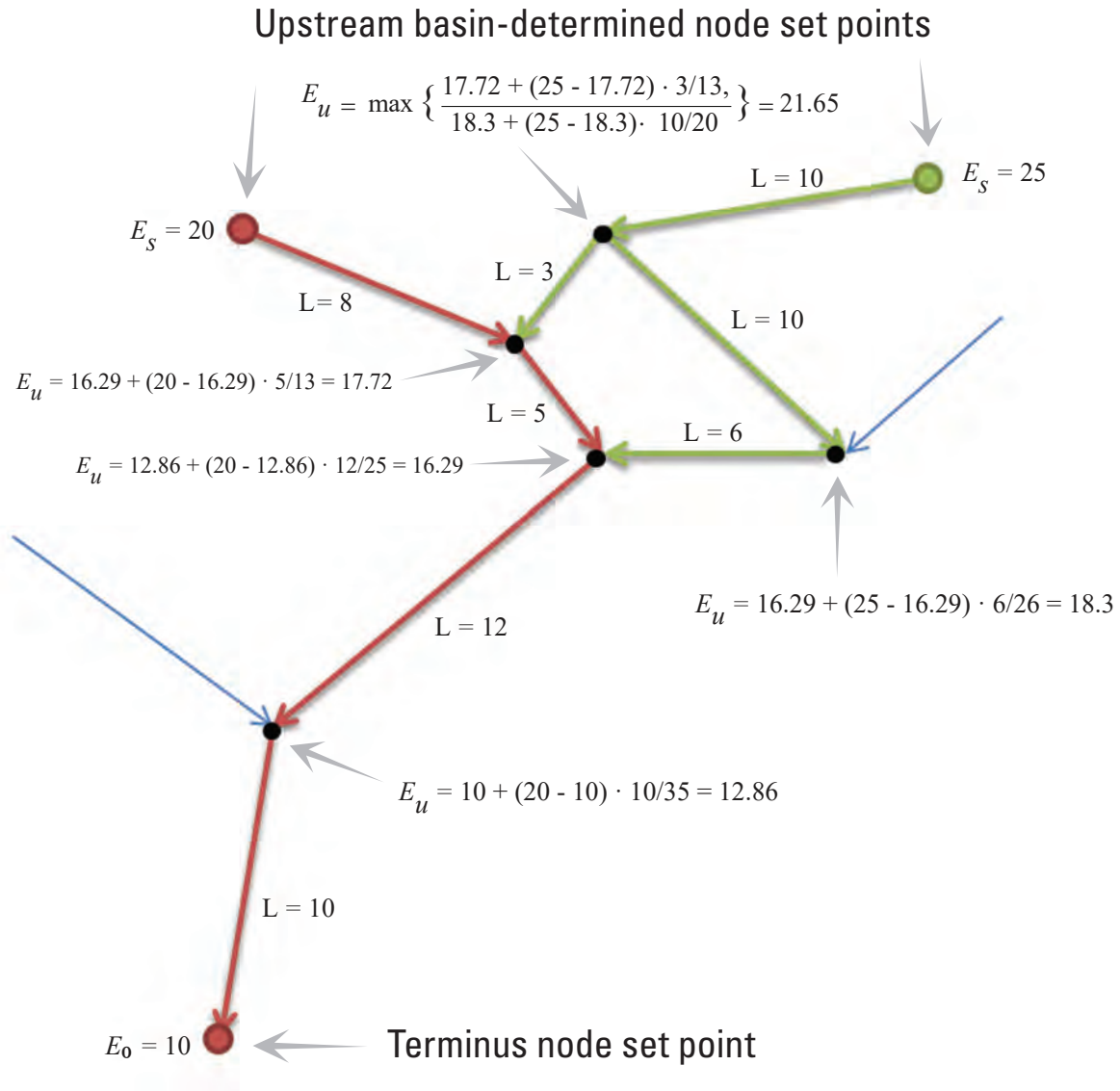


Figure 3.1. The calculation of node elevation by the interpolation method using basin-determined upstream set points.

Table 3.1. Regional regression results relating the logarithm of slope to various flowline predictors for regions in the conterminous United States.

[Observations are restricted to flowlines having a set point only for the downstream node. Regions correspond to the 21 NHDPlus production regions. Distance to the upstream set-point node corresponds to the longest possible channel path connecting the node to the flowline. Downstream slope is the minimum slope among all flowlines having an upstream node that connects to the flowline. Drainage area is diversion-flow apportioned drainage area. RMSE is root mean squared error. Adjustment factor is the average of the exponentiated model residuals.]

Region	Intercept	Coefficient estimates (standard errors)				Number of observations	R-square	RMSE	Adjustment factor
		Log distance to upstream set-point	Log down-stream slope	Log up-stream area	Log down-stream node elevation				
1	-1.997	-0.317	0.097	-0.38	0.133	5,387	0.286	1.825	3.454
	-0.329	-0.025	-0.011	-0.011	-0.013				
2	-4.461	-0.261	0.07	-0.342	0.317	7,645	0.398	1.655	2.793
	-0.245	-0.018	-0.01	-0.008	-0.009				
03N	-3.239	-0.395	0.106	-0.322	0.322	8,426	0.528	1.468	2.508
	-0.223	-0.016	-0.009	-0.006	-0.009				
03S	-3.143	-0.432	0.046	-0.286	0.258	3,673	0.435	1.537	2.534
	-0.338	-0.025	-0.012	-0.011	-0.012				
03W	-2.939	-0.28	0.089	-0.392	0.155	6,611	0.48	1.412	2.229
	-0.27	-0.019	-0.01	-0.008	-0.012				
4	-0.642	-0.364	0.173	-0.254	0	7,665	0.236	1.691	2.983
	-0.235	-0.02	-0.009	-0.009	0				
5	-6.16	-0.415	0.084	-0.407	0.665	8,964	0.452	1.517	2.363
	-0.471	-0.017	-0.009	-0.007	-0.037				
6	-7.309	-0.399	0.133	-0.445	0.813	3,151	0.478	1.535	2.677
	-0.671	-0.029	-0.014	-0.012	-0.053				
7	0.638	-0.471	0.145	-0.301	0	11,522	0.339	1.604	2.46
	-0.2	-0.016	-0.008	-0.006	0				
8	-0.357	-0.625	0.13	-0.208	0.212	11,148	0.381	1.555	2.532
	-0.22	-0.016	-0.008	-0.006	-0.009				
9	-0.12	-0.458	0.137	-0.239	0	2,431	0.334	1.606	2.68
	-0.394	-0.034	-0.016	-0.015	0				
10L	-8.291	-0.292	0.108	-0.218	0.629	12,299	0.356	1.364	1.956
	-0.275	-0.012	-0.008	-0.004	-0.019				
10U	-11.769	-0.134	0.178	-0.257	0.829	13,507	0.429	1.402	2.011
	-0.387	-0.011	-0.008	-0.004	-0.03				
11	-8.207	-0.251	0.14	-0.293	0.65	11,735	0.419	1.497	2.298
	-0.278	-0.014	-0.008	-0.005	-0.018				
12	-4.945	-0.242	0.098	-0.29	0.31	4,125	0.447	1.505	2.211
	-0.34	-0.023	-0.013	-0.009	-0.013				
13	-5.663	0	0.195	-0.291	0.212	1,860	0.459	1.449	2.283
	-0.432	0	-0.018	-0.01	-0.033				
14	-2.224	-0.08	0.122	-0.348	0	3,892	0.394	1.467	2.192
	-0.253	-0.021	-0.011	-0.008	0				
15	-2.141	-0.105	0.208	-0.304	0.056	3,165	0.42	1.402	2.646
	-0.521	-0.024	-0.014	-0.009	-0.034				
16	-15.456	0	0.251	-0.347	1.057	3,247	0.418	1.64	2.755
	-2.104	0	-0.012	-0.011	-0.173				
17	1.064	-0.407	0.288	-0.289	0.079	6,283	0.363	1.761	3.568
	-0.336	-0.024	-0.01	-0.009	-0.013				
18	-0.603	-0.281	0.292	-0.297	0.106	5,602	0.392	1.784	3.924
	-0.324	-0.024	-0.01	-0.01	-0.012				

