



# **Integrated Reporting Georeferencing Pilot Report**

**March 21, 2016**

## ACKNOWLEDGEMENTS

The U.S. Environmental Protection Agency (EPA) appreciates the time and effort that the Georeferencing Pilot Team and the state workgroup representatives put into this effort. Thank you to the following state representatives and contractors for participating in the discussions, providing feedback and data files for testing, or helping develop the methods and prototype. Without their combined contributions, this effort would not have been possible.

A team of contributors led by Wendy Reid, EPA Project Manager, wrote this report. This team included Tommy Dewald, Shera Reems, Deric Teasley, and Dwane Young, EPA Office of Water; and Tatyana DiMascio (ORISE participant).

The team would like to thank the following state representatives for participating in the Pilot Workgroup:

### State Pilot Workgroup Representatives

Alabama Department of Environmental Management – John Pate

Georgia Department of Natural Resources – Michael Basmajian, Susan Salter, Vicki Trent,

Hawaii Department of Health – Malie Beach-Smith

Kentucky Department for Environmental Protection – Scarlett Stapleton

Massachusetts Department of Environmental Protection – Thomas Dallaire

Maine Department of Environmental Protection – Susanne Meidel, Doug Suitor

Mississippi Department of Environmental Quality – Valerie Alley

Nevada Division of Environmental Protection – Dave Simpson

New Jersey Department of Environmental Protection – Leigh Lager

Pennsylvania Department of Environmental Protection – Gary Walters, Amy Williams

Texas Commission on Environmental Quality – Anne Rogers, Andrew Sullivan

Utah Department of Environmental Quality – Emilie Flemer, Mark Stanger

West Virginia Department of Environmental Protection – Chris Daugherty, James Laine

**Contractor Support:** INDUS Corporation, Horizon Systems Corporation, and RTI International

U.S. Environmental Protection Agency. 2016. *Integrated Reporting Georeferencing Pilot Report*. EPA 841-R-15-008. U.S. EPA, Office of Water, Washington, D.C. March 2016. <https://www.epa.gov/waterdata/water-quality-framework>

# CONTENTS

LIST OF TABLES.....	V
LIST OF FIGURES.....	V
DEFINITIONS FOR ACRONYMS AND TERMS .....	VI
<b>1. EXECUTIVE SUMMARY .....</b>	<b>1</b>
<b>2. INTRODUCTION .....</b>	<b>4</b>
<b>3. BACKGROUND.....</b>	<b>5</b>
PURSUING A COMMON SURFACE WATER GEOSPATIAL FRAMEWORK .....	5
CLEAN WATER ACT (CWA) REQUIREMENTS FOR INTEGRATED REPORT (IR) DATA SUBMISSION .....	5
TRADITIONAL IR DATA PROCESSING APPROACH .....	6
CHALLENGES IN THE TRADITIONAL IR GEOSPATIAL DATA PROCESSING APPROACH.....	7
<b>4. GOALS AND OBJECTIVES OF THE PILOT.....</b>	<b>8</b>
<b>5. STATE PARTICIPATION AND CONTRIBUTIONS.....</b>	<b>8</b>
UNDERSTANDING STATE DATA MANAGEMENT PRACTICES .....	9
Assessment Unit Delineation .....	9
Geospatial Practices .....	9
PILOT WORKGROUP PARTICIPANTS AND PILOT DATA.....	10
<b>6. INITIAL SET OF PILOT OPTIONS EXPLORED.....</b>	<b>11</b>
POINT-BASED INDEXING .....	11
WATERSHED POLYGON-BASED INDEXING .....	11
FEATURE-BASED INDEXING.....	11
<b>7. CATCHMENT-BASED INDEXING – RECOMMENDED PILOT OPTION .....</b>	<b>13</b>
CATCHMENT-BASED INDEXING.....	13
CATCHMENT-BASED INDEXING PROCESS.....	14
<b>8. RESULTS.....</b>	<b>18</b>
ISSUES ENCOUNTERED .....	18
EVALUATION CRITERIA .....	19
Costs .....	19
Timeliness .....	19
Data Quality and Use Cases .....	20
Completeness.....	22
<b>9. IMPLEMENTATION CONSIDERATIONS.....</b>	<b>23</b>
<b>10. RECOMMENDATIONS .....</b>	<b>23</b>
<b>11. REFERENCES .....</b>	<b>24</b>

<b>APPENDICES .....</b>	<b>25</b>
LIST OF TABLES .....	25
LIST OF FIGURES .....	25
APPENDIX A: BACKGROUND INFORMATION ON NHDPLUS .....	27
APPENDIX B: BACKGROUND INFORMATION ON NHDPLUS CATCHMENTS .....	28
Catchment Concepts .....	28
APPENDIX C: TECHNICAL SPECIFICATION FOR CATCHMENT-BASED INDEXING .....	30
Linear Event NHDPlusV2 (Linear-to-Catchment) Specification .....	30
Area Waterbody Event NHDPlusV2 (Waterbody-to-Catchment) Specification .....	43
HUC-Like Event NHDPlusV2 (HUC-to-Catchment) Specification .....	47
APPENDIX D: EPA IMPLEMENTATION CONSIDERATIONS .....	53
APPENDIX E: INFORMATION ON INITIAL PILOT OPTIONS .....	55
Point-based Indexing Details .....	55
Watershed Polygon-based Indexing Details .....	56
Feature-based Indexing Details .....	56
Elimination and Refinement of Options .....	56
APPENDIX F: QUESTIONS TO PILOT WORKGROUP REGARDING INITIAL PILOT OPTIONS .....	58

## LIST OF TABLES

Table 1: Summary of key use cases and whether the Catchment-based indexing approach meets their needs .....	20
---	----

## LIST OF FIGURES

Figure 1: This figure shows the standard processing of state assessment data. ....	6
Figure 2: This figure shows the evolution of the Pilot indexing options.....	13
Figure 3: This figure shows a simplified version of the Catchment-based indexing process for a linear Assessment Unit (e.g., a stream or river).....	14
Figure 4: The linear Catchment-based indexing process uses thresholds to exclude catchments that are assigned to an Assessment Unit due to a tiny piece of stream segment crossing into the adjacent catchment. ....	15
Figure 5: The area waterbody Catchment-based indexing process is based on three characteristics: the portion of the waterbody assessment unit that is located in the catchment; the portion of the catchment covered by the waterbody; or whether the catchment is associated with an NHD artificial path. ....	16
Figure 6: This figure shows the correspondence results between an area waterbody Assessment Unit (e.g., a lake) and the NHDPlus catchments.....	16
Figure 7: NHD artificial paths run through NHD waterbodies to connect the linear stream network flowing into and out of the lake.....	17
Figure 8: The HUC Catchment-based indexing process is based on two characteristics: the portion of the waterbody assessment unit that is located in the catchment; or the portion of the catchment covered by the waterbody. ....	17

## DEFINITIONS FOR ACRONYMS AND TERMS

Acronyms	
ATTAINS	Assessment, Total Maximum Daily Load (TMDL) Tracking and Implementation System
AU	Assessment Unit - A state-defined segment of waterbody that is determined for the purpose of tracking information on water quality conditions and restoration/protection efforts. For purposes of this Pilot, unless otherwise indicated, Assessment Unit and Event are synonymous. Assessment Unit is specific to the Clean Water Act Sections 303(d) and 305(b) programs, whereas Event is a more generic geospatial term.
CWA	Clean Water Act
EPA	United States Environmental Protection Agency
FTE	Full-Time Equivalent – The equivalent of a person working full time (40 hrs. per week) for a year
GIS	Geographic Information System
HEM Tool	Hydrography Event Management Tool
HEM EPA Add-on Tools	Tools to supplement the Hydrography Event Management Tool, for georeferencing EPA water program data
HUC	Hydrologic Unit Code
ID	Identifier – The unique alphanumeric code assigned to the Assessment Unit or feature
IR	Integrated Reporting – In 2002, EPA recommend that states submit an Integrated Report to fulfill reporting requirements under CWA Sections 305(b) and 303(d) regarding water quality assessment decisions and information on waters not supporting their water quality standards
LOE	Level of Effort
NHD	National Hydrography Dataset – a hydrologic reference layer. This term is sometimes used in conjunction with medium resolution or high resolution.
NHDPlus	National Hydrography Dataset <i>Plus</i> - Snapshot of the National Hydrography Dataset – a hydrologic reference layer in medium resolution (1:100,000), which includes additional attributes
NED	National Elevation Dataset
OW	The U.S. EPA Office of Water
PMO	Project Management Office
QA/QC	Quality Assurance / Quality Control
RAD	Reach Address Database – EPA Office of Water’s national geospatial repository
TMDL	Total Maximum Daily Load - a calculation of the maximum amount of a pollutant that a waterbody can receive and still meet water quality standards, and an allocation of that load among the various sources of that pollutant. [Source: <a href="http://www.epa.gov/tmdl">http://www.epa.gov/tmdl</a> ]
USGS	United States Geological Survey
VAA	Value Added Attributes that are part of NHDPlus
WBD	Watershed Boundary Dataset

Terms	
Assessment Unit (AU)	A state-defined segment of waterbody that is determined for the purpose of tracking information on water quality conditions and restoration/protection efforts. For purposes of this Pilot, unless otherwise indicated, Assessment Unit and Event are synonymous. Assessment Unit is specific to the Clean Water Act Sections 303(d) and 305(b) programs, whereas Event is a more generic geospatial term.
Attribute data	Information that describes the water quality status for each Assessment Unit, such as Assessment Unit ID
Catchment	The land surface that drains to each stream segment. Essentially these are small hydrologic management units that average about 1.1 square miles in size. Catchments are included in NHDPlus.
Event	A geospatial data record representing the geographic location of a natural or constructed feature, boundary, or the location of an occurrence in time. For purposes of this Pilot, unless otherwise indicated, Event is synonymous with Assessment Unit (AU). Assessment Unit is specific to the Clean Water Act Sections 303(d) and 305(b) programs, whereas Event is a more generic geospatial term.
Geospatial data	Geospatial data are information that identify the geographic location and characteristics of natural or constructed features and boundaries on the earth, typically represented by points, lines, polygons, and/or complex geographic features. This includes original and interpreted geospatial data, such as those derived through remote sensing including, but not limited to, images and raster data sets, aerial photographs, and other forms of geospatial data or data sets in both digitized and non-digitized forms. [Source: EPA's National Geospatial Data Policy, August 2005]
High Resolution GIS data	Geospatial data at a scale of 1:24,000 or better. This term is often used in conjunction with the NHD
Hydrologic Unit Code	A sequence of numbers or letters that identify a hydrologic feature such as a river, river reach, lake, drainage basin
Indexing	Also referred to as reach indexing. Georeferencing to NHD or NHDPlus
Level Path	A sequence of transport reaches that traces the main stem for a given flow of water [Source: <a href="http://nhd.usgs.gov/chapter1/chp1_data_users_guide.pdf">http://nhd.usgs.gov/chapter1/chp1_data_users_guide.pdf</a> ]
Medium Resolution	Geospatial data at a scale of 1:100,000. This term is often used in conjunction with the NHD and NHDPlus
NHD Event	A segment along NHD network
Non-NHD Event	A segment that is not located on NHD network
Stream Level	Identifies the main path to which a transport reach belongs [Source: <a href="http://nhd.usgs.gov/chapter1/chp1_data_users_guide.pdf">http://nhd.usgs.gov/chapter1/chp1_data_users_guide.pdf</a> ]
Water Quality Standards	Water Quality Standards define the goals for a waterbody by designating its uses, setting criteria to protect those uses, and establishing provisions such as antidegradation policies to protect waterbodies from pollutants. [Source: <a href="http://www.epa.gov/standards-water-body-health">http://www.epa.gov/standards-water-body-health</a> ]

This page intentionally left blank



# 1. EXECUTIVE SUMMARY

EPA, in collaboration with states, initiated an Integrated Reporting (IR) Georeferencing Pilot (Pilot) to investigate alternate ways for states to submit--and EPA to process--geospatial data related to Clean Water Act Sections 303(d) and 305(b) more efficiently, while maintaining or improving geospatial data availability and usefulness. The Pilot focused on four goals: 1) reduce EPA processing costs and state reporting burden for geospatial data, 2) publish geospatial data in a timelier manner, 3) maintain geospatial data quality to meet the requirements of the users, and 4) have a more complete national geospatial dataset. The Pilot examined technical issues and did not address programmatic issues. The Pilot did not focus on generation or submission to EPA of water quality assessment decisions (attribute data).

The Pilot primarily focused on evaluating more automated approaches to simplify the traditional highly manual IR geospatial data processing. EPA has customarily used a snapshot of the medium resolution (1:100,000-scale) National Hydrography Dataset (NHD) as the reference hydrography layer. This snapshot has been stored in the National Hydrography Dataset Plus (NHDPlus), along with additional attributes such as Strahler Stream Order, Level Path, and Time of Travel<sup>1</sup> which can be used to navigate upstream and downstream in the stream network. While EPA has been working with medium resolution geospatial data, more states have been submitting high resolution (1:24,000-scale or better) geospatial data to EPA. Thus, the Pilot also explored ways to increase direct use of high resolution geospatial data. Furthermore, the Pilot took into consideration states' capacity to process and submit geospatial data as they have limited geospatial staff and funding resources. The Pilot evaluated automated procedures that use NHDPlus catchments (essentially, small local drainage areas surrounding each medium resolution NHD stream segment and which average about 1.1 square miles in size) as a national *geofabric* to associate the mixed resolution state-submitted features to each other. Once these feature-to-catchment associations are established, existing NHDPlus-based tools and web-services can be used to support display, analysis, and reporting of both high and medium resolution NHD features and any other state-submitted geospatial features related to water quality assessment decision data.

The initial Pilot design explored relating the following three types of state-submitted inputs to NHD features: Points (based on monitoring locations), Watershed Polygons (based on hydrologic units (HUCs) that some states use as their Assessment Units<sup>2</sup>), and Features (based on lines representing streams or rivers and polygons representing lakes or reservoirs). Pilot states were not interested in Point-based indexing because they would need to reformat their own geospatial data to use this option. In addition, Watershed-based indexing received only minimal support from Pilot states because most states do not manage their water quality assessment data in that format. Both of these options would increase states' costs and require more time to format their geospatial data for any states who were not currently managing their geospatial data in that format. The Feature-based indexing option received the most support as it offered flexibility for states to submit their data in any resolution or format. However, the results of testing of the Feature-based indexing option revealed that it was too computer resource and time intensive for implementation. Based upon these findings, the fourth option, Catchment-based indexing, was explored and a prototype was developed for Esri's ArcMap 10.1 software platform.

The Catchment-based indexing approach simplified and refined the functions of the Feature-based indexing and Watershed Polygon-based indexing options. The Feature-based indexing option was too time intensive because it tried to compare state data to individual hydrologic network features (e.g., stream or lake) in both the high resolution NHD and the medium resolution NHDPlus and make an automatic conversion at the feature level. The Watershed Polygon-based indexing option focused on taking a state watershed polygon and extracting the underlying hydrologic network features from the medium resolution NHDPlus, so it took a state watershed

---

<sup>1</sup> Time of Travel is calculated using the Extended Unit Runoff Method.

<sup>2</sup> Assessment Unit - A state-defined segment of waterbody that is delineated for the purpose of tracking information on water quality conditions and restoration/protection efforts

polygon and converted it to linear features. However, the Catchment-based indexing approach works differently: rather than converting individual geospatial features, the Catchment-based indexing approach develops a crosswalk table between the state feature (e.g., stream, lake, or HUC) and the corresponding NHDPlus catchments. In that way, Catchment-based indexing does not require a conversion of the state geospatial data; rather, it develops a correspondence between the NHDPlus catchments and the state data such that both datasets can be used in conjunction with one another through database relationships. The Catchment-based indexing approach starts with a geospatial overlay of NHDPlus catchments onto any state geospatial input, and utilizes the logic built into the NHDPlus to refine the results. The output from the Catchment-based indexing approach is a crosswalk table detailing which state geospatial features correspond with which NHDPlus catchments. The catchments are used for analysis and the state geospatial data are used for display to show the location of the features within the catchments.

The Catchment-based indexing approach was tested on a broad range of data files from 21 states, representing streams, lakes, and HUCs, resulting in a total of 64 files run through the prototype. It took approximately 30 hours to process all 64 of the state files through the automated portion of the Catchment-based indexing process. Prior to processing, each file was manually converted into a standard database structure. Automated processing time for each file ranged from a few minutes to almost three hours, and the mean time to process a file through the prototype was less than half an hour. If this approach is implemented for processing data in the future, the output for each state will also need to undergo a manual QA/QC process to ensure that the automated process output accurately reflects the state data, which will add to the total processing time. It is estimated that this manual QA/QC of the output files will range from 5-10 hours per state. The total time to process a state's geospatial data (manual preprocessing time plus automated processing time plus manual QA/QC time) is estimated to range from 7-15 hours per state. This is a significant improvement over the traditional georeferencing process which ranges between 20-100 hours per state, reducing both EPA costs and the processing time required.

A variety of use cases for geospatial data representing assessed and impaired waters were considered during the Pilot. The Catchment-based indexing approach was evaluated to determine whether it satisfied the collection of typical use cases for all primary and most secondary customers for IR geospatial data. Based on this evaluation, the output from the Catchment-based indexing approach meets the collective needs of all use cases considered. Details on the use cases considered are provided in the Results section.