Table 3.2. Regional regression results relating the logarithm of flowline channel distance to upstream set-point node to various explanatory variables for production units in the conterminous United States.

[Observations are restricted to flowlines having a set-point only for the downstream node. Regions correspond to the 21 NHDPlus production regions. Distance to the upstream set-point node corresponds to the longest possible channel path connecting the node to the flowline. Downstream slope is the minimum slope among all flowlines having an upstream node that connects to the flowline. Drainage area is diversion-flow apportioned drainage area. RMSE is root mean squared error]

Region	Coefficient estimates (standard errors)				Number of observations	R-square	RMSE
	Intercept	Log down-stream slope	Log upstream area	Log downstream node elevation			
1	11.504	-0.032	0.126	-0.017	5,387	0.092	0.997
	-0.088	-0.006	-0.006	-0.007			
2	11.678	-0.021	0.143	-0.059	7,645	0.121	1.049
	-0.079	-0.006	-0.005	-0.006			
03N	11.925	-0.046	0.113	-0.102	8,426	0.171	1.014
	-0.082	-0.006	-0.004	-0.006			
03S	11.677	0.002	0.19	-0.04	3,673	0.206	0.997
	-0.104	-0.008	-0.006	-0.008			
03W	11.846	-0.012	0.151	-0.055	6,611	0.191	0.9
	-0.091	-0.006	-0.004	-0.007			
4	12.167	-0.039	0.116	-0.071	7,665	0.087	0.982
	-0.304	-0.005	-0.005	-0.029			
5	15.058	0.001	0.14	-0.347	8,964	0.15	0.942
	-0.245	-0.006	-0.004	-0.023			
6	12.35	-0.028	0.1	-0.091	3,151	0.079	0.937
	-0.346	-0.009	-0.007	-0.032			
7	11.83	-0.04	0.092	-0.025	11,522	0.087	0.909
	-0.253	-0.005	-0.003	-0.025			
8	11.983	-0.067	0.085	-0.059	11,148	0.113	0.948
	-0.071	-0.005	-0.003	-0.005			
9	18.142	-0.044	0.209	-0.666	2,431	0.307	0.947
	-1.022	-0.01	-0.008	-0.095			
10L	13.36	-0.056	0.022	-0.171	12,299	0.035	1.004
	-0.163	-0.006	-0.003	-0.014			
10U	13.656	-0.081	0.039	-0.216	13,507	0.059	1.115
	-0.285	-0.006	-0.003	-0.024			
11	13.918	-0.057	0.078	-0.256	11,735	0.118	1.013
	-0.137	-0.005	-0.003	-0.012			
12	12.537	-0.03	0.118	-0.094	4,125	0.156	1.025
	-0.124	-0.009	-0.006	-0.009			
13	12.227	-0.041	0.089	-0.072	1,860	0.111	1.102
	-0.329	-0.013	-0.008	-0.025			
14	14.948	-0.053	-0.016	-0.268	3,892	0.013	1.125
	-1.173	-0.009	-0.007	-0.095			
15	12.647	-0.045	0.004	-0.079	3,165	0.014	1.029
	-0.31	-0.011	-0.006	-0.025			
16	22.606	-0.068	0.018	-0.905	3,247	0.08	1.025
	-1.314	-0.007	-0.007	-0.108			
17	12.182	-0.028	0.087	-0.072	6,283	0.089	0.939
	-0.092	-0.005	-0.005	-0.007			
18	11.663	-0.06	0.065	-0.029	5,602	0.079	0.983
	-0.087	-0.006	-0.005	-0.006			

whereby the residual e is assumed to be independent of all predictor variables, including \tilde{d} . The log slope and distance equations can be expressed as

$$\tilde{s} = X_0\beta_0 + \tilde{d}\beta_d + e, \quad (3.2)$$

$$\tilde{d} = X_0\theta_0 + u. \quad (3.3)$$

As explained above, in applying these relations to the prediction of slope, the distance variable is not directly measured; only the distance to the most upstream node in the network, which is not a set-point node, is measured. The implication is that u , the distance equation residual, is distributed as a truncated distribution, with the lower bound given by $d_0 - X_0\theta_0$, where d_0 is the truncated distance to the most upstream node in the stream network. Upon substitution of the distance equation into the slope equation, slope is shown to be a function of the slope residual, u . Assuming u is distributed truncated normal with a lower bound of $d_0 - X_0\theta_0$, taking the expectation of slope, S , in real space, conditioned on the slope residual e , gives

$$E[S|e] = \int_{d_0 - X_0\theta_0}^{\infty} \exp(X_0\beta_0 + (X_0\theta_0 + u)\beta_d + e) \left\{ \frac{(2\pi\sigma_u^2)^{-\frac{1}{2}} \exp\left(-\frac{u^2}{2\sigma_u^2}\right)}{1 - \Phi\left(\frac{d_0 - X_0\theta_0}{\sigma_u}\right)} \right\} du, \quad (3.4)$$

$$E[S|e] = \exp\left(X_0(\beta_0 + \theta_0\beta_d) + (\beta_d\sigma_u)^2 + e\right) \frac{1 - \Phi\left(\frac{d_0 - X_0\theta_0 - \sigma_u^2\beta_d}{\sigma_u}\right)}{1 - \Phi\left(\frac{d_0 - X_0\theta_0}{\sigma_u}\right)},$$

where $\Phi(z)$ is the standard normal cumulative distribution function. The term in braces in the first line of equation 3.4 represents the truncated normal distribution having a lower bound of $d_0 - X_0\theta_0$. Note that the effect of using a truncated distribution is to lengthen the effective upstream distance to be greater than d_0 . Given that upstream distance has a negative coefficient in the slope equation, this lengthening results in a lower slope estimate than would be determined otherwise.

The regression-based estimate of slope, unconditioned on knowing the residual e , is determined by evaluating equation 3.4 for given values of the exogenous variables, X_0 , using the coefficient estimates contained in tables 3.1 and 3.2, where the exponentiated slope residual term, $\exp(e)$, is replaced by a smearing estimate given by the mean of the exponentiated residuals shown as the adjustment factor in table 3.1.

Owing to the nonlinear functional form associated with the coefficients, the replacement of the coefficients with estimated values induces a bias in the slope estimate. To correct this bias a parametric bootstrap is performed, where the coefficients β_0 , β_d , and θ_0 are randomly generated 100 times from a multivariate normal distribution with mean vector and covariance matrix given by their respective estimates obtained from the regression results for the log slope and distance models. In generating these random values, θ_0 is treated as being independent of β_0 and β_d , in accordance with the independence of e and u . Random values of the scale parameter σ_u^2 are generated by multiplying the estimate of σ_u^2 from the distance regression by 100 randomly generated values from a chi-square distribution with degrees of freedom given by the distance regression results; the result is normalized by the number of degrees of freedom.

The bias-corrected slope estimate is obtained assuming a ratio form for the bias. Under this approach (see Schwarz and others, 2006), a multiplicative bias-correction term is computed as the ratio of the estimate of x using the regression coefficient estimates to the mean value of the bootstrap-generated values; the $\exp(e)$ term cancels from this formulation.

Thus, if $S(\hat{\beta}_0, \hat{\beta}_d, \hat{\theta}_0, \hat{\sigma}_u^2; X_0)$ is the regression-based estimate of equation x, with e set to zero, $S(\beta_0^{(r)}, \beta_d^{(r)}, \theta_0^{(r)}, \sigma_u^{2(r)}; X_0)$ is the bootstrap-iteration r evaluation of this function using parametric bootstrap iteration r of R randomly generated sets of the coefficients $\{\beta_0^{(r)}, \beta_d^{(r)}, \theta_0^{(r)}, \sigma_u^{2(r)}\}$, and \hat{e}_i is the i -th of N estimated residuals from the regression model, the bias-corrected estimate of slope is

$$\hat{S} = N^{-1} \sum_{i=1}^N \exp(\hat{e}_i) \frac{S(\hat{\beta}_0, \hat{\beta}_d, \hat{\theta}_0, \hat{\sigma}_u^2; X_0)^2}{R^{-1} \sum_{r=1}^R S(\beta_0^{(r)}, \beta_d^{(r)}, \theta_0^{(r)}, \sigma_u^{2(r)}; X_0)}. \quad (3.5)$$

Equation 3.5 is first applied to determine slope at the most downstream flowlines having no upstream set point. Given the estimate of slope, and knowledge of the downstream node elevation and flowline length, it is straightforward to estimate the upstream node elevation. As with the interpolation method, if there are multiple flowlines meeting at a diversion, the node is assigned the largest elevation determined among the diverting flowlines. The method progresses upstream to determine slope and upstream node elevation based on revised values of the predictor variables, X_0 and d_0 .

The revised method obtains slopes based on set points available for both the upstream and downstream nodes in 71 percent of the flowlines. Slope was determined for 23 percent of the flowlines by interpolation between set points beyond their upstream or downstream nodes. Slope was assigned for 6 percent of flowlines on the basis of the regression method, a considerable improvement over the original NHDPlus slopes which have fill-in values (the default slope 0.00001) assigned to 14 percent of all flowlines. The method is computationally efficient and operates simultaneously on the entire 2.6 million flowlines composing the national NHDPlusV2 flow-directed network. Computation requires no more than four passes through the network and can be completed in about 2 minutes.

Figure 3.2 shows the method of slope determination for flowlines in region 08, a region of fairly low relief at the mouth of the Mississippi River. Most small tributaries have slope determined directly from upstream or downstream set points. Interpolated slopes tend to be associated with larger channels, which are of lower relief than tributary channels. The regression-based fill-in values are primarily located in tributary headwaters near the Gulf of Mexico, where relief is low, even for tributaries.

The distributions of the revised slopes with the original NHDPlus slopes are shown in figure 3.3 for comparison. The figure shows that the original NHDPlusV2 slope estimates (fig. 3.3A) have two distributional anomalies: a spike in the distribution associated with the default fill-in slope value, 0.00001, and a discontinuous bump in the distribution for mid- to upper-range slope estimates (see areas circled in red). The revised estimates (fig. 3.3B) successfully eliminated these distributional anomalies.

One reason the default slope is so prevalent in the original NHDPlusV2 estimates is that it is automatically applied for any flowline classified as a waterbody, the reasoning being that waterbodies are effectively “flat.” The revised approach does not distinguish waterbodies. Consequently, the revised slopes for waterbodies can be interpreted as the slopes that would prevail if these flowlines were converted to flowing streams. This interpretation may be problematic for some applications, particularly the estimation of travel time for waterbodies. However, given that travel time for waterbodies is typically estimated using a method that depends on volume and flow, and not using the typical method for streams that depends on slope, this problem is moot.

Figure 3.4 shows the revised and original NHDPlus slopes compared to slopes determined at approximately 2,846 sampling sites of the U.S. Environmental Protection Agency’s National River and Stream Assessment (NRSA) and Wadeable Stream Assessment (WSA) that are indexed to the NHDPlusV2 network. The comparison shows good agreement between the revised slope estimates and the measured slopes. The bulge of disagreement in the slope values at a NRSA/WSA slope of 0.01 is due to rounding error in the NRSA/WSA estimates. As can be expected, the revised slopes show greater disagreement for smaller NRSA/WSA values. The original NHDPlus default slopes, appearing as a horizontal grouping at the value 0.00001, are shown to distribute over four orders of magnitude of NRSA/WSA slope estimates, indicating considerable bias associated with these default values.