Another goal of the Pilot was to provide a more complete IR national geospatial coverage by examining ways to increase the number of states present in EPA's national geospatial datasets each reporting cycle. Completeness depends on the states' willingness to provide geospatial data to EPA, as well as EPA's ability to process the received files in a timely manner. EPA plans to provide the original state geospatial data for display in conjunction with the catchment representation. Therefore, states may be more willing to provide geospatial data to EPA due to the more hydrographically-representative portrayal of their data for visualization purposes, which can be used in conjunction with the upstream and downstream search capabilities provided through the NHDPlus catchments.

The Catchment-based indexing approach described in this document represents a new georeferencing paradigm. Historically, the IR geospatial data have been used for both display and programmatic purposes. Under this new option, these two types of uses would be addressed using separate approaches. Programmatic uses (such as cross-program analysis, upstream/downstream analysis, and programmatic measures) would use the Catchment-based indexing approach described in this report. Display purposes would be met with geospatial files provided by the states, without converting them to the medium resolution hydrographic network features in NHDPlus.

The implications for states should be minimal since the new approach will accept any existing state geospatial data as the input. This new approach assumes that states have some type of geospatial representation of their Assessment Units, and that streams are represented by lines and lakes are represented by polygons. At EPA, the

existing Office of Water geospatial software and database infrastructure will need to be examined in light of any new requirements presented by Catchment-based indexing, including the retention of state-submitted geospatial data for display in conjunction with their Catchment-based representation.

There were a couple of input types that were not explored or addressed during this Pilot. While this Pilot did not specifically address point features such as monitoring stations, the Catchment-based indexing approach should be easily adaptable to accommodate that input type. In addition, this Pilot did not address coastal areas where NHDPlus catchments do not exist. A separate strategy that handles coastal areas where catchments do not exist would need to be incorporated into the Catchment-based indexing approach going forward.

This Pilot was a research and development project that analyzed new approaches for georeferencing state water quality assessment decision data. Although the Catchment-based indexing approach does not address all of the issues related to the existing process for the submission and processing of state data (e.g., it does not resolve the time it takes states to create the original geospatial files for their purposes, or the time it takes states to submit the data, or Regional approval of that data in a timely fashion), this option is the most efficient approach among the options tested, leading to reduced processing costs and processing time. It effectively addresses all use cases considered. The Catchment-based indexing approach should also increase a state's ability and willingness to submit their geospatial data to EPA. Furthermore, this approach should address some, but not all, of the outstanding data processing issues, while maintaining data accuracy sufficient to meet the needs of users. Lastly, it includes the use of state-submitted geospatial data for display and visualization purposes in conjunction with the catchment representation and also leverages inherent NHDPlus capabilities such as an upstream or downstream search. For these reasons, the Pilot team recommends the Catchment-based indexing approach to EPA for future state geospatial data processing.<sup>3</sup>

---

<sup>3</sup> EPA is currently working with Alaska, for which NHDPlus catchments are not yet developed, and discussions are underway to determine an interim solution.

## 2. INTRODUCTION

Through the biennial Clean Water Act (CWA) Sections 303(d) and 305(b) Integrated Report, the EPA Office of Water (OW) compiles attribute and geospatial data submitted by states that describe their water quality assessment and impairment decisions. The attribute data are stored in the EPA Assessment, Total Maximum Daily Load (TMDL) Tracking and Implementation System (ATTAINS).<sup>4</sup> These data include the water quality standards assessment decisions made by states for designated use support, causes of impairment, probable sources contributing to impaired waters, and other relevant information, such as the Assessment Unit ID (identifier) and size (length or areal extent). The geospatial data are stored in the EPA Watershed Assessment Tracking and Environmental Results (WATERS) system.<sup>5</sup> These data enable visualization of assessment and impairment decisions using maps and provide a key mechanism to track progress in the restoration of impaired waters, support national analyses of waterbody conditions, and facilitate communication and decision making among EPA, states, the public, and others. A couple of the primary goals for ATTAINS include providing snapshots of water quality reporting at the national, regional, state, watershed, or local levels regarding status of actions to address priority waters, as well as responding to specific inquiries about assessments and impairments. ATTAINS supports communication of information related to CWA programs through tools such as the How's My Waterway App,<sup>6</sup> and the WATERS framework. These tools are available through the EPA Water Data and Tools website.<sup>7</sup>

EPA assists states in their submission of 303(d) and 305(b) datasets for inclusion in ATTAINS. The effectiveness of ATTAINS depends on several factors: the efficiency and timeliness of the data submission process, the quality of data submitted by states to EPA, and, in turn, the accuracy of the representation of water quality conditions produced by EPA from the state-submitted data.

The traditional EPA process for integrating state geospatial data for water quality assessment and impairment decisions has been in place for well over a decade. It is time-consuming, labor intensive, and has not consistently delivered complete coverage of state-submitted data. In addition, it does not fully leverage the state-submitted geospatial data which is increasingly more detailed. Facing more limited state and Federal resources, EPA seeks efficiencies and improved methods to process state georeferenced water quality assessment and impairment decisions at the federal level.

In collaboration with state partners, EPA initiated an Integrated Reporting Georeferencing Pilot (Pilot) to investigate alternate ways for states to submit, and EPA to process, geospatial data more efficiently, while maintaining or improving geospatial data availability and usefulness. The main goal of the Pilot was to investigate automated approaches to simplify and expedite the traditional geospatial data processing. The Pilot also was intended to explore more effective ways to process state-submitted high resolution geospatial data, while continuing to leverage the robust display, analysis and reporting capabilities that exist at EPA. Finally, the Pilot took into consideration states' capacity to process and submit data based upon their already limited geospatial staff and funding resources.

This document first provides background information and challenges for the traditional process followed by a description of Pilot goals and state participation. The report then discusses the original Pilot options and how those options evolved into the Catchment-based indexing approach. The Catchment-based indexing results are then presented, followed by a summary and recommendations. The appendices provide additional background

---

<sup>4</sup> ATTAINS:

<http://www.epa.gov/waterdata/assessment-and-total-maximum-daily-load-tracking-and-implementation-system-attains>

<sup>5</sup> WATERS: <http://www.epa.gov/waterdata/waters-watershed-assessment-tracking-environmental-results-system>

<sup>6</sup> How's My Waterway App: <http://www.epa.gov/waterdata/how-s-my-waterway>

<sup>7</sup> Water Data and Tools: <http://www.epa.gov/waterdata>

information on NHDPlus and catchments, technical details of the Catchment-based indexing approach, and more information on the original Pilot options.

### **3. BACKGROUND**

#### **Pursuing a Common Surface Water Geospatial Framework**

In the early 1980s, EPA's Office of Water first envisioned a common surface water geospatial framework to support its display, analysis, and reporting needs. The framework was intended to promote consistency and data integration across states and between EPA water programs. The first national digital stream network database, known as Reach File Version 1 (RF1) at 1:500,000-scale, was developed in the mid-1980s and successfully demonstrated the framework concept. In the late 1990s, the Office of Water, the US Geological Survey (USGS) and state cooperators collaborated to produce the initial National Hydrography Dataset (NHD) at 1:100,000-scale (known as medium resolution NHD). In 2006, the Office of Water and USGS first released NHDPlus, which enhanced the capabilities of medium resolution NHD by integrating its stream network with the landscape by defining the local drainage area (catchment) for each stream segment. NHDPlus also provided many additional attributes (e.g., stream order, flow volume and velocity, temperature, and precipitation) and analytical capabilities. Lastly, a continued interest in a more-detailed stream network led to the 2007 completion of high resolution NHD at 1:24,000-scale or better by USGS, the US Forest Service, and many states, which is a mixed resolution dataset based on the most-detailed data available. (Appendix A provides more information on NHDPlus.)

To date, the Office of Water has employed the medium resolution NHD to support its programmatic needs, which include the georeferencing of water quality Assessment Units and impairments. As state use of more-detailed stream networks (predominantly, high resolution NHD) has grown, EPA has found it to be increasingly challenging to relate state-submitted stream features directly to the corresponding medium resolution NHD stream features. Furthermore, the state-submitted stream features have only been used to support georeferencing to the medium resolution NHD and then set aside. The medium resolution features are then used for display, analysis, and reporting.

#### **Clean Water Act (CWA) Requirements for Integrated Report (IR) Data Submission**

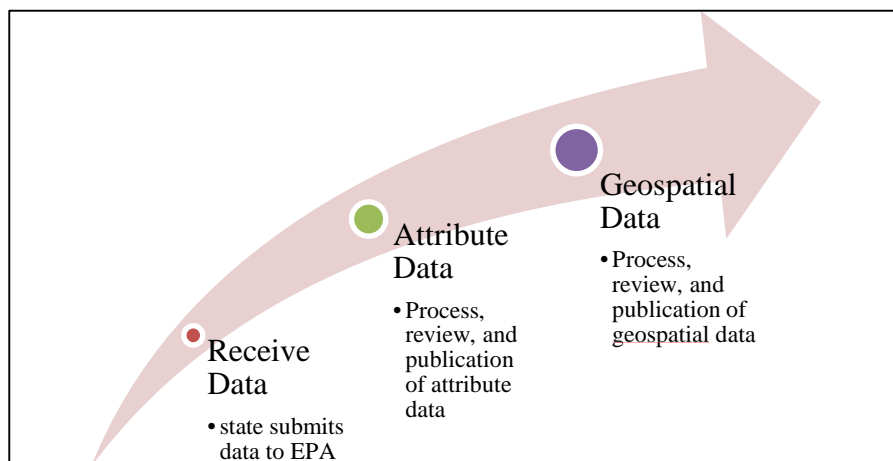
Under CWA Section 303(d), states are required to submit to EPA a list of impaired and threatened waters still needing a Total Maximum Daily Load(s)—a calculation of the maximum amount of a pollutant that a waterbody can receive and still meet water quality standards, and an allocation of that load among the various sources of that pollutant. The supporting regulation (40 CFR 130.7) requires states to submit this information to EPA on April 1 of every even numbered year. Under CWA Section 305(b) and its supporting regulation (40 CFR 130.8), states are required to report to EPA on the status of the Nation's waters on April 1 of every even numbered year. Beginning with the 2002 reporting cycle, EPA recommended that states combine their Section 303(d) and 305(b) submissions into a single Integrated Report (IR).<sup>8</sup>

---

<sup>8</sup> The 303(d), 305(b) and 314 Integrated Reporting Guidance for 2002 through the current reporting cycle are available from this website: <http://www.epa.gov/tmdl/identifying-and-listing-impaired-waters>

## Traditional IR Data Processing Approach

In the 2006 Integrated Reporting guidance, EPA began requesting that states submit the associated attribute data (i.e., water quality assessment and impairment decisions, such as Assessment Unit ID, designated uses, whether the water is supporting or not supporting the water quality standard, and any causes of impairment) and geospatial data (i.e., representation of the location and extent of their Assessment Units related to their water quality assessment and impairment decisions) as part of a state's Integrated Report submission. Currently, EPA accepts a number of different attribute and geospatial data formats as listed in the 2014 Integrated Reporting memo. To help track Assessment Unit changes from one reporting cycle to the next, EPA recommends that states provide a crosswalk between the current and past Assessment Unit IDs when any ID changes have occurred. Assessment Units can refer to linear water features (e.g., streams, rivers or coastlines), area water features (e.g., lakes, ponds or reservoirs), or watersheds (e.g., Hydrologic Unit Codes [HUCs] used in the Watershed Boundary Dataset [WBD]).



**Figure 1: This figure shows the standard processing of state assessment data.**

Figure 1 shows the general steps involved in the standard processing of IR data. Upon receipt of attribute and geospatial data, EPA processes the information in two steps. The first step focuses on processing the attribute data. Once the attribute data have been reviewed and released for publication, work begins on the second step, which focuses on processing the geospatial data. For this piece, EPA uses the Hydrography Event Management (HEM) Tool<sup>9</sup> and the supplemental HEM EPA Add-On Tools to process and QA/QC the state submissions.<sup>10</sup>

EPA uses the NHD layer from the medium resolution NHDPlus as its georeferencing framework. If the geospatial data submitted by a state are not from medium resolution NHD, they are converted to the NHDPlus network. When an Assessment Unit lies on a small stream not found in the medium resolution NHDPlus stream network, the point where that Assessment Unit flows into the NHDPlus network serves as the surrogate location for the Assessment Unit. A check is then performed to confirm that all of the Assessment Unit IDs present in the attribute data are included in the resulting geospatial file. Any missing Assessment Unit IDs are investigated during a subsequent review process. The geospatial files are then posted to a password-protected review website, and the states and Regions are given the opportunity to review the geospatial data for accuracy. Once the states and Regions have reviewed and cleared the data for release to the public, the geospatial data files are published to the EPA WATERS geospatial downloads<sup>11</sup> where they can be viewed and downloaded.

<sup>9</sup> The Hydrography Event Management (HEM) Tool and the supplemental HEM EPA Add-on Tools are available from USGS: <http://nhd.usgs.gov/tools.html>

<sup>10</sup> EPA contractors benefitted from the increased efficiency in data management provided by the HEM and EPA Add-on Tools which accommodate processing data at the state level as compared to the previous tool (Reach Indexing Tool for ArcView 3.x) which worked at a HUC8 level.

<sup>11</sup>EPA Office of Water's Geospatial Data Downloads: <http://www.epa.gov/waterdata/waters-geospatial-data-downloads>

## Challenges in the Traditional IR Geospatial Data Processing Approach

The delays before publication of geospatial data can be lengthy due to the programmatic and technical steps involved leading up to the submission and processing of the data.<sup>12</sup> The state must submit, and the EPA Region must approve, the state CWA Section 303(d) list before the water quality assessment decision data (attribute data) are submitted, processed, reviewed, and published. Once the attribute data are published, work on the geospatial data begins.

The traditional geospatial process is primarily manual. It is costly and can take many months to process and review the work, depending upon the complexity of the state input data and the number of states' data in the queue. The delays in publication of the geospatial data mean that the attribute data may be published sooner without the corresponding geospatial location information being available to the public at the same time. In addition, the limited funds available mean that some of the state geospatial data may not get processed for a reporting cycle. Thus, the users who are applying geospatial tools for analyses may not have access to the most complete or recent information in a timely manner.

EPA uses the NHD layer from the medium resolution NHDPlus as its georeferencing framework because of its national consistency. Also, EPA has developed a robust suite of display, analysis and reporting tools that access these data within WATERS.<sup>13</sup> The tools include upstream and downstream queries, the ability to delineate a watershed anywhere in the country along with its characteristics, as well as numerous related web services. At the same time, states are more frequently using higher resolution geospatial data to support their programs than in the past. As a result, the geospatial data that EPA receives from states are a mixture of resolutions that EPA must then standardize in a reasonable, cost-effective manner. The association of these state-submitted geospatial data with individual medium resolution features is becoming increasingly challenging.

---

<sup>12</sup> In 2013, the EPA completed the effort, "Reducing Reporting Burden under Clean Water Act Sections 303(d) and 305(b)." Through this effort, states identified several steps in the Integrated Report process that are a significant burden, which included: 1) state review and use of available data to make assessment decisions, 2) state preparation of data and associated geospatial information and entry into an assessment database, 3) state preparation and submission of final 303(d) lists and 305(b) reports to EPA, and other relevant documentation, 4) state preparation or refinement of its assessment and listing methodology, and 5) state response to public comments. See <http://www.epa.gov/tmdl/identifying-and-listing-impaired-waters> for the full report. The Integrated Report Georeferencing Pilot was one effort that responded to the "state preparation of data and associated geospatial information..." burden identified by states and was designed to address the technical issues that EPA and states face when preparing, submitting, processing, and displaying geospatial data. This Pilot was not designed to focus on the programmatic issues of the CWA 303(d) and 305(b) data submission process, such as overlapping Assessment Units, defining Assessment Units, use of probability survey data in lieu of targeted monitoring data for 305(b), delays in data submissions, re-segmentation, and limited resources for training and monitoring. Therefore, this document will not discuss these issues; however, EPA does recognize that they are a critical component of the overall discussion on improving reporting and reducing burden under CWA Sections 303(d) and 305(b). As such, the EPA initiated a series of workgroups to respond to the recommendations from the above mentioned effort, which included: identifying the data elements that states should track and report on to EPA, developing tools that reduce state and EPA burden in preparing, submitting, and reviewing data, identifying better measures to show success in restoring waters, and developing tools to assist states in the analysis of water quality data compared with water quality standards to make Assessment Unit decisions.

<sup>13</sup> For more information on these tools, please see:

<http://www.epa.gov/waterdata/waters-watershed-assessment-tracking-environmental-results-system>

## 4. GOALS AND OBJECTIVES OF THE PILOT

The purpose of the Georeferencing Pilot was to investigate opportunities for simplifying the traditional approach of processing state geospatial data by EPA, while maintaining high quality datasets with a focus on four goals. Each goal is followed by a suite of questions related to the goal, but these questions do not represent the full universe of questions that may be considered.