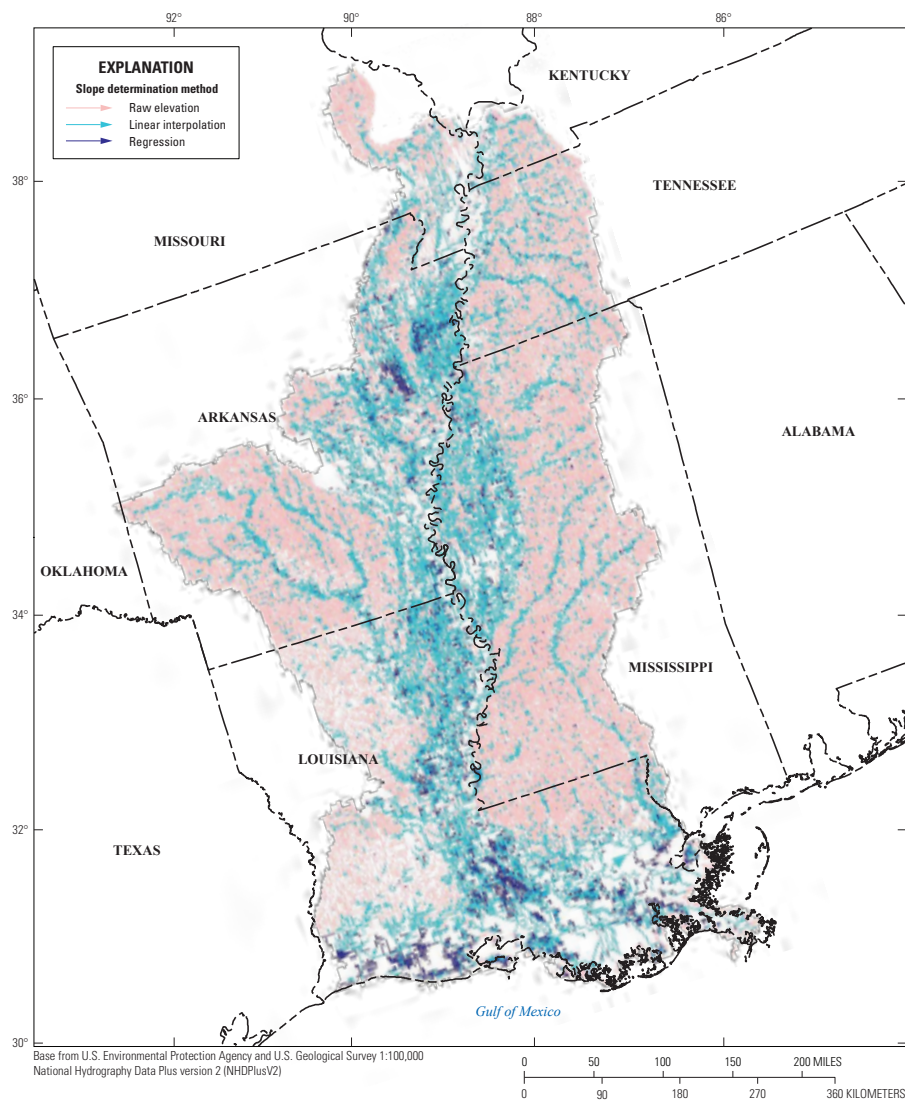


Figure 3.2. NHDPlusV2.1 streams in production unit 08 in the United States, and the specific method used to determine slope using the revised method.

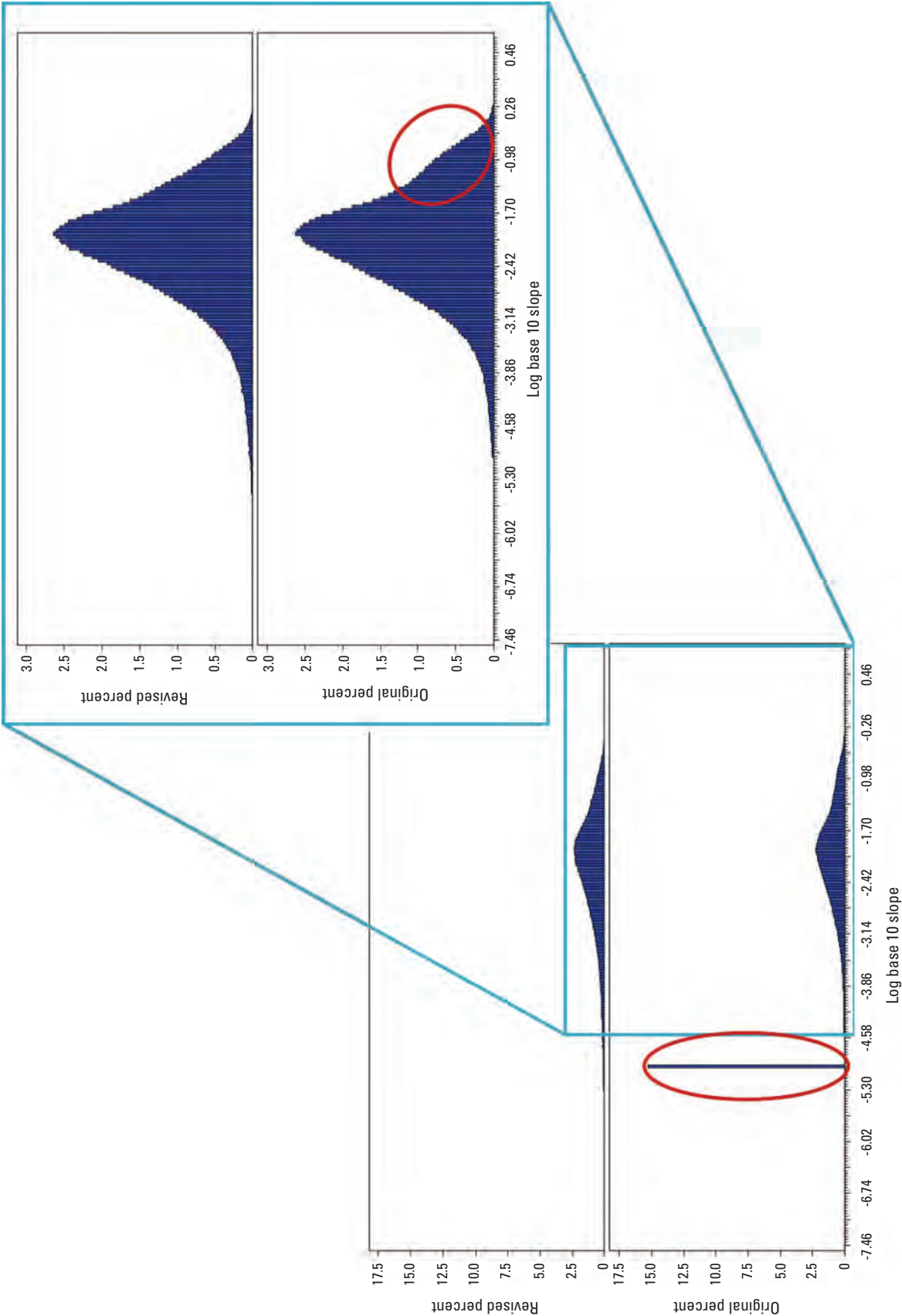


Figure 3.3. Slope distributions of the original NHDPlusV2 and revised methods.

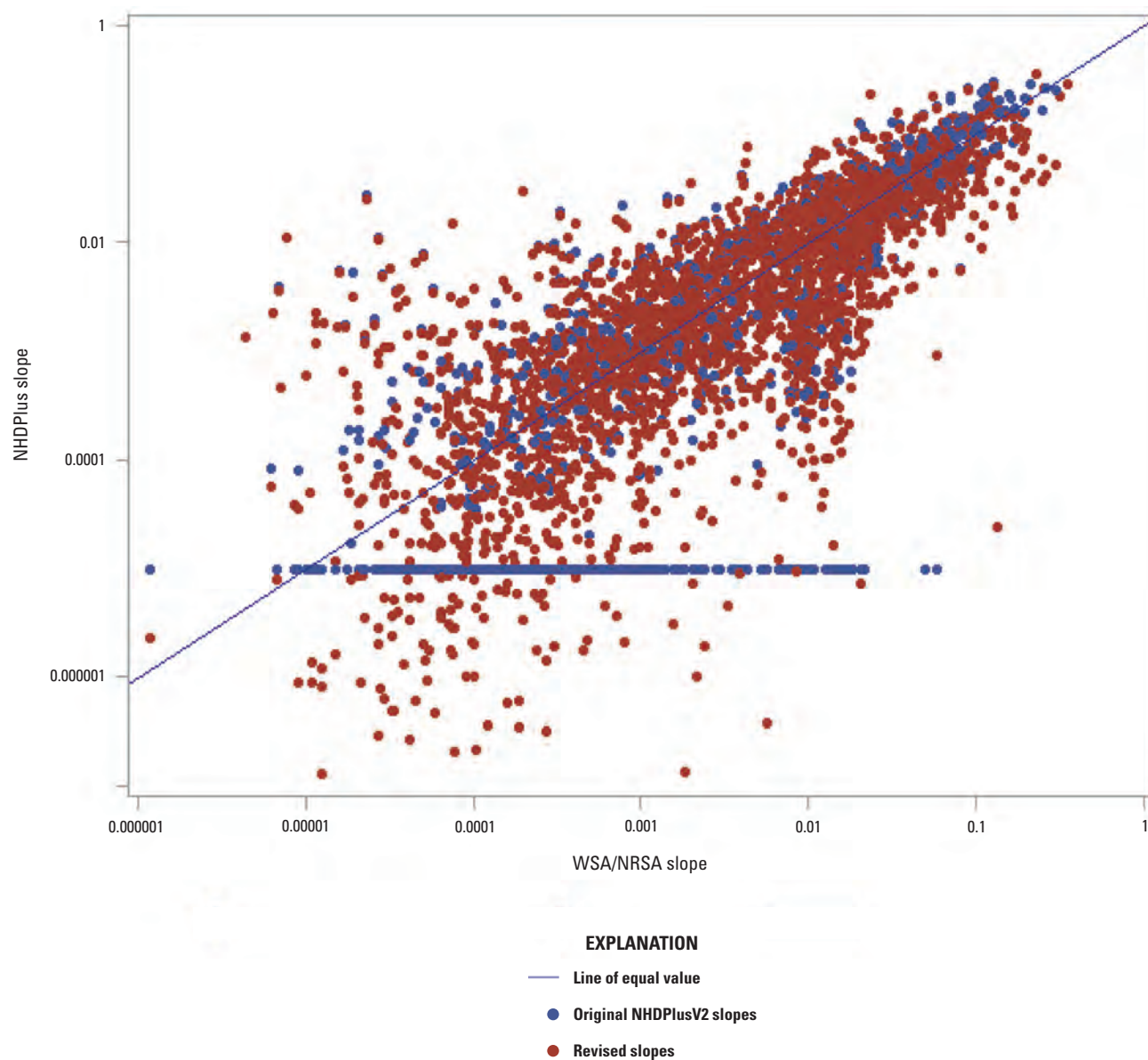


Figure 3.4. NHDPlus slopes determined according to the original NHDPlusV2 method and the revised method in relation to slopes measured at 2,846 sites indexed to NHDPlusV2 from the U.S. Environmental Protection Agency's Wadeable Streams Assessment and National River and Stream Assessment. WSA, Wadeable Streams Assessment; NRSA, National River and Stream Assessment.

References Cited

- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., 2012, NHDPlus Version 2: User Guide: U.S. Environmental Protection Agency, 182 p., accessed May 1, 2018, at https://s3.amazonaws.com/nhdplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf.
- Schwarz, G.E., Hoos, A.B., Alexander, R.B., and Smith, R.A., 2006, The SPARROW surface water-quality model: Theory, application, and user documentation: U.S. Geological Survey Techniques and Methods, book 6, chap. B3, accessed month, day, year, at <http://pubs.usgs.gov/tm/2006/tm6b3/>.
- U.S. Environmental Protection Agency, National Aquatic Resource Surveys, Wadeable Streams Assessment: U.S. Environmental Protection Agency, accessed March 30, 2017, at <https://www.epa.gov/national-aquatic-resource-surveys/wadeable-streams-assessment>.
- U.S. Environmental Protection Agency, National Aquatic Resource Surveys, National Rivers and Streams Assessment: U.S. Environmental Protection Agency, accessed March 30, 2017, at <https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys>.

Appendix 4. Description of Attributes in E2NHDPlusV2_us

Table 4.1. Description of attributes in E2NHDPlusV2_us.

[km, kilometer; km², square kilometer; cm, centimeter; mm, millimeter; deg, degree; neg, negative value; deg C, degrees Celsius; cfs, cubic feet per second; char, characters; %, percent; f/s, foot per second; f, foot]

E2NHDPlusV2_us variable name	NHDPlusV2.1 variable name	Description
ComID	ComID	Unique feature identifier
FL_ComID	ComID	Comid for flowline (neg=closed catch)
FL_GNIS_ID	GNIS_ID	Flowline GNIS identifier
FL_GNIS_Na	GNIS_Name	Stream GNIS name (<=65 char)
LENGTHKM	LENGTHKM	Flowline length (km)
REACHCODE	REACHCODE	Reach code
WB_Comid	WBAreaComI	Comid for waterbody
FL_FCode	FCode	Flowline feature code
nhdplusreg		NHDPlus region code
FL_FTypeAb		Flowline feature type abbreviation
IfWDig		If flowline in digitized network (0/1)
StreamLeve	StreamLeve	Strm lev1 (lowst=main,1=cst,4=grnd/brdr)
StreamOrde	StreamOrde	Stream order (1-10)
StreamCalc	StreamCalc	Major path stream order (1-10)
FromNode	FromNode	Original NHDPlusV2 from node
ToNode	ToNode	Original NHDPlusV2 to node
Hydroseq	Hydroseq	Hydro sequence (upstream order)
LevelPathI	LevelPathI	Levl path (hydsq most dnstrm FL in path)
TerminalPa	TerminalPa	Hydseq value of terminal FL in path
TotLngthKm	ArbolateSu	Total upstream flowline length (km)
Divergence	Divergence	Divergnce code (0=none,1=prime,2=second)
StartFlag	StartFlag	If flowline is headwater
TermFlag	TerminalFl	If terminate at coast/grnd/border
FromMeas	FromMeas	Reachcode measure at dnstrm node (0-100)
ToMeas	ToMeas	Reachcode measure at upstrm node (0-100)
IncAreaKm2	AreaSqKm	Incremental drainage area (km ²)
TotAreaKM2	TotDASqKM	Total upstream area (km ²)
CumAreaKm2	DivDASqKm	Cumulative div-based drainage area (km ²)
DivFrac	DivFrac	Diversion fraction
MaxElRawCm	MaxElevRaw	Maximum catchment raw elevation (cm)
MinElRawCm	MinElevRaw	Minimum catchment raw elevation (cm)
MaxElSmoCm	MaxElevSmo	Max catchment smoothed elevation (cm)
MinElSmoCm	MinElevSmo	Min catchment smoothed elevation (cm)
SLOPE	SLOPE	Channel slope (vert length/horiz length)
HWNodeAKm2	HwNodeSqKm	Area upstream of headwater node (km ²)
CatAreaKm2	AreaSqKm	Incremental catchment area (km ²)
IfSink		If closed catchment
PrecipV	PrecipV	Mean annual precipitation (mm*100)
TempV	TempV	Mean annual temperature (deg C*100)
RunOffV	RunOffV	Mean annual runoff (mm)