1. Costs – Level of investment should be appropriate for accuracy of the state’s Assessment Unit characterization.
  - Are there potential cost reductions for EPA to process state geospatial data without significantly reducing quality of data representing state’s Assessment Units?
  - Are there potential cost reductions to states that might increase state willingness to submit data and increase efficiencies for state Level of Effort (LOE)?
2. Timeliness – Display geospatial data concurrently with attribute data
  - Are there opportunities to reduce the time delay associated with the geospatial data processing and its delivery to the public and stakeholders?
  - Are there opportunities to reduce the time delay associated with geospatial data submission from the states?
3. Quality – Maintain geospatial data quality at the levels appropriate for intended uses
  - As the Pilot addresses other goals, such as costs and timeliness, is the data quality maintained at an acceptable level to support already established activities by the current users?
  - Are there opportunities to improve the level of accuracy with which state geospatial data are represented in the EPA’s national ATTAINS datasets?
  - Are there opportunities to improve EPA’s ability to track restoration of impaired waters?
4. Completeness – Increase the number of states that are present in the EPA’s national ATTAINS datasets each reporting cycle
  - Are there opportunities to increase state willingness to provide data to EPA?
  - Are there opportunities to address technical issues that have prevented EPA from processing certain state geospatial data for inclusion in the national ATTAINS datasets?
  - Are there opportunities to allow the use and display of geospatial data from additional resolutions beyond the medium resolution NHD/NHDPlus?

The Georeferencing Pilot focused on the increased use of automation—leveraging advances in technology and geospatial data—to meet these goals, and answer their associated questions.

## 5. STATE PARTICIPATION AND CONTRIBUTIONS

States had several opportunities to contribute to the Pilot, either by providing background information or participating more directly in the Pilot workgroup. After developing the initial Pilot concept, EPA recruited Regional and state partners during the fall of 2011. As part of a complementary, but separate project, EPA gathered information on state data management practices prior to the Pilot kickoff. The state input helped inform the Pilot goals and objectives. The Pilot kickoff with state participants occurred in the spring of 2012. Pilot participants helped to shape which options should be considered by indicating their willingness to support specific options and providing estimates for their level of effort to ultimately implement the options. The options considered in the Pilot are discussed in section 6.



## Understanding State Data Management Practices

In preparation for the Pilot design, EPA sought to better understand state CWA Sections 303(d) and 305(b) data management practices. EPA offered states the opportunity to provide information on how states delineate their Assessment Units, manage their 303(d) and 305(b) geospatial data, and manage their assessment decisions (attribute data). About two-thirds of 56 states and territories provided input.

### Assessment Unit Delineation

Assessment Unit questions focused on the process used to delineate the Assessment Unit boundaries, and the precision of the resulting boundaries. The Assessment Unit delineation questions confirmed that states consider many factors when delineating Assessment Units and that they place considerable emphasis on the precision of Assessment Unit boundaries.

One of the Assessment Unit questions asked how many states collect the Latitude and Longitude coordinates with the assessment decision data. An option considered for the Pilot involved submitting the Latitude and Longitude coordinates of the monitoring stations that are tied to the Assessment Unit, and providing information on how far to navigate upstream or downstream from that location to designate the boundaries of the Assessment Unit. Thus, it was important to learn how many states collect the Latitude and Longitude coordinates with the assessment decision data, collect the data but store it elsewhere (which may be more difficult to directly associate with the appropriate Assessment Unit), or don't collect this data. Responses showed that all states who provided input are collecting the Latitude and Longitude coordinates.<sup>14</sup> However, fewer than half of the respondents store the coordinates with the Assessment Unit decision data (attribute data). The remaining states store the coordinates in another database or spreadsheet, which may be more challenging to connect with the appropriate Assessment Unit decision data.

It was also important to understand whether states used the sampling results from the monitoring stations as the primary factor to determine the Assessment Unit location and boundaries. Based upon the responses, most states use more than just the monitoring location to define the extent of an Assessment Unit. Nine states did not consider monitoring stations at all during this process. Twenty-seven states (out of the 37 states who responded), considered five or more factors when defining the extent of an Assessment Unit, and six of those states considered nine or more factors. The five factors considered most frequently during Assessment Unit delineation included:

- Hydrologic modification such as channelization or a dam
- Large tributary or diversion
- Multiple monitoring stations
- Point source influences
- Single monitoring station

States were also asked about the precision of their Assessment Unit boundaries. Nineteen states indicated that they use a high level of precision (plus or minus a few yards), 14 states use a medium level of precision (plus or minus 100 yards) and only one state uses a low level of precision (plus or minus a mile) in delineating their Assessment Unit boundaries.

### Geospatial Practices

It was important to better understand state capacity for georeferencing, and how states manage their geospatial data. The questions on geospatial practices addressed which reference hydrography layers were used by states to georeference their assessment data, as well as the spatial resolution of those layers. Based on the responses, 22 states (out of the 35 states who responded to the question) georeference water quality assessment data to a version of the NHD. Also, 24 states (out of the 38 states who responded to the question) georeference to high resolution

---

<sup>14</sup> Some stations that are monitored by outside organizations may not contain the location coordinates.

geospatial data of some format, 11 states georeference to medium resolution geospatial data of some format and 3 states use a mixture of resolutions.

In the 2006 Integrated Reporting Guidance, EPA recommended that states use the medium resolution NHD to report Assessment Unit geographic information. However, at the time, EPA also recognized that states may consider moving to higher resolution once developed.<sup>15</sup> To identify the number of states that have moved to higher resolution to meet their internal needs, yet still report to EPA using the medium resolution NHD, EPA asked states if they maintain their geospatial data at both the medium resolution and a higher resolution. Based on the responses, four states maintain and submit to EPA geospatial data at the medium resolution, and maintain geospatial data at a higher resolution to meet their internal needs, placing an increased burden on states to manage geospatial data at two resolutions. Most states, however, have completely moved away from using medium resolution NHD and only use some form of high resolution NHD to meet their internal needs, as well as to report to EPA. While this approach reduces state burden, it increases EPA burden to process high resolution data under the traditional georeferencing process.

Most states now use geospatial data in their state water programs. In the past, there were many states that didn't use any geospatial data with their water quality assessment decision data, so creating geospatial data for EPA imposed a burden. Based upon the state responses, most states now routinely use geospatial data to georeference assessment data for their own purposes, and, as a result, the burden has now shifted from the initial creation of the geospatial data, to the use of a specific format and resolution (such as medium resolution NHDPlus) for the geospatial data provided to EPA.

State staff resources to georeference assessment decision data are limited. Nineteen states (out of the 38 states who responded to the question) have less than one full time equivalent (FTE) assigned to georeferencing for water, and of those states, only three have that FTE solely dedicated to 303(d) and 305(b) georeferencing—the others must share the partial FTE with other programs.

## **Pilot Workgroup Participants and Pilot Data**

Thirteen states participated in the Pilot workgroup:

- EPA Region 1: Maine, Massachusetts
- EPA Region 2: New Jersey
- EPA Region 3: Pennsylvania, West Virginia
- EPA Region 4: Alabama, Georgia, Kentucky, Mississippi
- EPA Region 6: Texas
- EPA Region 8: Nevada, Utah
- EPA Region 9: Hawaii

Pilot states participated in phone and web conferences, provided feedback on Pilot options, shared information on the level of effort that would be necessary for their state to modify their geospatial data to implement the different Pilot options, supplied geospatial data files for testing, and helped review the Pilot outputs.

For the Pilot purposes, EPA supplemented the test files received from Pilot states with past geospatial data submissions from other states to provide a broader geographic coverage, as well as to test for specific state geospatial data issues that have provided difficulties using EPA's traditional georeferencing methods. The states for which this supplemental data were used included Colorado, Florida, Idaho, North Carolina, New Hampshire, Ohio, Rhode Island, South Carolina, and Wisconsin.

---

<sup>15</sup> 2006 Integrated Reporting Guidance: <http://www.epa.gov/tmdl/identifying-and-listing-impaired-waters>

## **6. INITIAL SET OF PILOT OPTIONS EXPLORED**

The initial design of the Pilot was intended to explore three options for state-submitted geospatial data: Point-based indexing, Watershed Polygon-based indexing, and Feature-based indexing. These options focused on the type of Assessment Units represented by the state geospatial data (i.e., sampling location, watershed boundary, or linear and area waterbody features such as streams or lakes). The Pilot considered how a more-automated georeferencing process working with these different types of state-submitted geospatial data might meet the Pilot goals of improved cost effectiveness, timeliness, quality and completeness. For more information on these initial options, refer to Appendix E: Information on Initial Pilot Options.

### **Point-based Indexing**

The concept of Point-based indexing was based on the thought that states could simply provide the Latitude and Longitude coordinates for a monitoring station and supplemental information describing how far upstream or downstream from this point the Assessment Unit boundaries extended to create a geospatial feature, in a tabular format as part of their attribute data submissions. The state data management practices confirmed that there are more considerations involved in the Assessment Unit delineation than just monitoring site location. The Pilot states also responded that generating point coordinates from existing geospatial linear and polygon events would be a greater burden than providing their existing geospatial data. Furthermore, the Pilot team subsequently concluded that the ability to satisfactorily define the upstream and downstream extent of the Assessment Unit boundary from a monitoring site using supplemental upstream and downstream information from states would not produce the desired precision conveyed in the state data management practices. For this approach to produce satisfactory results, the state-supplied supplemental information would need to reference a common surface water network and, thus, would effectively impose on states the use of one format and resolution for their geospatial data. This effect would run counter to the desire to more effectively leverage the mixed resolution geospatial data already used by states for their own purposes. In summary, this approach would be less precise and more burdensome than traditional methods, and, as a result, the Point-based indexing option was dropped from the Pilot due to low state interest.

### **Watershed Polygon-based Indexing**

The Watershed Polygon-based indexing option arose from the fact that some states manage their Assessment Units on a watershed scale. Instead of providing the linear and area waterbody features within the watershed polygon, states would provide the watershed polygon boundary. From this boundary, the underlying hydrographic features would be identified. The Watershed Polygon-based indexing option received limited interest among Pilot states because few states currently use Watershed Polygons as their Assessment Units, since many states prefer to work at a more-detailed level with individual surface water features. However, the option was retained and evaluated in order to accommodate the states that do manage their Assessment Units at the watershed level.

### **Feature-based Indexing**

The Feature-based indexing option was based on the fact that states manage geospatial data at different resolutions. More frequently, states now use a higher resolution reference layer than the medium resolution NHDPlus that EPA employs. Feature-based indexing was designed to address two key issues: 1) to allow states to submit high resolution geospatial data that they use in their own programs and 2) for EPA to begin using state-submitted features for display and analysis while georeferencing the data to the medium resolution NHDPlus to meet EPA's reporting and analysis needs.

The Feature-based indexing option received the most interest from states in the Pilot workgroup because most Pilot states are already using high resolution data, and Feature-based indexing would encourage all such states to submit the geospatial data in the resolution they already use. The Feature-based indexing option explored more-automated methods than the traditional processing techniques to relate high resolution state-submitted features

with the medium resolution NHD features within NHDPlus. More specifically, Feature-based indexing took advantage of the common content (reach codes, network relationships, stream names, etc.) shared by high resolution NHD and medium resolution NHD. These common NHD fields help establish a correspondence between EPA's high resolution NHD snapshot and the medium resolution NHD snapshot in NHDPlus. Any state-submitted high resolution geospatial data was first overlaid on EPA's high resolution NHD snapshot and then related to medium resolution NHD using this High-Resolution-to-Medium-Resolution NHD correspondence. This approach included attempting to identify the subset of high resolution features that corresponded to each medium resolution feature. State-submitted medium resolution NHD geospatial data would still be directly overlaid on EPA's medium resolution NHD (in NHDPlus). Initial results showed that the approach was so excessively computer resource intensive and time consuming that it was not feasible to implement within EPA's computing environment.

## 7. CATCHMENT-BASED INDEXING – RECOMMENDED PILOT OPTION

### Catchment-Based Indexing

The initial set of options (Point-based indexing, Watershed Polygon-based indexing, and Feature-based indexing) explored in the Pilot proved to be unsatisfactory due to either low interest from states or the high level of resources required to implement and operate.

After considering the lessons learned from these initial approaches, the Pilot team formulated a fourth option that would provide many of the same benefits offered by the popular Feature-based indexing option, while addressing its implementation and operational shortcomings. Figure 2 shows how the original Pilot options evolved into the Catchment-based indexing approach. Since the desired conditions were not met in the previous options, the best components of Feature-based indexing were combined with other existing geospatial technology to test the idea of referencing to NHDPlus catchments. Since catchments are already linked to the underlying NHD network within NHDPlus, using catchments as a framework allows EPA to leverage existing NHDPlus-based tools for analysis and reporting purposes.<sup>16</sup>

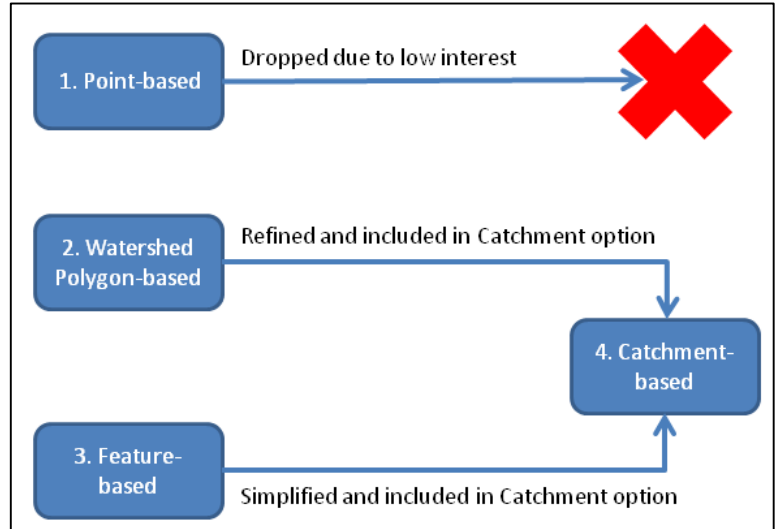


Figure 2: This figure shows the evolution of the Pilot indexing options.

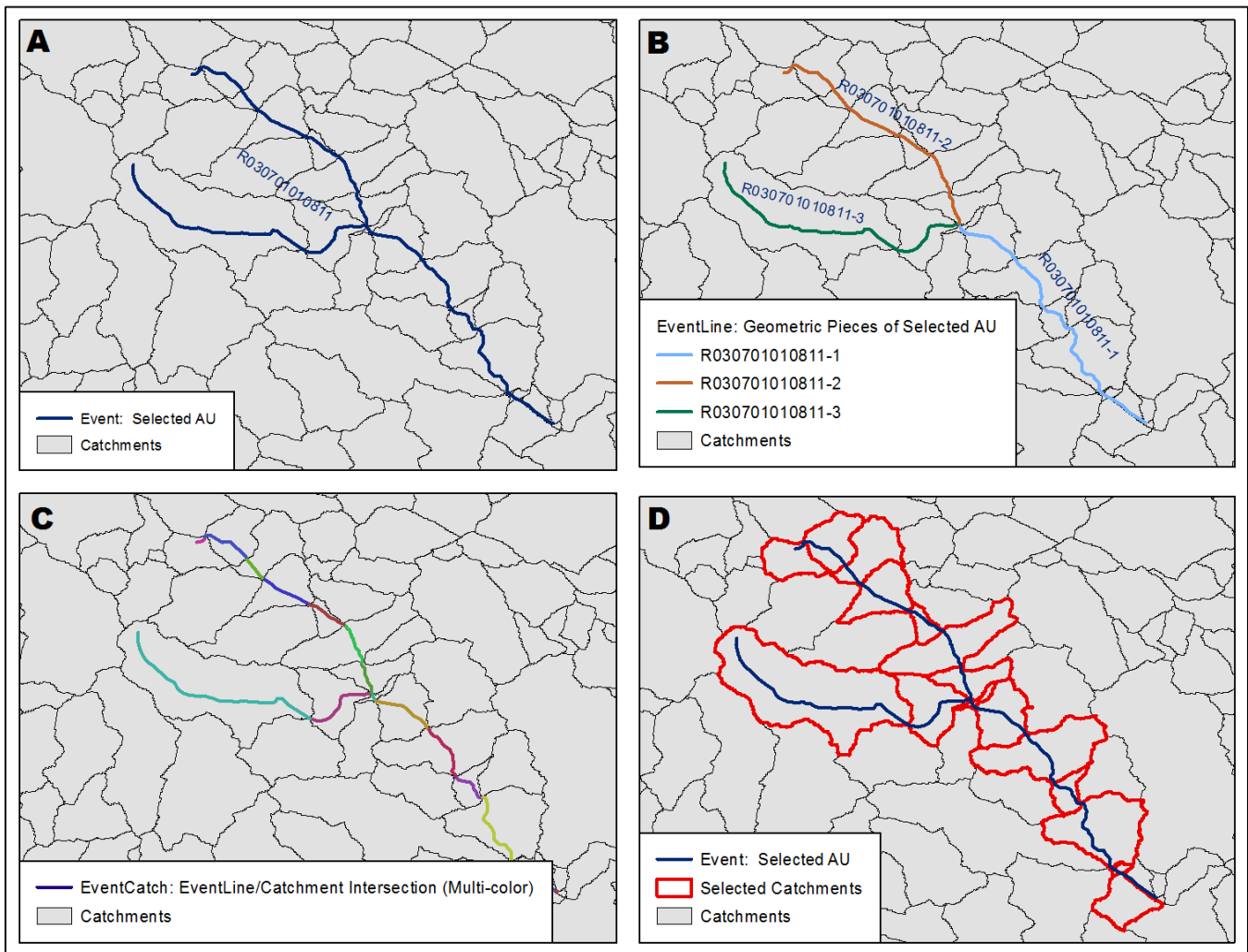
The new approach, known as Catchment-based indexing, helps to address longstanding state concerns about downgrading their high resolution geospatial data by mapping it onto medium resolution NHD features. Catchment-based indexing is fundamentally different from the other options in that it accepts any Assessment Unit type (e.g., points, watersheds, lines, or area waterbody features) as an input and then associates them with NHDPlus catchments instead of individual medium resolution NHD surface water features. NHDPlus catchments are defined as the local drainage areas for individual surface water features within NHDPlus (see Appendix B for more specific information on NHDPlus catchments). Initial responses from EPA, state and other data users on this new approach have been positive to date.