Table 4.1. Description of attributes in E2NHDPlusV2_us.—Continued

[km, kilometer; km², square kilometer; cm, centimeter; mm, millimeter; deg, degree; neg, negative value; deg C, degrees Celsius; cfs, cubic feet per second; char, characters; %, percent; f/s, foot per second; f, foot]

E2NHDPlusV2_us variable name	NHDPlusV2.1 variable name	Description
MAFlowUcfs	Q0001E	Cum mean annual flow (unit mthd,cfs)
MAVelUfps	V0001E	Stream velocity at mean ann. flow (f/s)
IncFlwUcfs	Qincr0001E	Incr catch mn ann flow (unit mthd,cfs)
WB_GNIS_ID	GNIS_ID	Waterbody GNIS identifier
WB_GNIS_Na	GNIS_Name	Waterbody GNIS name (<=65 char)
WB_AreaKm2	AreaSqKm	Waterbody area (km ²)
WB_ElevaFt	Elevation	Waterbody elevation (f)
WB_FCode	FCode	Waterbody feature code
WB_FTypeAb		Waterbody feature type abbreviation
ifarea		If NHDAREA feature (0/1)
alt_reg		WB feature alternative NHDPlusreg
cfromnode		Compactified from node
ctonode		compactified to node
termground		If all dwnstrm flowpaths term to ground
divunrslvd		If unresolved divrsion path
orig_CumAreaKm2		Pre-modification CumAreaKm2
inc_CumAreaKm2		Difference incremental area (km ²)
orig_MAFlowUcfs		Pre-modification MAFlowUcfs
inc_MAFlowUcfs		Difference incremental flow (cfs)
ModDivFrac		Modified value of diversion fraction
ModFDir		If modified to reverse flow direction
orig_divergence		Pre-modification divergence code
orig_divfrac		Pre-modification divergence fraction
orig_hydroseq		Pre-modification hydro sequence
incusarkm2		Amount incareakm2 in US (km ²)
CUMUSARKM2		Amount cumareakm2 in US (km ²)
PCTFORGNAR		Percent cumareakm2 not in US (%)
ifintterm		If flowline terminates to ground (0/1)
ifextterm		If flwline terminates at US border (0/1)
pctunrslvd		Indx non-div2grnd unrslvd drnage (0-100)
divmajorpa		If major path on non-trmgrnd unrslvd div
div2ground		If termground w non-termground upstream
divindetrm		If non-trmgrnd & completes div2grnd loop
divgroup		Hydro-independent diversion group id
divfrslvd		If div-affected FL reslvd w mod divfrac

Appendix 5. Description of Selected Ancillary Geospatial Dataset Variables Assigned to the Catchments and Flowlines of NHDPlusV2.1

Table 5.1. Description of selected ancillary geospatial dataset variables assigned to the catchments of and flowlines of NHDPlusV2.1.

Theme	Dataset	Variable	Citation
Atmospheric Deposition	Community Multi-scale Air Quality (CMAQ) Wet Atmospheric Deposition	2012 Wet	Zhang, W., Capps, S.L., Hu, Y., Nenes, A., Napelenok, S.L., and Russell, A.G., 2012, Development of the high-order decoupled direct method in three dimensions for particulate matter: enabling advanced sensitivity analysis in air quality models: Geoscience Model Development, v. 5, no. 2, p. 355–368, https://doi.org/10.5194/gmd-5-355-2012 , 2012.
Atmospheric deposition	Community Multi-scale Air Quality (CMAQ) Dry Atmospheric Deposition	2012 Dry	Schwede, D.B., and Lear, G.G., 2014, A novel hybrid approach for estimating total deposition in the United States, U.S. Environmental Protection Agency Papers, accessed June 20, 2019, at http://digitalcommons.unl.edu/usepapapers/219
Bankful width/depth	Bankfull_width_area_depth	Bankful Width/Depth	Wieczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2019, Select Attributes for NHDPlus Version 2.1 Reach Catchments and Modified Network Routed Upstream Watersheds for the Conterminous United States (ver 2.0, October 2019): U.S. Geological Survey data release, https://doi.org/10.5066/F7765D7V ; Bankfull Hydraulic Geometry Related to Physiographic Divisions.
Climate	Parameter-elevation Regressions on Independent Slopes Model (PRISM)	Mean Temperature and Precipitation, 2000–2014	PRISM, PRISM Group at Oregon State University, http://www.prism.oregonstate.edu/
Climate	Parameter-elevation Regressions on Independent Slopes Model (PRISM)	Average number of wet days (CAT_CWD)–30 year average	Easterling, D.R., Karl, T.R., Lawrimore, J.H., and Del Greco, S.A., 1999, United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data for 1871–1997, https://info.ornl.gov/sites/publications/Files/Pub57691.pdf
Climate	Water Balance Model Output	Actual Evapotranspiration, Potential evapotranspiration, Precipitation, Runoff, Snowfall, Soil Moisture, Temperature. 2000–2014, AVERAGE.	Wolock, D.M., and McCabe, G.J., 2018, Water Balance Model inputs and outputs for the conterminous United States, 1900–2015: U.S. Geological Survey data release, https://doi.org/10.5066/F71V5CWN .
Climate	Water Balance Model Output	Actual Evapotranspiration, Potential evapotranspiration, Precipitation, Runoff, Snowfall, Soil Moisture, Temperature. 2000–2014, detrended to 2012.	Wolock, D.M., and McCabe, G.J., 2018, Water Balance Model inputs and outputs for the conterminous United States, 1900–2015: U.S. Geological Survey data release, https://doi.org/10.5066/F71V5CWN .
Conservation Practices	National Agricultural Statistics Service (NASS)	CPRAC1: Land artificially drained by ditches 2012. CPRAC4: Land drained by tile 2012. CPRAC7: Land on which conservation tillage was used 2012. CPRAC10: Land on which conventional tillage was used 2012. CPRAC13: Land on which no-till practices were used 2012. CPRAC16: Land planted to a cover crop (excluding CRP) 2012. CPRAC19: Land under a conservation easement 2012.	U.S. Department of Agriculture, National Agricultural Statistics Service, 2016, Land Use Practices Data at the National, State, and County Level, accessed June 2017 at https://www.nass.usda.gov/Publications/AgCensus/2012/ .