<sup>16</sup> Under CWA Sections 303(d) and 305(b), states have a significant amount of flexibility in how they report their information to the EPA. As a result, it is difficult for the EPA to synthesize the variety of state-submitted geospatial data to accurately reflect the status of the nation's waters in a geographic map. To accommodate existing flexibility in state reporting of geospatial information to EPA, while, at the same time, providing a simpler, nationally consistent approach for integrating these data, EPA proposes to use NHDPlus catchments as EPA's common unit to represent the status of those Assessment Units. States would still use Assessment Units for tracking and reporting their water quality assessment decisions. The detailed state-submitted geospatial information at the Assessment Unit level would be retained and used for display purposes. The EPA is exploring the use of catchments to analyze state-reported data, so this process will continue to evolve over the next several Integrated Reporting cycles as the EPA and states learn and receive feedback on how this data will be used for activities such as reporting on strategic measures, analyses, and displaying information for public consumption.

## Catchment-based Indexing Process

Catchment-based indexing associates each state-submitted Assessment Unit with its corresponding set of NHDPlus catchments. A conceptual description of the process is presented here and the detailed specification is provided in Appendix C: Technical Specification for Catchment-based Indexing. The process varies somewhat depending upon the type of geographic input (linear features such as streams, area waterbody features such as lakes, or watershed area features such as HUCs).

For linear Assessment Units (i.e., streams or rivers), the first step in Catchment-based indexing is to divide the Assessment Unit (shown in Figure 3A), which can be a complex set of interconnected lines, into individual Assessment Unit line segments (or pieces of geometry) that can be processed separately (as shown in Figure 3B). The individual pieces of geometry are then overlaid on the catchments to identify an initial set of catchments that will be further evaluated. The Assessment Unit shown in Figure 3A was initially associated with 19 different

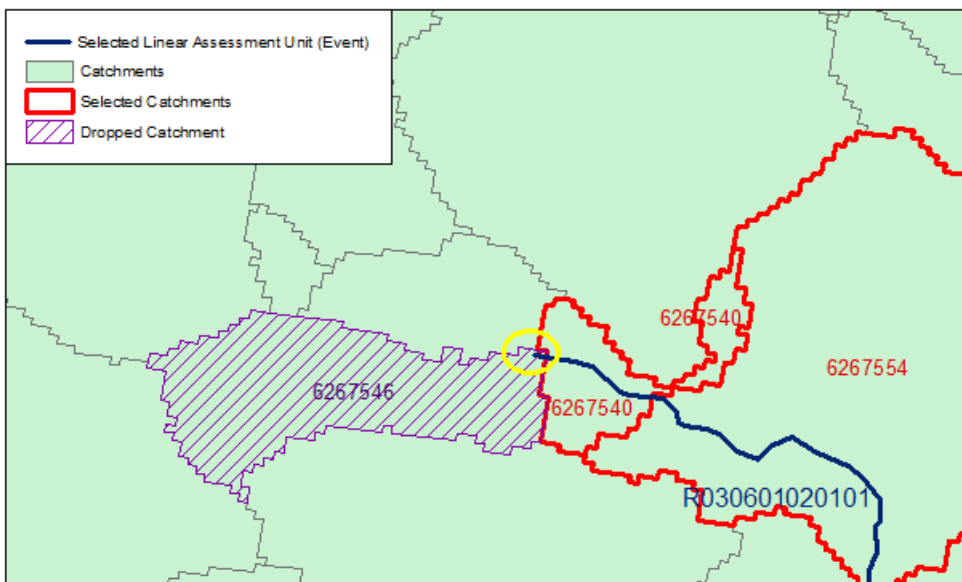


**Figure 3:** This figure represents a simplified version of the Catchment-based indexing process for a linear Assessment Unit (i.e., stream or river). In this example, one state-submitted Assessment Unit (in graphic A) is separated into three individual pieces of geometry (in graphic B). The Assessment Unit intersects with 19 catchments to create 19 uniquely colored segments (EventCatch, as shown in graphic C). Those segments and catchments are evaluated using logic built into NHDPlus to determine which catchments should be associated with the Assessment Unit in the final output file and which ones should be excluded. Graphic D shows the final catchments associated with the example Assessment Unit. More details regarding this process are located in Appendix C.

catchments. Each portion of the piece of geometry that was associated with a different catchment is displayed in a different color in Figure 3C.

Next, the catchments in the initial set are evaluated one stream level path at a time to confirm whether they should be included in the results file, and whether any new catchments that were not initially selected should be included. The NHDPlus Level Path attribute, which identifies a specific stream, is used in combination with the NHDPlus Hydrologic Sequence (Hydroseq) Number attribute, which identifies whether one catchment is upstream of another, to evaluate the catchments on each level path of the piece of geometry - starting with the main level path and then moving on to any upstream or downstream level paths.

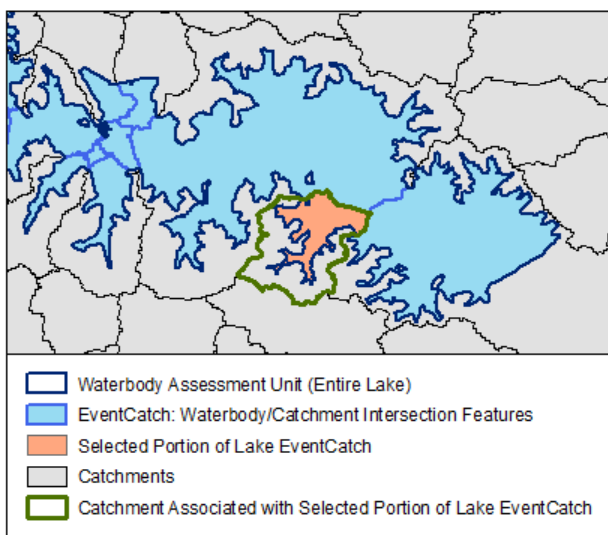
Once all level paths associated with an Assessment Unit piece of geometry have been evaluated, the process confirms whether catchments at the upstream and downstream ends of the Assessment Unit piece of geometry were correctly included and adds or removes them from the results file as appropriate. For example, Figure 4 shows that catchment 6267546 was initially included by the catchment overlay process. The Assessment Unit segment in that catchment is very short (36 meters). To address these situations, catchments containing pieces of the Assessment Unit that were less than 100 meters in length were rejected. The stair-stepped appearance of the red catchment boundaries is an artifact of the 30 meter gridded elevation used to delineate the boundaries. The 100 meter threshold was selected because it was approximately the length of three 30 meter grid cells. The final results from the Figure 3A example Assessment Unit are shown in Figure 3D.



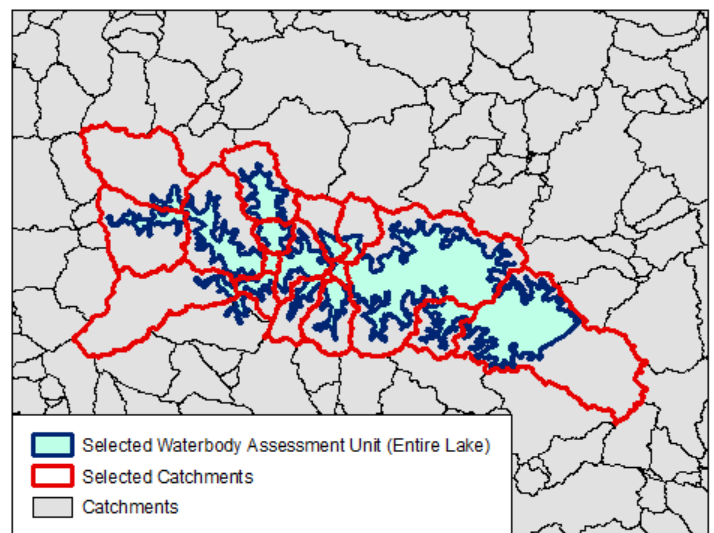
**Figure 4: The linear Catchment-based indexing process uses thresholds to exclude catchments that are assigned to an Assessment Unit due to a tiny piece of stream segment crossing into the adjacent catchment. In this example, catchment 6267546 was initially included because of a very short overlapping piece of the Assessment Unit (36 meters). For purposes of this Pilot, a 100 meter minimum length threshold was applied to determine whether the catchment should be removed from the output catchment file. In this case catchment 6267546 was dropped from the final output file.**

For area waterbody Assessment Units (i.e., lake, pond, or reservoir), the first step in Catchment-based indexing is to overlay the state-submitted area waterbody features on the catchments. For the purposes of this Pilot, catchments were added to the results file when the piece of the area waterbody Assessment Unit overlapping a specific catchment was greater than or equal to 1% of the total area of the Assessment Unit, or the waterbody

Assessment Unit covered at least 50% of the catchment, as shown in Figure 5.<sup>17</sup> These thresholds may need to be refined in the future. The small threshold of 1% was selected for the first threshold (percent of the total Assessment Unit area that falls in the catchment) so that long, slender waterbodies such as two-dimensional streams could be corresponded to catchments without losing connecting pieces. The threshold of 50% was selected for the second threshold (percent of the catchment covered by the area waterbody Assessment Unit) so that if the waterbody covered at least half of the catchment, the catchment would be included. Catchments containing part of the area waterbody Assessment Units were also added to the results file if the catchment was associated with an Artificial Path. The selected catchments that correspond to the example area waterbody Assessment Unit are shown in Figure 6. Note that NHDPlus catchments are delineated for NHD waterbodies by treating the artificial paths within the waterbodies as extensions of the stream networks flowing into and out of the waterbody and ignoring the shorelines, as shown in Figure 7 (on page 17).



**Figure 5: The area waterbody Catchment-based indexing process is based on three characteristics. The catchment is included in the output file if any of the following is true: 1) the Selected Portion of Lake EventCatch size is greater than or equal to 1% of the overall Waterbody Assessment Unit size; or 2) the Selected Portion of Lake EventCatch is greater than or equal to 50% of the associated catchment size; or 3) the catchment is associated with an artificial path.**



**Figure 6: This figure shows the correspondence between an area waterbody Assessment Unit (e.g., a lake) and the NHDPlus catchments. In this case, the process made associations with catchments for all but a small piece of the lake.**

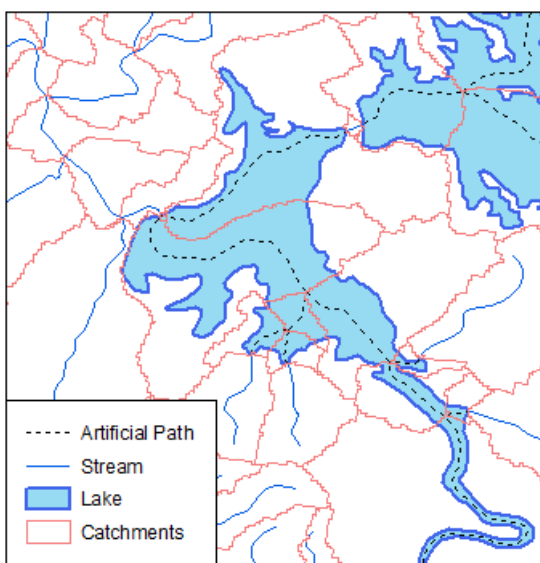
For area polygons representing watersheds, such as USGS Hydrologic Units (HUCs), the first step in Catchment-based indexing is to overlay the state-submitted HUC features on the catchments. While the boundaries of the Watershed Boundary Dataset (WBD) were referenced during production of NHDPlus, the catchments do not always conform to 12-digit hydrologic unit (HUC12) boundaries. Where HUC12s and catchments share boundaries, they usually align, but may not always align at the pour points. More information is provided in Appendix B: Background Information on NHDPlus Catchments. For the purposes of this Pilot, catchments were added to the results file when the piece of the HUC Assessment Unit overlapping a specific catchment was greater

<sup>17</sup> These tolerances for processing area waterbody Assessment Units may need to be refined in the future.

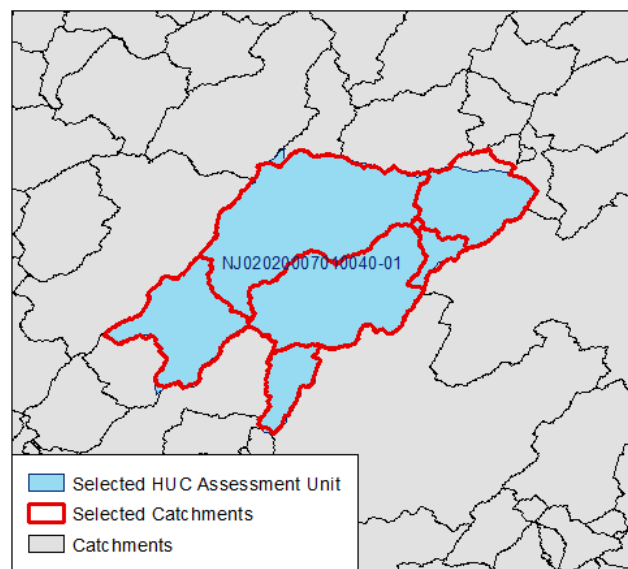


than or equal to 1% of the total area of the HUC Assessment Unit, or the HUC covered at least 50% of the catchment.<sup>18</sup> These thresholds were selected to match the thresholds used for the area waterbody process. The HUC process varies slightly from the area waterbody Assessment Unit process, because the HUC process does not need to account for artificial paths. Instead, the HUC process is solely a geospatial process that relies on the thresholds.

For standardized watershed-derived polygons such as USGS or WBD HUC12s, the process works well. For state-defined HUC14s or non-standardized land-based (HUC-like) polygons, the process works well in some cases but may be less effective in other cases, depending upon how closely the non-standardized polygons align with the NHDPlus catchments. An example of the correspondence between catchments and a state-defined HUC14 is shown in Figure 8. Additional examples are provided in Appendix C: Technical Specification for Catchment-based Indexing.



**Figure 7:** NHD artificial paths run through NHD waterbodies to connect the linear stream network that flows into and out of the lake. This figure shows artificial paths (the dotted lines) flowing through a waterbody (the blue polygon). These artificial paths connect the linear stream network (blue lines) flowing into and out of the lake. The catchments (shown in pink here) are associated with the linear stream network as well as with artificial paths.



**Figure 8:** The HUC Catchment-based indexing process is based on two characteristics. The catchment is included in the output file if either of the following is true: 1) the portion of the HUC inside the catchment is greater than or equal to 1% of the overall HUC size; or 2) the portion of the HUC inside the catchment is greater than or equal to 50% of the associated catchment size. This figure shows the correspondence between a watershed Hydrologic Unit (HUC) or similar (HUC-like) polygon and NHDPlus catchments. This example uses a state-defined HUC14 from New Jersey.

<sup>18</sup> These tolerances for processing Hydrologic Units (HUCs) or similar state polygons representing continuous land surface (HUC-like) may need to be refined in the future.

## 8. RESULTS

As discussed in Section 7, because the desired conditions were not met in the initial suite of Pilot options, EPA combined the best components of Feature-based and Watershed Polygon-based indexing with other existing geospatial technology to test the idea of referencing to NHDPlus catchments. The prototype Catchment-based indexing procedures were developed for different input scenarios including: linear features to catchments (Linear-to-Catchment) for rivers and streams; area waterbody features to catchments (Waterbody-to-Catchment) for lakes, ponds, or reservoirs; and Hydrologic Unit boundaries to catchments (HUC-to-Catchment) for polygons representing sections of the land such as watersheds.

While this Pilot did not specifically address point features such as monitoring stations, the Catchment-based indexing approach should be easily adaptable to accommodate that input. Also, this Pilot did not address coastal areas where NHDPlus catchments do not exist. A separate strategy that handles coastal areas where catchments do not exist would need to be incorporated into the Catchment-based indexing approach going forward. In addition, EPA is currently working with Alaska, for which NHDPlus catchments are not yet developed, and discussions are underway to determine an interim solution.

NHDPlus catchments are based on hydrography rather than political boundaries, so some catchments cross state boundaries. For purposes of the Pilot, entire catchments were used. However, moving forward, depending on the needs of the program, it may be more representative to clip catchments to state boundaries so that only the portion of the catchment that falls within a state is attributed to the state. To accomplish this, a step could be added to the automated process to use entire catchments during the process for connectivity purposes but clip catchments at state boundaries at the end of the process.

The prototype procedures were developed using Esri's ArcGIS for Desktop 10.1, and were evaluated using state-submitted input data files. The prototype procedures were refined and improved based upon these initial processing results. The procedures identify errors which are set aside for manual review and indexing. In addition, a subset of the automated indexing results is flagged for manual QA/QC review.

### Issues Encountered

When applying the Catchment-based indexing approach during the Pilot, several data issues were encountered that caused individual Assessment Units to be rejected during the automated indexing process. These Assessment Units were manually reviewed to determine the cause for the rejection and to identify potential improvements to the indexing process to address the causes. Several improvements were made to the indexing process during the Pilot. The remaining data-related issues involve area waterbody Assessment Units that are represented by a series of individual lines that close on themselves, linear Assessment Units that follow one shoreline of wide rivers, and linear Assessment Units that follow the minor path at stream divergences (see Appendix C for more details). These issues represent a small portion of the total state data processed. For example, around 12,400 Event records (approximately 1.4%) of the approximately 904,000 Event records (representing approximately 146,000 Assessment Units) from 21 states that were processed by the Catchment-based indexing prototype were lines representing area waterbodies and were rejected due to the self-closing line issue<sup>19</sup>. This self-closing line issue may be addressed by providing Best Practice strategies to states that recommend not using lines to represent polygons. In addition, subsequent files that involve self-closing lines may be converted to polygons and processed through the Area Waterbody Polygon (Waterbody-to-Catchment) process. Any such issues that cannot ultimately be addressed by improvements to the indexing process would be manually reviewed and indexed. The process also encountered an apparent file size limitation where one large state-submitted file needed to be split into two files in order to be processed successfully.

---

<sup>19</sup> The Linear-to-Catchment process requires two endpoints, whereas a line that represents a polygon, such as a lake shoreline, only has one endpoint because the line starts and ends at the same point. This self-closing line issue causes an error.

## Evaluation Criteria

The evaluation criteria are based on the Pilot goals: 1) reduce EPA processing costs and state reporting burden, 2) publish geospatial data in a timelier manner, 3) maintain data quality to meet the requirements of the users, and 4) have a more complete national dataset.

### Costs

One of the goals for the ATTAINS program is to ensure that the level of investment for georeferencing is appropriate for characterizing a state's Assessment Units. As a result, one of the goals for the Pilot was to identify potential cost reductions to EPA for processing state geospatial data related to water quality assessment decisions, without significantly compromising data quality. The traditional process used to georeference water quality assessment and impaired water data is mostly manual, which results in a high level of effort and costs, ranging from 20 hours to more than 100 hours per state. The Catchment-based indexing approach is largely automated, with limited manual pre-processing and post-processing steps, and, as a result, is intended to reduce EPA's costs for processing state-submitted geospatial data. The pre-processing steps involve examining the state input file to determine which specific Catchment-based indexing method to apply (i.e., linear, area waterbody, or HUC) and then converting the state data into a standard NHD Event format, before initiating the automated processing. The Pilot established that the pre-processing steps should take 1 to 1.5 hours for most states. The post-processing steps include the manual indexing of Assessment Units that were rejected by the automated processing and conducting a QA/QC of the collective results. During the Pilot, any manual indexing was significantly faster since there were a limited number of Assessment Units to process when compared to the existing process. Additionally, there was considerable time saved by indexing to catchments rather than more-detailed surface water features. Thus, the estimated time for the post-processing steps ranged from 5 to 40 hours depending upon the quality and number of features in the state input data, though the estimated time to conduct QA/QC for most state files will likely range from 5-10 hours based on the Pilot data.

The Pilot also examined potential cost reductions for states to submit geospatial data to EPA. The Catchment-based indexing approach allows states to submit the geospatial data they already use for their own purposes—as long as it contains the appropriate Assessment Unit identification information—rather than needing to reformat their existing data or create new data. This approach will save states time and effort with regards to reporting burden.

### Timeliness

Another goal of the Pilot was to enable EPA to process and publish water quality assessment geospatial data in a timelier manner. Geospatial data related to water quality assessments and impaired waters are used by a variety of EPA programs and external organizations. One of the main issues these users have identified with the traditional georeferencing process is the total time between the submission of state input files and the publication of the resulting geospatial data for a reporting cycle. In addition to the programmatic delays, which the Pilot did not address, the traditional highly manual indexing process is time-consuming. As noted in the preceding Cost section, the highly automated Catchment-based indexing approach resulted in a significant decrease in time to process many state input files, which should allow for faster publication of the resulting geospatial data.

The Catchment-based indexing approach was tested using input files containing around 904,000 Event records (which represent approximately 146,000 Assessment Units) from 21 states, including several states which had multiple input files for streams and lakes (i.e., linear or area waterbody), resulting in a total of 64 files run through the prototype. It took approximately 30 hours to process all 64 of the state files through the automated portion of the Catchment-based indexing process. Processing time for each file ranged from a few minutes to almost three hours, and the mean time to process a file through the prototype was less than half an hour. In addition to the automatic processing time, the output for each state will need to undergo a QA/QC process to ensure that the automated process accurately reflects the state data. Estimates for the time it would take to QA/QC the output

files range from 5-10 hours per state. The total time to process a state’s geospatial data (preprocessing time plus automated processing time plus QA/QC time) is expected to range from 7-15 hours per state. This is a significant improvement over the traditional georeferencing process which ranges between 20-100 hours per state. As QA/QC is performed, it is also expected that improvements to the automated process can be implemented which will have the potential to reduce the total time even further.

### Data Quality and Use Cases

The Pilot evaluated options to reduce the time and cost associated with the traditional georeferencing process for state geospatial data, while maintaining the data quality required to meet users’ needs. This was accomplished by evaluating whether the output from the Catchment-based indexing approach satisfied the collection of typical use cases for all primary and most secondary customers for ATTAINS geospatial data.

The use cases that were considered and the evaluation results are summarized in Table 1. While the number and variety of use cases for ATTAINS geospatial data go far beyond those listed here, the Pilot team felt that these use cases were representative of the larger set of uses and appropriate to meet the purposes of the Pilot. Based on this evaluation, the output from the Catchment-based indexing approach meets the collective needs of all use cases considered.

**Table 1: Summary of key use cases and whether the Catchment-based indexing approach meets their needs**

Use Case	Primary Functions	Technical Applicability of Catchment-based Indexing Approach	Satisfactory?
Viewing of where waters have been assessed, where waters are impaired, and where TMDLs have been done	Display	Information can be displayed at a catchment level. In addition, underlying state data can be displayed (a mixture of resolutions and types of referenced hydrography) for segment-level details	Yes
National Water Quality Inventory Report to Congress (National 305(b) Report)	Analysis-based calculations	For determining assessed watershed areas, analyses may use catchments. For determining length of assessed waters, may use underlying state data of mixed resolutions and types	Yes
Co-occurrence of impaired waters with other geographic features, including Federal Lands, drinking water intakes, or environmental justice communities	Analysis-based calculations	Depending on the scale and needs of the analysis, it would be possible to determine co-occurrence of catchment areas with the features of interest, or to use the underlying state data for stream-level calculations	Yes
Nitrogen and Phosphorus Pollution Data Access Tool (NPDAT)	Display; analysis-based information on either catchment, HUC8, or HUC12 levels	Displays can use either catchment or the underlying state data. Assessment and impairment data by catchment is consistent with existing summary units	Yes
Office of Enforcement and Compliance Assurance (OECA) – screening system, looking for facilities in compliance	Display; analysis on segment level	The new option will provide more timely data. Catchments can be used to identify corresponding permitted facilities.	Yes

Use Case	Primary Functions	Technical Applicability of Catchment-based Indexing Approach	Satisfactory?
STOrage and RETrieval (STORET)	Analysis: ability to tie ambient water quality monitoring stations and the raw data in STORET to the assessed water	STORET data are point based and can be associated with catchments so there is no impact. EPA is evaluating making a tighter link between monitoring locations and Assessment Units.	Yes
Strategic Plan Measures	Analysis-based calculations to track water quality and restorations	Supports or enables the revision of Program Measures. New measures will rely on Catchment-based indexing approach	Yes
National data extracts; downloads	Display, Download	Catchment level dataset will be available on the national level to provide a consistent resolution for analysis purposes. In addition, the original geometry and resolution--as well as limited attribute data--for the original state-submitted geospatial files should be available for download, though these state-submitted files should be used only for visualization because the files would be of mixed-resolution across the country	Yes
How's My Waterway Mobile Application	Display; Web Services; Co-occurrence analysis; Communication to the general public	Display of assessed and impaired waters would be possible with catchments, or with underlying state data on stream level; Web services can be redesigned to work with the new datasets; Communication to the public will still require explanation of data caveats, but the ability to display state-submitted data should improve transparency	Yes
Electronic Multi-Sector General Permit Notice of Intent (eNOI)/ Construction General Permit (CGP)	Co-occurrence of construction general permit area and impaired waters in catchments	Using the catchments to make the linkages between a General Permit and an assessed/impaired water will provide equivalent functionality to what currently exists.	Yes
Upstream/downstream relationship	Analysis of upstream or downstream locations	Catchments carry upstream/downstream attributes through their relationship with the underlying NHDPlus features	Yes

The Pilot examined the level of accuracy with which state geospatial data are represented in EPA's national ATTAINS geospatial datasets and the ability for EPA to accurately display or represent state waters. The traditional approach for synthesizing state water quality assessment geospatial data nationally has been to georeference the state data to the surface water features within the medium resolution NHDPlus. The medium resolution NHDPlus provides national consistency in resolution and, also, a suite of related geospatial products that EPA uses to support a robust set of display, analysis and reporting tools. On the other hand, states are using a variety of increasingly high resolution geospatial data to represent their water quality assessment data. Over the years, this situation has led to increased dissatisfaction by states with the EPA georeferenced data, since many high resolution surface water features are either not present in the medium resolution NHDPlus, or are represented in less detail.

The Catchment-based indexing approach improves upon this situation by associating state geospatial data of any resolution with catchments, which are small drainage areas. Instead of high resolution state Assessment Units being forced to fit onto individual medium resolution features, they are associated with the set of catchments that correspond to the Assessment Unit. In addition to the catchment representation of state-submitted geospatial data and related catchment summaries, EPA also intends to provide the original state-submitted geospatial data for display and visualization purposes. Users can then display the submitted surface water features for the Assessment Units, in combination with the catchment representation, while leveraging the underlying NHD network for upstream and downstream search capabilities.

The data quality goal of the Pilot also pursued opportunities to improve EPA's ability to track the restoration of impaired waters. In the past, states have often changed the identifiers (IDs) of some Assessment Units between reporting cycles, which has made it difficult to track the status of individual waters over time. Also, some states use separate Assessment Unit IDs to track different attributes for the same spatial location, which results in overlapping Assessment Units, and makes it difficult to summarize data without double-counting those Assessment Units. These issues can be better addressed with Catchment-based indexing because the catchments can become EPA's tracking units for reporting water quality and restoration of impaired waters, without interfering with how states track Assessment Units for their own purposes since states will still be tracking data at the Assessment Unit level.

### **Completeness**

Another goal of the Pilot was to provide a more complete ATTAINS national coverage by examining ways to increase the number of states present in EPA's national geospatial datasets each reporting cycle. Completeness depends on the states' willingness to provide geospatial data to EPA, as well as EPA's ability to process the received files in a timely manner.

States may be more willing to provide geospatial data to EPA due to the more hydrologically-representative portrayal of their data through the catchments as described in the preceding section on Data Quality. Again, EPA plans to provide the original state geospatial data for display in conjunction with the catchment representation, which also provides upstream and downstream search capabilities. In addition, the traditional highly manual georeferencing process requires a high level of effort to process and georeference state files which contain a large number of Assessment Units. Due to EPA's limited resources allocated for data processing, some state files which required a high level of effort may not be georeferenced and published for a reporting cycle using the traditional process, which leads to gaps in the national coverage. In comparison, the automated procedures used for Catchment-based indexing make it feasible to georeference states with large numbers of Assessment Units relatively quickly and lead to an increase in the completeness of the ATTAINS national geospatial datasets.

## **9. IMPLEMENTATION CONSIDERATIONS**

The Catchment-based indexing approach described in this document represents a new georeferencing paradigm. Historically, the ATTAINS program has used a single geospatial dataset for both display and programmatic purposes. Under this new option, these two types of business needs would be addressed using separate approaches. Programmatic purposes (such as cross-program analysis, upstream/downstream analysis, and programmatic measures) would use the Catchment-based indexing approach described in this report. Display purposes would be met with geospatial files provided by the states, without converting them to the medium resolution hydrographic network features in NHDPlus.

The implications for states should be minimal since the new approach will accept any existing state geospatial data as the input as long as it includes the necessary Assessment Unit identifiers. Existing review tools should support catchment display with little modification. At EPA, the existing Office of Water geospatial software and database infrastructure will need to be examined in light of any new requirements presented by Catchment-based indexing, including the retention of state-submitted geospatial data for display in conjunction with the Catchment-based representation. Any additional software or database functionality to support Catchment-based indexing will leverage the existing software and database infrastructure. More information regarding potential EPA Office of Water implementation issues are provided in Appendix D.

## **10. RECOMMENDATIONS**

The Pilot was a research and development project that analyzed new approaches for georeferencing state water quality assessment decision data. This new approach assumes that states have some type of geospatial representation of their Assessment Units. Although the Catchment-based indexing approach does not address all of the issues related to the existing process for the submission and processing of state data (e.g., it does not resolve states submitting data or Regional approval of that data in a timely fashion), this option is the most efficient approach among the options tested, leading to reduced processing costs and processing time. It effectively addresses all use cases considered. The Catchment-based indexing approach would also increase state's ability and willingness to submit their data to EPA. Furthermore, this approach would address some, but not all, of the outstanding data processing issues, while maintaining data accuracy sufficient to meet the needs of users. Lastly, it includes the use of state-submitted geospatial data for display and visualization purposes in conjunction with the catchment representation and also leverages inherent NHDPlus capabilities such as upstream/downstream search.

For these reasons, the Pilot team recommended that EPA move toward Catchment-based indexing for future geospatial processing of state water quality assessment decision data. In order for the approach to be successful, the Office of Water Project Management Office (OW PMO), which manages the Office's central geospatial software and database infrastructure, will need to be engaged regarding the implementation and subsequent operation of the new process. Additionally, the team will need to address the lack of catchment data in Alaska, as well as off-shore areas in light of program measures. A comprehensive QA/QC approach and tools to support QA/QC should be developed. Furthermore, since this approach presents a shift from historical practices, the implementation process should include an appropriate education and training component. EPA should work with states to develop "best practices" recommendations to minimize processing complications due to issues with state-submitted data. EPA should also develop educational materials for the users describing the new approach, changes, and examples of potential uses.

## 11. REFERENCES

- Brakebill, J.W., D.M. Wolock, and S.E. Terziotti. (2011). *Digital Hydrologic Networks Supporting Applications Related to Spatially Referenced Regression Modeling*. Journal of the American Water Resources Association, Vol. 47 (5), 916-931. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2011.00578.x/pdf>. [Accessed 16 December 2013].
- U.S. Environmental Protection Agency. (2005). Guidance for 2006 Assessment, Listing and Reporting Requirements Pursuant to Sections 303(d), 305(b) and 314 of the Clean Water Act. Available at: <http://www.epa.gov/tmdl/identifying-and-listing-impaired-waters>. [Accessed 16 December 2013].
- U.S. Environmental Protection Agency. (2013). Reducing Reporting Burden under Clean Water Act Sections 303(d) and 305(b). Available at: <http://www.epa.gov/tmdl/identifying-and-listing-impaired-waters>. [Accessed 16 December 2013].
- U.S. Environmental Protection Agency, Office of Water, Watershed Assessment Tracking and Environmental Results (WATERS). [Online]. Available at: <http://www.epa.gov/waterdata/waters-watershed-assessment-tracking-environmental-results-system>. [Accessed 16 December 2013].
- U.S. Environmental Protection Agency, Office of Water, Water Quality Assessment and TMDL Information (ATTAINS). [Online] Available at: <http://www.epa.gov/waterdata/assessment-and-total-maximum-daily-load-tracking-and-implementation-system-attains>. [Accessed 16 December 2013].
- U.S. Geological Survey. National Hydrography Dataset (NHD). [Online] Available at: <http://nhd.usgs.gov/>. [Accessed 16 December 2013].
- U.S. Geological Survey. National Hydrography Dataset Data Dictionary. [Online] Available at: [http://nhd.usgs.gov/NHDDataDictionary\\_model2.0.pdf](http://nhd.usgs.gov/NHDDataDictionary_model2.0.pdf) [Accessed 16 December 2013].
- U.S. Geological Survey. The Hydrography Event Management (HEM) Tool. [Online] Available at: <http://nhd.usgs.gov/tools.html>. [Accessed 16 December 2013].
- U.S. Geological Survey. *Evaluation of Catchment Delineation Methods for the Medium-Resolution National Hydrography Dataset*. By C.M. Johnston; T.G. Dewald; T.R. Bondelid; B.B. Worstell; L.D. McKay; A. Rea; R.B. Moore; and J.L. Goodall. Scientific Investigations Report 2009-5233. Available at: <http://pubs.usgs.gov/sir/2009/5233/pdf/sir2009-5233.pdf>. [Accessed 16 December 2013].
- U.S. Geological Survey and U.S. Environmental Protection Agency, NHDPlus Version 2 User Guide. [Online]. [ftp://ftp.horizon-systems.com/NHDPlus/NHDPlusV21/Documentation/NHDPlusV2\\_User\\_Guide.pdf](ftp://ftp.horizon-systems.com/NHDPlus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf). [Accessed 16 December 2013].