Table 5.1. Description of selected ancillary geospatial dataset variables assigned to the catchments of and flowlines of NHDPlusV2.1.
—Continued

Theme	Dataset	Variable	Citation
Conservation Practices	National Resources Inventory (NRI)	Grouped practices identified as improving or worsening water quality	Wieczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2019, Select Attributes for NHDPlus Version 2.1 Reach Catchments and Modified Network Routed Upstream Watersheds for the Conterminous United States (ver. 2.0, October 2019): U.S. Geological Survey data release, https://doi.org/10.5066/F7765D7V ; Percent of Agriculture Land Subject to select national resource inventory conservation practices, 1997–2012.
Conservation Practices	Irrigated Agriculture Dataset for the United States (MIrAD-US)	2012 Irrigation	Brown, J.F., and Pervez, M.S., 2014, Merging remote sensing data and national agricultural statistics to model change in irrigated agriculture: Agricultural Systems, accessed May 17, 2016, at doi:10.1016/j.agry.2014.01.004. Data Accessed May 17, 2016, at https://earlywarning.usgs.gov/USirrigation
Conservation Practices	National Resources Inventory (NRI)	1992 Tile Drainage	Nakagaki, N., Wieczorek, M.E., and Qi, S.L., 2016, Estimates of subsurface tile drainage extent for the conterminous United States, early 1990s: U.S. Geological Survey data release, http://dx.doi.org/10.5066/F7RB72QS .
Conservation Practices	U.S. Department of Agriculture, Farm Service Agency (FSA)	2012 Percent Farm land in Conservation Reserve Program (CRP) (fraction of CRP acres/total cropland)	United States Department of Agriculture, Farm Service Agency, 2012, Conservation Reserve Program Statistics; CRP Practice Acres by County, accessed March 9, 2018, at https://www.fsa.usda.gov/programs-and-services/conservation-programs/reports-and-statistics/conservation-reserve-program-statistics/index
Conservation Practices	NRCS CEAP model output	LOSS in TP, TN, and Sediment due to	U.S. Department of Agriculture, Natural Resources Conservation Service, Resource Assessment Division, Conservation Effects Assessment Project (CEAP)—Cropland, 2003–06 Farmer Survey CEAP-1 APEX output at HUC-4 resolution, conterminous U.S., July 2016.
Distance to Streams	NHDPlus Flow Direction	Length	Wieczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2018. Select attributes for NHDPlus Version 2.1 reach catchments and modified network routed upstream watersheds for the conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/F7765D7V . Mean Flow Path Distance to NHDPlus Version 2 Streams.
Enhanced Vegetation Index (EVI)	Moderate Resolution Imaging Spectroradiometer (MODIS)	2012 Seasonal greenness	Vermote, E., 2015, MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006 [MOD09A1]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MOD09A1.006.
Fertilizer	U.S. Geological Survey Estimates of Fertilizer Use.	2012 N and P, Conditional and Unconditional	Stewart, J.S., Brakebill, J.W., Schwarz, G.E., and Preston, S.D., 2018, Dataset used for estimating catchment-level nitrogen and phosphorus fertilizer use from commercial fertilizer sales data for the conterminous U.S., 2012. U.S. Geological Survey data release, https://doi.org/10.5066/F73J3C50
Fertilizer	U.S. Geological Survey Estimates of County Fertilizer Use.	2012 N and P	Brakebill, J.W., and Gronberg, J.M., 2017, County-level estimates of nitrogen and phosphorus from commercial fertilizer for the conterminous United States, 1987–2012: U.S. Geological Survey data release, https://doi.org/10.5066/F7H41PKX .

Table 5.1. Description of selected ancillary geospatial dataset variables assigned to the catchments of and flowlines of NHDPlusV2.1.
—Continued

Theme	Dataset	Variable	Citation
Geology	Surficial Material of the US.	All Classes	Soller, D.R., Reheis, M.C., Garrity, C.P., and Van Sistine, D.R., 2009, Map database for surficial materials in the conterminous United States: U.S. Geological Survey Data Series 425, scale 1:5,000,000 [https://pubs.usgs.gov/ds/425/].
Geology	Lithology	Lithologic classes	Anning, D.W., and Ator, S.W., 2017, Generalized lithology for the conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/F7R78D4N .
Hydrologic	Base Flow Index (bfi48grd)	Value (base flow index value as percent)	Wolock, D.M., 2003, Base-flow index grid for the conterminous United States. U.S. Geological Survey digital data, accessed December 21, 2015, at http://water.usgs.gov/lookup/getspatial?bfi48grd .
Land Use/ Cover	National Land Cover Database (NLCD)	All Land Classes	Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information: Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p. 345–354, accessed April 22, 2015, at https://www.mrlc.gov/nlcd2011.php .
Land Use/ Cover	Cropland Data Layer (CDL)	All Land Classes	USDA National Agricultural Statistics Service Cropland Data Layer, 2012, Published crop-specific data layer [Online], accessed April 22, 2015, at https://nassgeodata.gmu.edu/CropScape/
Land Use/ Cover	U.S. Conterminous Wall-to-Wall Anthropogenic Land Use Trends (NWALT), 1974–2012	All Land Classes	Falcone, J.A., 2015, U.S. conterminous wall-to-wall anthropogenic land use trends (NWALT), 1974–2012: U.S. Geological Survey Data Series 948, 33 p. plus appendixes 3–6 as separate files, accessed February 12, 2015, at http://dx.doi.org/10.3133/ds948
Land Use/ Cover	National Land Cover Database (NLCD) Percent Impervious	Percent Impervious	Xian, G., Homer, C., Dewitz, J., Fry, J., Hossain, N., and Wickham, J., 2011, The change of impervious surface area between 2001 and 2006 in the conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 77, no. 8, p. 758–762, accessed December 21, 2015, at https://www.mrlc.gov/nlcd11_data.php .
Land Use/ Cover	U.S. Geological Survey Landslide Hazard	Landslide susceptibility	Godt, J.W., Landslide Hazards: U.S. Geological Survey, accessed April 30, 2014, at https://landslides.usgs.gov/hazards/nationalmap/
Land Use/ Cover	National Land Cover Database (NLCD) Tree Canopy	Percent canopy in 100m buffer of stream network	Tipton, J., Moisen, G., Patterson, P., Jackson, T.A., and Coulston, J., 2012, Sampling intensity and normalizations: Exploring cost-driving factors in nationwide mapping of tree canopy cover, <i>in</i> McWilliams, W., and Roesch, F.A., eds., 2012, Monitoring Across Borders: 2010 Joint Meeting of the Forest Inventory and Analysis (FIA) Symposium and the Southern Mensurationists. e-Gen. Tech. Rep. SRS-157. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: p. 201–208.