# APPENDICES

## List of Tables

Table 2:	Stream and Catchment Numbers and Sizes for RF1 and NHDPlus.....	29
Table 3:	Summary of Pilot Workgroup Participants' Willingness to Submit Data for the Original Pilot Options and Estimate of State Level of Effort (LOE) Required to Format and Submit Data for each Option.....	57

## List of Figures

Figure 9:	This figure shows NHDPlus stream segments and catchments for a lake and surrounding area.....	28
Figure 10:	This figure shows an example that compares the Watershed Boundary Dataset (WBD) HUC12s and NHDPlus catchments. ....	29
Figure 11:	This figure shows a single linear Assessment Unit (Event) in the left image that is split into three pieces of geometry in the right image during Step B of the Linear-to-Catchment process. At this stage, the pieces retain the same Event ID. ....	31
Figure 12:	A linear Event (Assessment Unit) that outlines part of the shoreline of a two-dimensional feature (e.g., lake or wide river) will be processed using the Linear-to-Catchment process but may provide unexpected results.....	33
Figure 13:	A linear Event (Assessment Unit) that circles an area waterbody (e.g., lake) and circles back to the same starting point will yield an error if processed using the Linear-to-Catchment process, because it only has one endpoint. ....	33
Figure 14:	In Step D of the Linear-to-Catchment process, the Event ID of the individual pieces of geometry from Step B are updated by appending a "-<subID>", in this case a sequential number, so that each piece can be processed separately. The original Assessment Unit ID is retained in the "STOrigID" field.....	34
Figure 15:	In Step E of the Linear-to-Catchment process, the EventLine file which contains the unique pieces of geometry for the Event (Assessment Unit) is intersected with catchments using the ArcGIS "Identity" function to get unique EventLine/Catchment pairs which are contained in the layer EventCatch.....	35
Figure 16:	In Step H of the Linear-to-Catchment process, the Event-Catchment pairs (EventCatch) are dissolved into unique EventID/LevelPathID combinations (EventLPI). ....	36
Figure 17:	Step K.6 of the Linear-to-Catchment process pulls together all the previous information and displays all the catchments associated with the example linear Event (Assessment Unit). ....	39
Figure 18:	This figure shows a linear Assessment Unit that follows one shoreline of a two dimensional stream or river. Since NHDPlus uses the centerline of the stream, rather than a stream bank, the shoreline may be associated with a different catchment than the stream centerline.....	40
Figure 19:	This figure shows a linear Assessment Unit that follows a stream bank of a two-dimensional stream. Two catchments that intersect the shoreline Assessment Unit are excluded because those catchments are associated with tributaries instead of the stream centerline. ....	41
Figure 20:	The Linear-to-Catchment process may miss downstream segments of an Event (Assessment Unit) when it follows a shoreline of a two-dimensional stream or river. ....	41
Figure 21:	If a linear Assessment Unit (Event) follows the shoreline of a two-dimensional stream and loops out around a tributary before rejoining the main stream, the Linear-to-Catchment process may miss catchments related to the tributary.....	42
Figure 22:	If a linear Assessment Unit (Event) includes a stream divergence, the Linear-to-Catchment process selects the catchments along the path that is designated as the main path in the NHD. ....	42

Figure 23: After the "Identity" function in Step C.2 of the Waterbody-to-Catchment process, the area Waterbody Assessment Unit (Entire Lake) in the left image is broken into pieces corresponding to the portions of the Assessment Unit that fall into each catchment (creating EventCatch) in the right image. .... 44

Figure 24: For an area waterbody event such as a lake, a catchment is included in the output file when any of the following are true: 1) the portion of the Assessment Unit (Event) that falls into the catchment is at least 1% of the overall Assessment Unit size; or 2) the portion of the Assessment Unit in the catchment covers at least 50% of the catchment size; or 3) the catchment containing part of the Assessment Unit is associated with an artificial path from the NHD Flowline. .... 46

Figure 25: This figure shows the catchments associated with the example area waterbody Assessment Unit (Event) using the area Waterbody-to-Catchment process..... 46

Figure 26: After the "Identity" function in Step D of the HUC-to-Catchment process, the HUC Assessment Unit (Event) in the left image is broken into pieces corresponding to the portions of the Assessment Unit that fall into each catchment (creating EventCatch) in the right image. .... 48

Figure 27: For a land-based area Assessment Unit (Event) such as a watershed or Hydrologic Unit (HUC), a catchment is included in the output file when either of the following are true: 1) the portion of the Assessment Unit that falls into the catchment is at least 1% of the overall Assessment Unit size; or 2) the portion of the Assessment Unit in the catchment covers at least 50% of the catchment size..... 49

Figure 28: This figure shows the catchments associated with the example HUC Assessment Unit (Event) using the area HUC-to-Catchment process..... 50

Figure 29: This figure shows a HUC14 Assessment Unit that corresponded well with the catchments, aligning almost perfectly..... 51

Figure 30: This figure shows another HUC14 Assessment Unit that corresponded well with the catchments. ... 51

Figure 31: This figure shows a HUC14 Assessment Unit that corresponded fairly well with the catchments. ... 51

Figure 32: While most catchments are smaller than HUCs, this figure shows a catchment that is larger than a HUC14. .... 51

Figure 33: This figure shows a HUC14 example that corresponded well except for one catchment..... 52

Figure 34: This figure shows another HUC14 example that corresponded well except for one catchment..... 52

Figure 35: With Point-based indexing, it is important to know whether a point is associated with a tributary on high resolution NHD or the mainstem on medium resolution NHDPlus..... 55

## Appendix A: Background Information on NHDPlus

The NHDPlus is a suite of geospatial products that build upon and extend the capabilities of the medium resolution National Hydrography Dataset (NHD) by integrating it with the National Elevation Dataset (NED) and the Watershed Boundary Dataset (WBD). Interest in estimating NHD stream flow volume and velocity to support pollutant fate-and-transport modeling was the driver behind the joint USEPA and USGS effort to develop the initial NHDPlus, which was first released in late 2006. NHDPlus has been used in a wide variety of applications since that time. This widespread positive response prompted the multi-agency NHDPlus team to design an enhanced NHDPlus Version 2 that was completed in October, 2012<sup>20</sup>.

NHDPlus Version 2 both improves and extends Version 1 data content by leveraging the significantly updated ingredient national datasets. The medium resolution NHD has benefited from over five thousand updates, including more names, more lakes and a more complete network, primarily resulting from a national review performed by modelers and editors from the USGS National Water Quality Assessment Program. Over sixty percent of the 30m NED has been updated based upon re-sampling of the growing collection of 10m elevation data. Where NHDPlus Version 1 used the WBD data for the handful of certified states that were available at the time, Version 2 includes the now complete national coverage for the Watershed Boundary Dataset. For Version 2, the process used to integrate the snapshots of these three national geospatial ingredient datasets, as described in USGS Scientific Investigations Report 2009-5233, has been enhanced to improve the hydro-enforcement and resulting catchments delineations.

The Version 2 data model accommodates the ability to specify the percent of water that travels down each path at major divergences as well as water additions, removals and inter-basin transfers. Version 2 catchment attributes will again include PRISM temperature and precipitation. Over 30,000 USGS streamflow gages, an increase of 7,000, have been located on the NHD network and were used when producing mean annual and mean monthly streamflow volume and velocity estimates for all networked flowlines in Version 2. These flow estimates account for the effects of evapotranspiration and are adjusted based upon their network relationships with streamflow gages in the downstream vicinity.

---

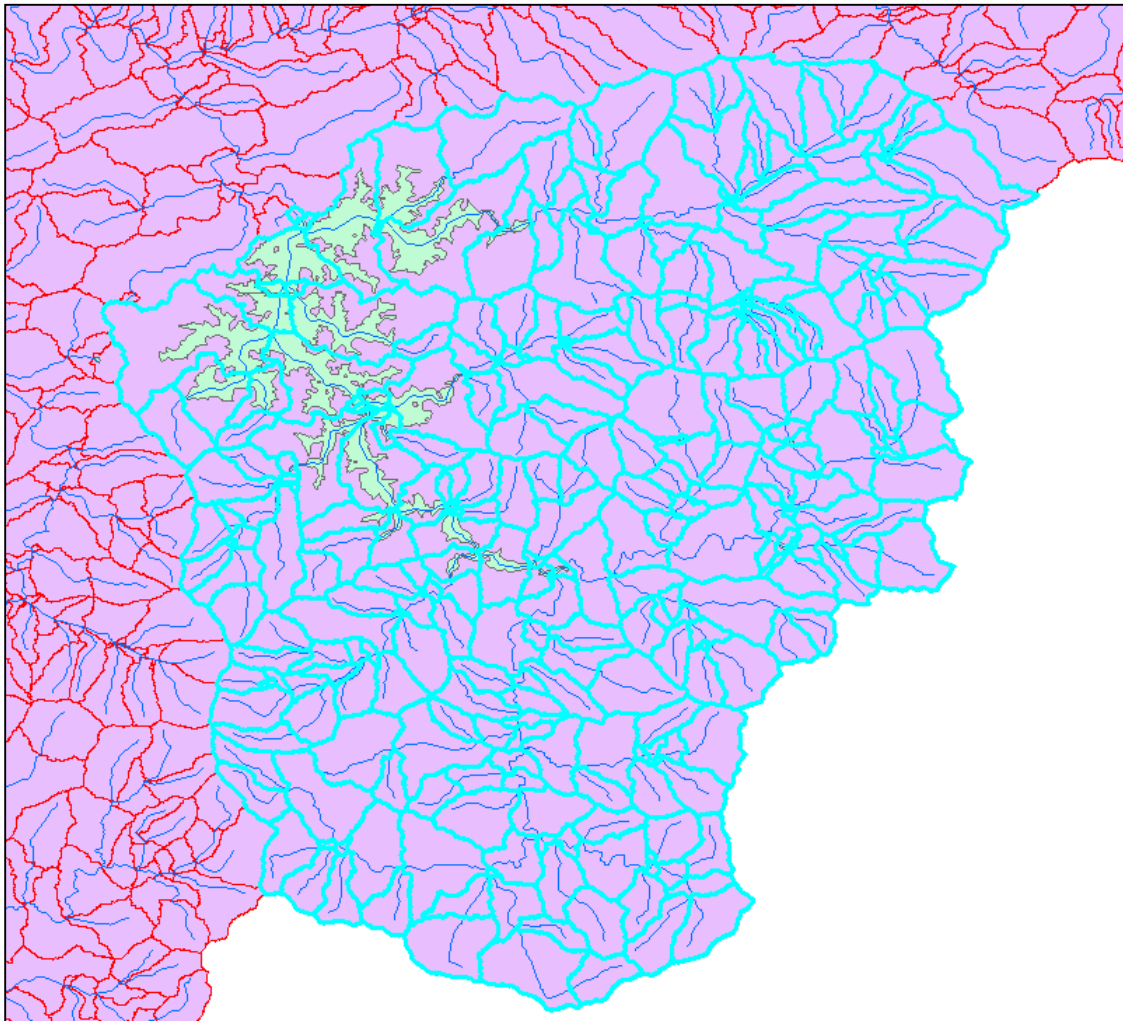
<sup>20</sup> EPA is currently working with Alaska, for which NHDPlus catchments are not yet developed and discussions are underway to determine an interim solution.

## Appendix B: Background Information on NHDPlus Catchments

### Catchment Concepts

Catchments represent the local drainage area for the individual stream segments of a specific network, such as Reach File Version 1 (RF1) at 1:500,000-scale and the medium resolution NHD component of NHDPlus at 1:100,000-scale. Individual stream segments most often span from network confluence-to-confluence. Isolated stream segments are not connected to the network and, therefore, do not have catchments.

Figure 9 shows the stream segments (blue) and associated catchments (red and cyan) for a lake (green) and its surrounding area. Note that the catchments upstream from any given location on the network can be combined to form the watershed above the location. In the figure, the catchments upstream from the outlet of the lake (cyan) form the watershed for the lake.



**Figure 9:** This figure shows NHDPlus stream segments and catchments for a lake and surrounding area.

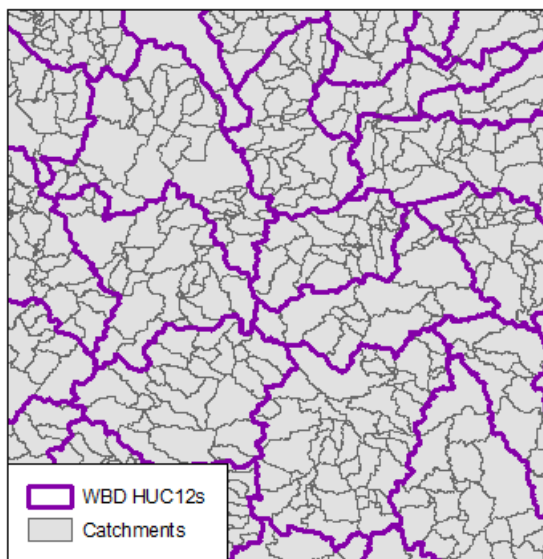
Table 2 shows the numbers and sizes of stream segments and catchments for RF1 and medium resolution NHD. These figures are approximations (+/- 10%) provided for purposes of comparison. Catchments are smaller where there are lots of streams and larger where there are fewer streams, such as dry areas in the west.

**Table 2: Stream and Catchment Numbers and Sizes for RF1 and NHDPlus**

Stream Network	Map Scale	Map Accuracy	Total Stream Miles (mi)	# of Stream Segments	Stream Segment Average Length (mi)	Catchment Average Area (sq mi)
Reach File Version 1 (RF1)	1:500K	+/- 254m	600,000	60,000	10	50
Medium Resolution NHD	1:100K	+/- 50m	3,200,000	2,600,000	1.2	1.1

For additional information on the process used to develop NHDPlus catchments, see the NHDPlus User Guide<sup>21</sup> and USGS Scientific Investigations Report 2009–5233 entitled ‘Evaluation of Catchment Delineation Methods for the Medium-Resolution National Hydrography Dataset’.<sup>22</sup>

As described in the documents just referenced, during the production of NHDPlus the boundaries of 12-digit hydrologic units (HUC12s) from the Watershed Boundary Dataset (WBD) were used to improve the common



boundaries of adjacent catchments. Figure 10 shows a comparison of WBD HUC12s and catchments. While this technique greatly improves alignment between NHDPlus catchment and WBD HUC12 boundaries, the catchments do not always conform to the HUC12 boundaries. An analysis comparing NHDPlus catchments to HUC12 boundaries shows that when aggregating NHD catchments into WBD HUC12s for the lower 48 states (which includes over 3 million square miles), 97% of the total area is in the correct HUC12. EPA and USGS continue to coordinate on how to minimize, and when possible to eliminate, the differences over time. The NHDPlus Version 2 User Guide describes the primary reasons for the current differences, most of which are definitional or scale-related.

**Figure 10: This figure shows an example that compares the Watershed Boundary Dataset (WBD) HUC12s and NHDPlus catchments.**

<sup>21</sup> NHDPlus User Guide:

[ftp://ftp.horizon-systems.com/NHDPlus/NHDPlusV21/Documentation/NHDPlusV2\\_User\\_Guide.pdf](ftp://ftp.horizon-systems.com/NHDPlus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf)

<sup>22</sup> USGS Scientific Investigations Report 2009–5233 entitled ‘Evaluation of Catchment Delineation Methods for the Medium-Resolution National Hydrography Dataset’: <http://pubs.usgs.gov/sir/2009/5233/pdf/sir2009-5233.pdf>

## Appendix C: Technical Specification for Catchment-based Indexing

Catchment-based indexing involves georeferencing state geospatial data to NHDPlus Version 2 Catchments. The process to georeference to catchments varies depending upon the type of input file: linear files (representing rivers and streams), area waterbody files (representing lakes, ponds, or reservoirs), or watershed boundary files (representing Hydrologic Unit Codes). This Appendix includes separate technical Specification documents for each of those input file types. Unless otherwise stated, in this Appendix, NHDPlus refers to NHDPlusV2. For information on a standard Event feature class, see the National Hydrography Dataset Data Dictionary<sup>23</sup> from USGS.

### **Linear Event NHDPlusV2 (Linear-to-Catchment) Specification** **Specifications for Linking State Linear Events to NHDPlusV2 Catchments**

Note: Specification is based on review of “Stream and Coastal Stream” linear events from Alabama, Colorado, Georgia, Pennsylvania, Idaho, and North Carolina.

All lengths are computed in USGS Albers projection.

The specification document includes technical steps for the process. *Additional information or explanation regarding a step is provided in italics.*

Input:

Temp folder path (input form)

Name/Location of Catchments shapefile (input form)

Name/Location of input event shapefile (input form).

*For purposes of the Pilot, Event is typically interchangeable with Assessment Unit. So, Event shapefile refers to a state’s Assessment Unit shapefile. However, a state’s Assessment Unit ID is stored in the state Original ID field (STOrigID); and the EventID may either match the STOrigID value, or it may be a subset of the STOrigID field, as noted in Step D below.*

Location for output dbf named <inputfilename>\_2Catch (input form)

Name/Location for output “failed” shapefile (input form)

PlusFlowlineVAA.dbf (national)

Development environment:

.NET

Microsoft SQL Server 2012 Localdb

ArcGIS 10.1

Windows 7 virtual machine

---

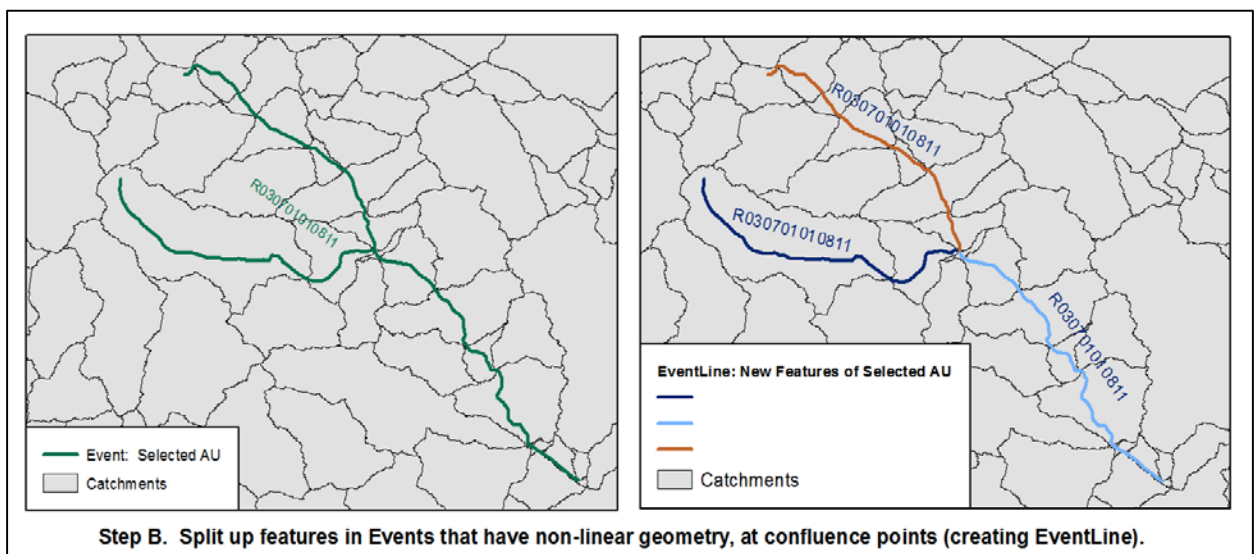
<sup>23</sup> USGS: National Hydrography Dataset Data Dictionary: [http://nhd.usgs.gov/NHDDataDictionary\\_model2.0.pdf](http://nhd.usgs.gov/NHDDataDictionary_model2.0.pdf)

- A. **(Manual) Load State’s linear events into a standard linear Event feature class** associating corresponding fields as needed. This step places the state data into a common table format that can then be run through the automated process. A standard Event feature class is in the NHDLineEvent Feature Class format with the following fields added. Fields listed in numbers A.3 through A.6 will be populated later in the process:
1. EventID [Text(100)] - If STOrigID is unique, populate EventID with STOrigID, else create unique values for EventID. Field may be updated in Step D.
  2. STOrigID [Text(100)] – State Original ID: Populate with state’s original event id.
  3. EvtLen [Double] - Event Length: length of event in meters. Field will be populated in Step C.
  4. EvCatLen [Double] – Event Catchment Length: Length of event/catchment piece in meters. Field will be populated in Step F.
  5. EvLPILen [Double] – Event Level Path ID Length: Length of event/LPI (Level Path Identifier) piece in meters. Field will be populated in Step I.
  6. EvLPIPct [Double] – Event Level Path ID Percent: Percent of event on the LPI (Level Path Identifier). Field will be populated in Step J.

For Prototype only: Create state catchment subset shapefile.

*The first step is manual. It involves loading the state’s linear events into a standard linear NHD Event feature class structure with a few additional fields that allow the storage of the state’s Assessment Unit ID. For the Pilot, the catchments for the state were subset from the national NHDPlus catchments.*

- B. **Split up features in Events that have non-linear geometry** (i.e., gaps, less than 2 endpoints, more than two end points, intersecting geometry) into true linear geometry (creating **EventLine**) with a clear beginning and clear ending. Use, for example, ArcGIS Geoprocessing tool “Feature to Line”. Figure 11 shows a single Event (AU) that is broken into three pieces of geometry by the “Feature to Line” step.



**Figure 11:** This figure shows a single linear Event (Assessment Unit), R030701010811, in the left image that is split into three pieces of geometry in the right image during Step B of the Linear-to-Catchment process. At this stage, the pieces retain the same Event ID.

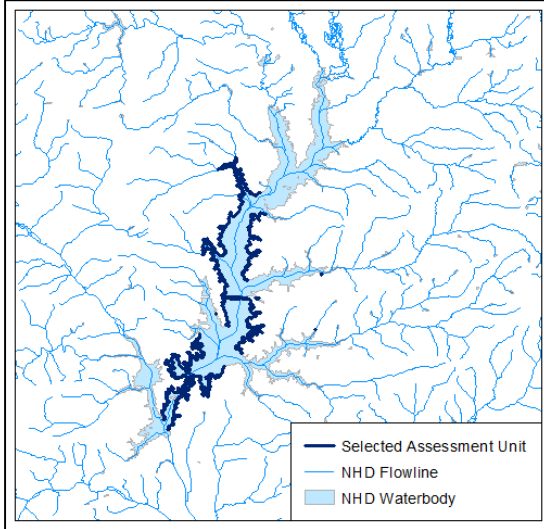
**Check lines for two and only two endpoints. Any lines with less or more endpoints, write to “Failed” shapefile, delete from EventLine.**

*In some cases, a line starts and ends on the same point—it has only one endpoint--creating a circular line. The process cannot evaluate that situation so it writes that spatial record to a “Failed” shapefile and deletes that record from the EventLine layer. See Problem Cases for examples.*

*Assessment units sometimes contain multiple geometric pieces. Often these pieces arise at locations with convergences or divergences in the hydrography. For example, some of those pieces may represent portions of the Main Flow Path, while others may represent portions of a tributary. This step takes state events and splits them into true linear shapes using an ArcGIS function called “Feature to Line”. This allows the prototype to analyze individual pieces of the Assessment Unit during future steps (e.g., analyzing the main stem separately from the individual tributaries), as shown in Figure 11. Each new piece retains its attributes, such as the Assessment Unit ID and name, so that it can be regrouped later.*

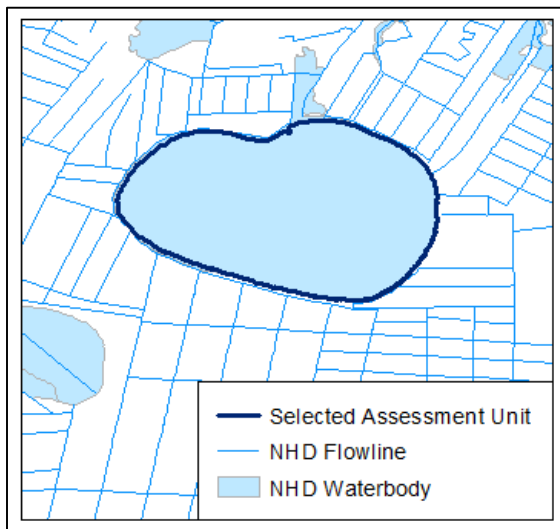


**Problem Case 1:** Figure 12 provides an example of a linear Assessment Unit (Event) that outlines part of the shoreline of a lake or two-dimensional stream or river feature. The Linear-to-Catchment process uses the stream or lake artificial paths which approximate the stream or lake centerline. The Linear-to-Catchment process will process this type of event, but may provide unexpected results. A clear definition for how this event should relate to NHDPlus is needed.



**Figure 12:** This figure shows a linear Event (Assessment Unit) that outlines part of the shoreline of a lake or two-dimensional stream or river feature. The Linear-to-Catchment process uses the stream or lake artificial paths which approximate the stream or lake centerline. The Assessment Unit shown in this example is comprised of multiple linear segments, and one segment also includes a transect across the lake. The Linear-to-Catchment process will process this type of event, but may provide unexpected results.

**Problem Case 2:** Figure 13 provides an example of a linear Assessment Unit (Event) that circles an area waterbody such as a lake and circles back to the same starting point. This event is a legitimate line, but it only has one endpoint and the event will not split into desired linear segments. The Linear-to-Catchment process cannot process this Assessment Unit because it only has one endpoint, so the Assessment Unit would yield an error. Two-dimensional water features should be represented by polygons, not lines. If this Assessment Unit had been submitted as a polygon, the Area Waterbody-to-Catchment process would have processed it successfully.



**Figure 13:** This figure shows a linear Event (Assessment Unit) that outlines an area waterbody (e.g., lake) and circles back to the same starting point. The Linear-to-Catchment process cannot process this Assessment Unit because it only has one endpoint, so the Assessment Unit would yield an error.

C. Using USGS Albers projection, compute length of features in meters and Populate EventLine.EvtLen.

D. Append a “-<subID>” to EventLine.EventID to make it unique (i.e., -001, -002, -003). The feature object ID can be used for this. Figure 14 provides an example.

*The state Original ID (Assessment Unit ID) is stored in the STOrigID field. Prior to this step, the EventID may match the STOrigID. Since Events (Assessment Units) can have multiple pieces of geometry, following Step B, above, these new pieces of geometry from the original Event (Assessment Unit) need unique IDs for the duration of the process because each step is evaluated based on a uniquely identified record. Following Step D, if the Assessment Unit contained multiple pieces of geometry following Step B, then the EventID field is updated with a suffix (or subID) to distinguish the individual pieces of geometry. Once the process is complete, the pieces of the Event can be recombined using the Assessment Unit ID (STOrigID).*

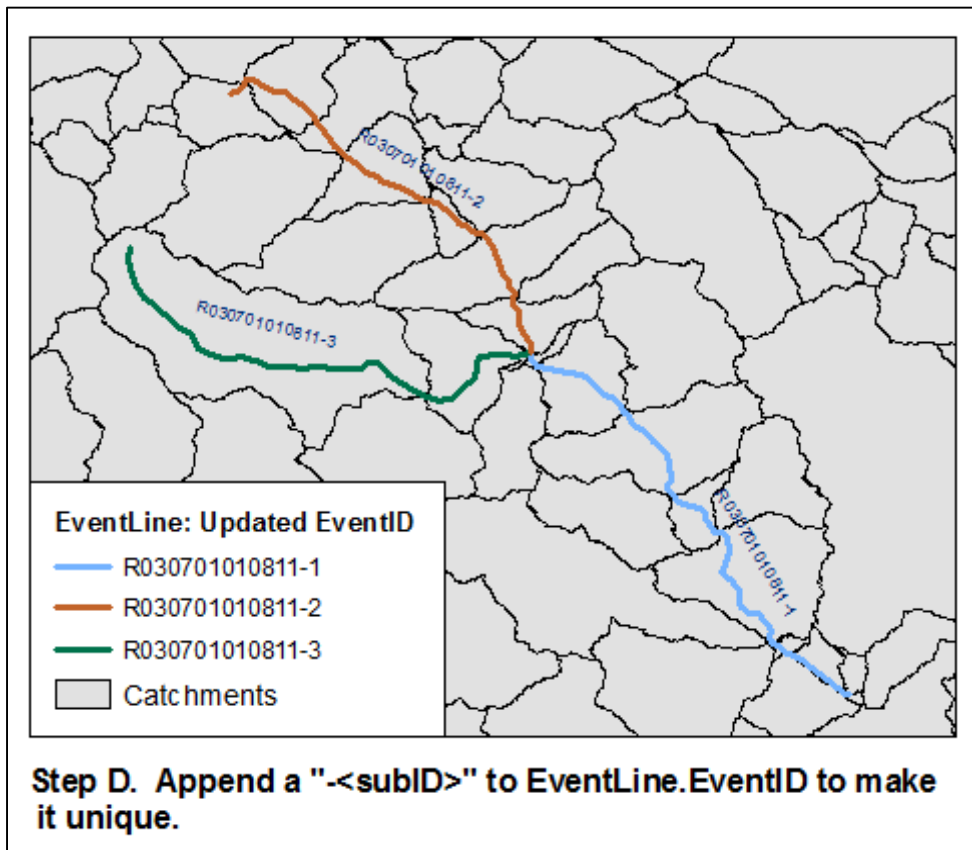


Figure 14: In Step D of the Linear-to-Catchment process, the Event ID of the individual pieces of geometry from Step B are updated by appending a “-<subID>”, in this case a sequential number, so that each piece can be processed separately. The original Assessment Unit ID is retained in the “STOrigID” field.

- E. **Intersect EventLine with Catchment creating EventCatch** that contains unique Eventline/Catchment intersection features. In EventCatch retain attributes of both EventLine and Catchment. Use ArcGIS Geoprocessing tool “Identity”. Figure 15 provides an example.

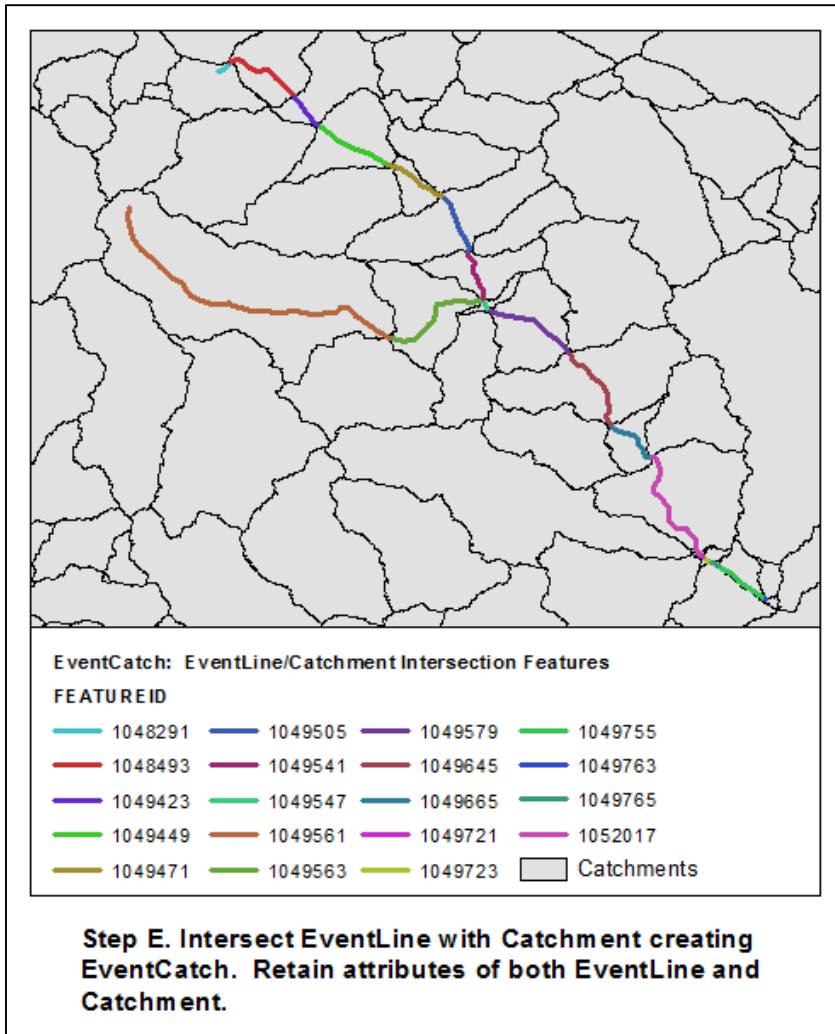


Figure 15: In Step E of the Linear-to-Catchment process, the EventLine file, which contains the unique pieces of geometry for the Event (Assessment Unit), is intersected with catchments using the ArcGIS "Identity" function to get unique Eventline/Catchment pairs which are contained in the layer EventCatch. The legend shows the catchment's ID (FeatureID) number. In this example, the "Identity" function results in 19 Event-Catchment pairs.

The "Identity" function in ArcGIS overlays and associates the Catchments with the Events, so that it can provide information on the unique EventLine/Catchment pairs ("EventCatch"), as illustrated in Figure 15.