Table 5.1. Description of selected ancillary geospatial dataset variables assigned to the catchments of and flowlines of NHDPlusV2.1.—Continued

Theme	Dataset	Variable	Citation
Manure	U.S. Geological Survey County Estimates of Manure	2012 N and P	Gronberg, J.M., and Arnold, T.L., 2017, County-level estimates of nitrogen and phosphorus from animal manure for the conterminous United States, 2007 and 2012: U.S. Geological Survey Open-File Report 2017–1021, 6 p. [Also available at https://doi.org/10.3133/ofr20171021 .]
Municipal Discharge	U.S. Environmental Protection Agency (EPA) Integrated Compliance Information System (ICIS) and Permit Compliance System (PCS).	2012 N, P, flow	Skinner, K.D., and Wise, D.R., 2018, Point-source nutrient loads to streams of the conterminous United States, 2012: U.S. Geological Survey data release, https://doi.org/10.5066/P9PYVPFT .
Natural Phosphorus	U.S. Geological Survey Estimates of Natural Phosphorus (Pmapnatl)	Bed Sediment Phosphorus	Nardi, M.R., 2014, Watershed Potential to Contribute phosphorus from geologic materials to receiving streams, conterminous United States. U.S. Geological Survey, accessed June 20, 2017, at https://water.usgs.gov/GIS/metadata/usgswrd/XML/pmapnatl.xml .
Natural Phosphorus	U.S. Geological Survey Estimates of Natural Phosphorus	Natural occurring phosphorous in 4 different soil horizons (A horizon, C horizon and A horizon, C horizon per Geologic Mapping Unit)	Terziotti, S., 2019, Distribution of phosphorus in soils and aggregated within geologic mapping units, conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/P918DF1E .
Physiography	U.S. Geological Survey Physiographic Divisions (Physio)	Physiographic Regions	Fenneman, N.M., and Johnson, D.W., 1946, Physiographic divisions of the conterminous U.S., accessed October 2, 2017, at https://water.usgs.gov/lookup/getspatial?physio
Population Census	U.S. Geological Survey Estimates of Housing Density	2000, 2010	Falcone, J.A., 2016, U.S. block-level population density rasters for 1990, 2000, and 2010: U.S. Geological Survey data release, http://dx.doi.org/10.5066/F74J0C6M .
Reservoir Information	NHDPlus, National Inventory Dams (NID), and National Land Cover Data (NLCD)	Surface area	Wieczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2018, Select attributes for NHDPlus Version 2.1 reach catchments and modified network routed upstream watersheds for the conterminous United States: U.S. Geological Survey data release, https://doi.org/10.5066/F7765D7V . Attributes for Surface Water Impoundments Used in NAWQA SPARROW Model
Reservoir Information	National Hydrologic Database Plus (NHDPlus)	Lake residence time	McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., “NHDPlus Version 2: User Guide”, 2012, accessed March 4, 2015, at https://s3.amazonaws.com/nhdplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf
Septic	U.S. Geological Survey Estimates of Populations on Septic (SepN_CONUS_bg_2010)	Estimate of nitrogen from septic	LaMotte, A.E., 2018, Estimated nitrogen from septic for the conterminous United States, 2010: U.S. Geological Survey data release, https://doi.org/10.5066/P9QTGSI7
Soil properties	State Soil Geographic Database (STATSGO)	Pct Clay, Sand, Silt, KFACT, grain size, permeability, rainfall runoff factor, soil thickness, depth to water table, available water capacity, mean salinity	Wolock, D.M., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 656, XX p., accessed May 4, 2008, at https://water.usgs.gov/lookup/getspatial?muid

Table 5.1. Description of selected ancillary geospatial dataset variables assigned to the catchments of and flowlines of NHDPlusV2.1.
—Continued

Theme	Dataset	Variable	Citation
Soil proper- ties	State Soil Geographic Database (STATSGO2)	Hydrologic Groups, Soil Ph	Soil Survey Staff, U.S. Department of Agriculture, Natural Resources Conservation Service, General Soil Map (STATSGO2), 2010, accessed October 14, 2010, at http://websoilsurvey.sc.egov.usda.gov/App/ WebSoilSurvey.aspx
Water use	U.S. Geological Survey County Estimates of Water Use	2010 surface water withdrawal, groundwater withdrawal, irrigation	Maupin, M.A., Kenny, J.F., Hutson, S.S., Lovelace, J.K., Barber, N.L., and Linsey, K.S., 2014, Estimated use of water in the United States in 2010: U.S. Geological Survey Circular 1405, 56 p. [Also available at https://dx.doi.org/10.3133/cir1405 .]
Water use	U.S. Geological Sur- vey Estimates of Population Served by Public Supply Water Use	Population Served	Price, C.V., and Maupin, M.A., 2014, Documentation for the U.S. Geological Survey Public-Supply Database (PSDB)—A database of permitted public-supply wells, surface-water intakes, and systems in the United States: U.S. Geological Survey Open-File Report 2014–1212, 22 p. [Also available at https://dx.doi.org/10.3133/ ofr20141212 .]
Water use	U.S. Geological Survey Estimates of Thermoelectric Water Use	Total Withdrawal	Diehl, T.H., Harris, M.A., Murphy, J.C., Hutson, S.S., and Ladd, D.E., 2013, Methods for estimating water con- sumption for thermoelectric power plants in the United States: U.S. Geological Survey Scientific Investigations Report 2013–5188, 78 p. [Also available at http://dx.doi.org/10.3133/sir20135188 .]
Watersheds	National Watershed Boundary Dataset (WBD)	HUC12	Moore, R.B., 2019, Crosswalk between NHDPlus V2.1 and its WBD Snapshot: U.S. Geological Survey data release, https://doi.org/10.5066/P9CFXHGt
Wildfire	U.S. Geological Survey Estimates of Areas Burned by Wildfire (Gaplf2011)	2000–2012	U.S. Geological Survey Gap Analysis Program, 2016, GAP/LANDFIRE National Terrestrial Ecosystems 2011, accessed December 22, 2017, at https://doi.org/10.5066/F7ZS2TM0 .

For additional information, contact:
NAWQA Science Team
U.S. Geological Survey
12201 Sunrise Valley Drive, MS 413
Reston, VA 20191-0002

Email: gs-w_opp_nawqa_science_team@usgs.gov

Publishing support provided by the
West Trenton and Pembroke Publishing Service Centers