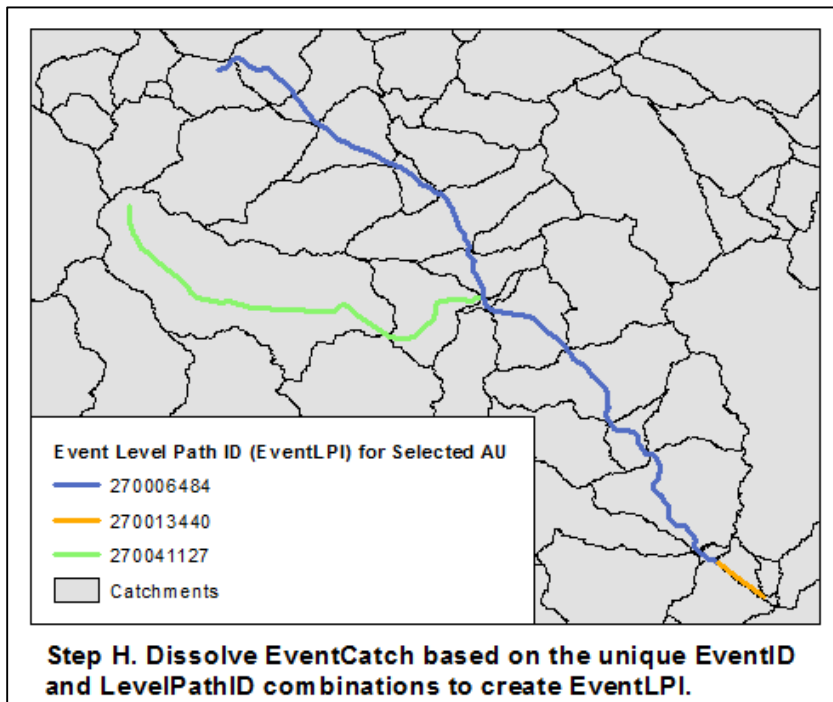
- F. **In USGS Albers projection, compute length of features in meters and Populate EventCatch.EvCatLen.** This is the EventCatch feature length in meters.

**G. Join EventCatch.FeatureID and PlusFlowlineVAA.ComID.**

*Join the NHDPlus Flowline Value Added Attributes (VAA) to the EventCatch layer to obtain specific attributes needed for the next steps (e.g., Level Path ID, Hydrosequence number).*

**Note:** H & I can be accomplished in a tabular manner by “combining” records with the same EventID and LevelPathID into a single table entry summing the EvCatLen values into EvLPILen for all combined records. This makes the rest of the specification a tabular exercise with no spatial functions.

**H. Combine geometries of EventCatch for each unique EventCatch.EventID/ EventCatch.LevelPathID combination, creating EventLPI that contains a single geometry for each unique EventID/LevelPathID combo (see Figure 16 for an example). Use, for example, ArcGIS tool “Dissolve”.**



**Figure 16:** In Step H of the Linear-to-Catchment process, the Event-Catchment pairs (EventCatch) are dissolved into unique EventID/LevelPathID combinations (EventLPI). In this example, 19 Event-Catchment pairs (EventCatch) from Step E were dissolved into 3 Event-Level Path ID pairs (EventLPI) shown here. The legend lists the Event Level Path ID (EventLPI). This example depicts the concept of Level Paths for the entire Assessment Unit (STOrigID).

**I. In USGS Albers projection, compute length of features in meters and Populate EventLPI.EvLPILen.**

*Populate the Event Level Path ID Length (EvLPILen) in the Event Level Path ID (EventLPI) table.*

**J. Populate EventLPI.EvLPIPct = 100 \* EventLPI.EvLPILen/ EventLPI.EvtLen.**

*Determine the percent of the Event that is contained in the Event Level Path ID segment by dividing the Event Level Path ID Length (EvLPILen) by the Event Length (EvtLen). Populate the Event Level Path ID Percent (EvLPIPct) attribute with that information. This data will be used to help determine whether pieces are missing from the output file in later steps.*

The objective of the remaining steps is to build the final correspondence table called FinalEventCatch - EventID/FeatureID (i.e., each EventID and all the catchment FeatureIDs that belong to it). FinalEventCatch fields: EventID, FeatureID, STOrigID

- K. These steps create a list of entries for **FinalEventCatch** for a single event – this list is called **“ThisEventCatchList”**. ThisEventCatchList holds EventID (the ID for that piece of geometry), FeatureID (catchment’s ComID), FromNode, ToNode, Hydroseq, LevelPathID, all EventCatch attributes and all EventLPI attributes.

**For each unique EventLPI.EventID (called “ThisEventID”):**

1. **Find the event’s main EventLPI.LevelPathID** which is the one with max(EventLPI.EvLPIPct), call this **“MainLPI”**.

*The Main Level Path ID (“MainLPI”) is the one with the largest percent of the total event length.*

2. **Add the main level path entries to ThisEventCatchList for MainLPI:**

- a. Select EventCatch where EventCatch.EventID = ThisEventID and EventCatch.LevelPathID = MainLPI. In selected EventCatch, find Max and Min EventCatch.Hydroseq. Refer to these as MaxSeq and MinSeq. Save EventCatch.EventID/ EventCatch.FeatureID/ EventCatch.FromNode/ EventCatch.ToNode/ EventCatch.Hydroseq/ EventCatch.STOrigID for each selected EventCatch in ThisEventCatchList.

*Select the Event-Catchment pair (EventCatch) where the Level Path ID is equal to the Main Level Path ID (MainLPI). Determine the maximum and minimum Hydrosequence numbers of that Main Level Path. Add the fields listed from the EventCatch layer, above (EventID, FeatureID, FromNode, ToNode, Hydroseq, and STOrigID) to the output file “ThisEventCatchList.”*

- b. **Fill in the missing pieces of MainLPI.** Select **PlusFlowlineVAA** where **PlusFlowlineVAA.LevelPathID = MainLPI** and **MinSeq < PlusFlowlineVAA.Hydroseq < MaxSeq**. For each selected **PlusFlowlineVAA.ComID** that is not in the selected EventCatch.FeatureID, add ThisEventID/ **PlusFlowlineVAA.ComID/ PlusFlowlineVAA.FromNode/ PlusFlowlineVAA.ToNode/ PlusFlowlineVAA.Hydroseq** to ThisEventCatchList.

*On the Main Level Path (MainLPI), select any Hydrosequence numbers that fall between the maximum and minimum Hydrosequence numbers determined in the previous step. Add those to the output file "ThisEventCatchList."*

3. In addition to MainLPI which was added in K.2 above, **add in other LPIs that flow into the top or out of the bottom** of the current Event path in ThisEventCatchList. *An event may change Level Paths at a confluence point or when the feature type or feature attributes change (e.g., Perennial vs. Intermittent streams). A piece of geometry (EventID) may change Level Paths at the top or bottom of an Assessment Unit. This step includes the catchments associated with those top and bottom ends. For example, an Assessment Unit may start on a tributary and then follow the mainstem, which would cause a change in the Level Path at the top of a piece of geometry.*  
**For each EventLPI.LevelPathID (called "ThisLPI") not yet added to ThisEventCatchList where EventLPI.EventID = ThisEvent:**
  - a. **Find top and bottom of the current event path in ThisEventCatchList.** Top is (MaxFromNode = ThisEventCatchList.FromNode for Max(ThisEventCatchList.Hydroseq). Bottom is (MinToNode = ThisEventCatchList.ToNode for Min(ThisEventCatchList.Hydroseq).
  - b. **Select EventCatch where EventCatch.EventID = ThisEventID and EventCatch.LevelPathID = ThisLPI.** Find top and bottom of the event path for ThisLPI from selected EventCatch. Top is (ThisFromNode = EventCatch.FromNode for Max(EventCatch.Hydroseq). Bottom is (ThisToNode = EventCatch.ToNode for Min(EventCatch.Hydroseq).
  - c. **If MaxFromNode = ThisToNode or MinToNode = ThisFromNode, then add selected** EventCatch.EventID/ EventCatch.FeatureID/ EventCatch.FromNode/ EventCatch.ToNode/ EventCatch.Hydroseq **to ThisEventCatchList.**
  - d. **In selected EventCatch, find Max and Min EventCatch.Hydroseq.** Refer to these as MaxSeq and MinSeq. Using PlusFlowlineVAA, Select PlusFlowlineVAA where PlusFlowlineVAA.LevelPathID = ThisLPI and MinSeq <= PlusFlowlineVAA.Hydroseq <= MaxSeq. **For each selected PlusFlowlineVAA.ComID that is not in the selected EventCatch, add the ThisEventID/ PlusFlowlineVAA.ComID/ PlusFlowlineVAA.FromNode/ PlusFlowlineVAA.ToNode/ PlusFlowlineVAA.Hydroseq to ThisEventCatchList.**

4. Repeat Step K.3 until no additional LevelPathIDs can be added to ThisEventCatchList.

*Recall that Step B split up the features that had non-linear geometry into true linear geometry so that we could process the geometric pieces separately. Step K.3 needs to be applied to each remaining geometric piece.*

5. Add ThisEventCatchList.EventID/FeatureID/STOrigID to FinalEventCatch.

*This step adds the columns EventID, FeatureID, STOrigID from ThisEventCatchList to FinalEventCatch.*

6. Using FinalEventCatch, examine the group of entries for each STOrigID (note, each entry is for one catchment that is linked to the STOrigID). Look at each endpoint (tops and bottoms) for the STOrigID. If the length of the endpoint piece is < 100 meters, remove the catchment from the STOrigID entries. *This step removes catchments that only contain very small pieces (less than 100 meters) of a state's original event (STOrigID). This threshold of 100 meters was arbitrarily chosen because it was approximately the size of three 30 meter grid cells of the catchments. Figure 17 shows the final set of catchments associated with our example Assessment Unit.*

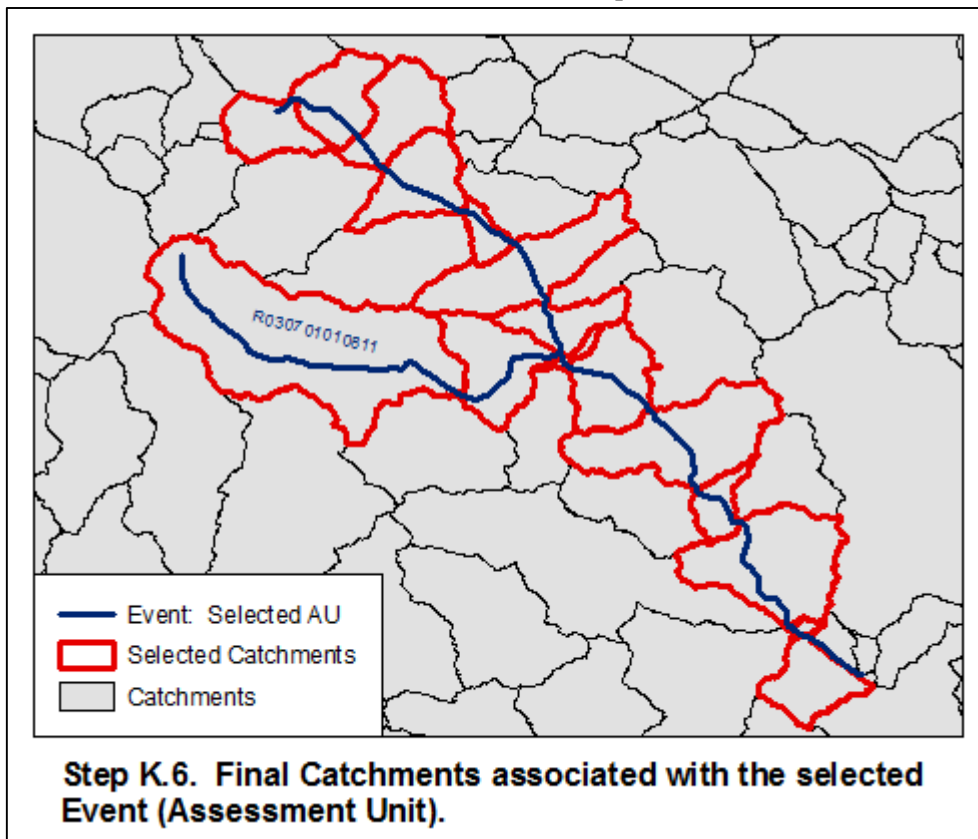


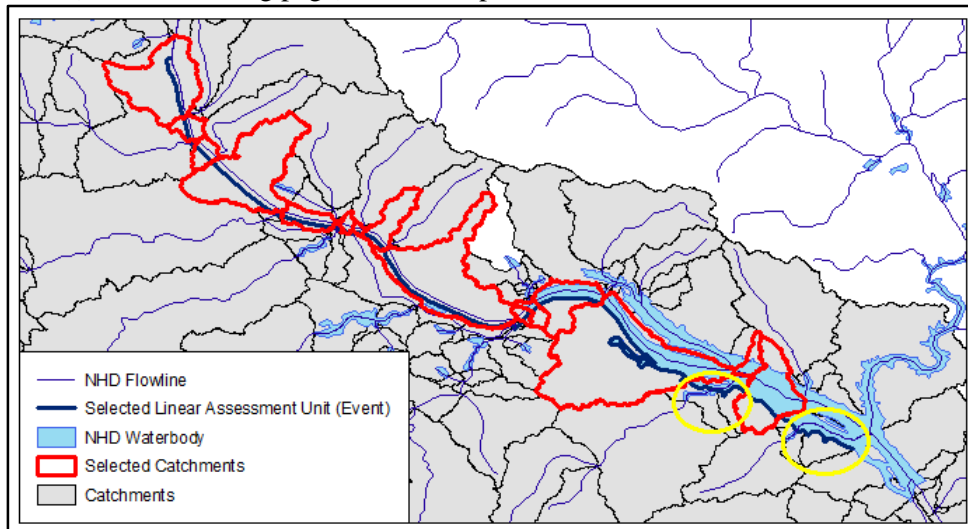
Figure 17: Step K.6 of the Linear-to-Catchment process pulls together all the previous information and displays all the catchments associated with the example linear Event (Assessment Unit), as shown here.

## Recommendations for Enhancement to the Specification and Prototype Process:

**Issue 1:** Occurring at the bottom of the main levelpath, a catchment is excluded that should be part of the event. This happens when events have been delineated on banks of wide rivers and those banks are not in a main levelpath catchment at the bottom of the event nor are they in catchments in an immediately downstream levelpath.

Solution: There is no identified technical solution for automatically addressing these cases. However, we may be able to address these cases through training or best practices.

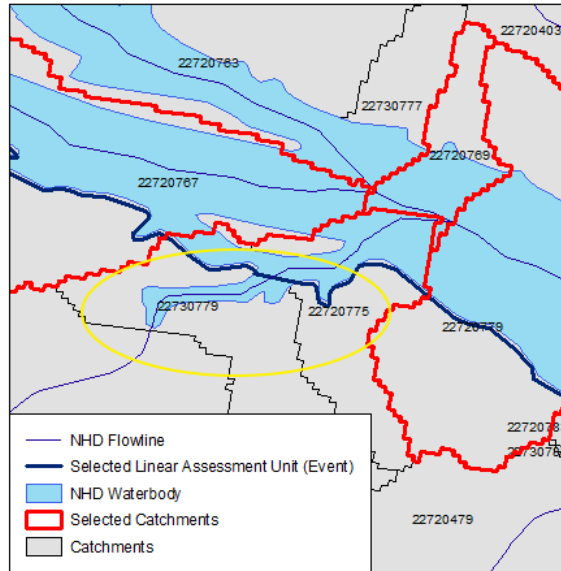
Example: The event (Assessment Unit) shown in Figure 18 is along the outside bank of a two dimensional stream/river. Correctly, catchments along the levelpath of the artificial paths within the stream/river are included using the Linear-to-Catchment process--this prevents gaps along the Assessment Unit. Correctly, catchments not along the main levelpath are excluded using the Linear-to-Catchment process because they are associated with tributaries rather than the main stem depicted by the Assessment Unit. Also, interior missing catchments along the levelpath are correctly captured by the Linear-to-Catchment process (see Figure 19 on the following page). Figure 20 (on the following page) shows a separate issue. The issue arises at the bottom of the event, where



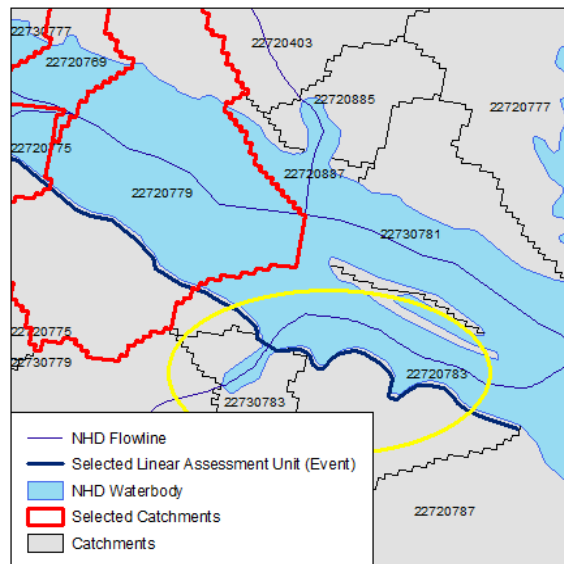
**Figure 18:** This figure shows a linear Assessment Unit that follows one shoreline of a two dimensional stream or river. Since NHDPlus uses the centerline of the stream, rather than a stream bank, the shoreline may be associated with a different catchment than the stream centerline. If that catchment is not associated with the main level path or a level path immediately upstream or downstream, the catchment may be excluded from the output file (see areas circled in yellow).

FeatureID (catchment ID) 22730781 is excluded. That catchment is part of the main levelpath, but the linear event does not touch that catchment at all. It is also not a “missing piece” with a Hydroseq between the max and min Hydroseq along the main levelpath of the identified catchments for the event. Therefore, the catchment size associated with the event is smaller than may be expected, because the process excluded catchment 22730781 (which is associated with the main level path but does not touch the linear event) as well as catchments 22730783 and 22720783 (which touch the linear event but are associated with a tributary rather than the main stem).





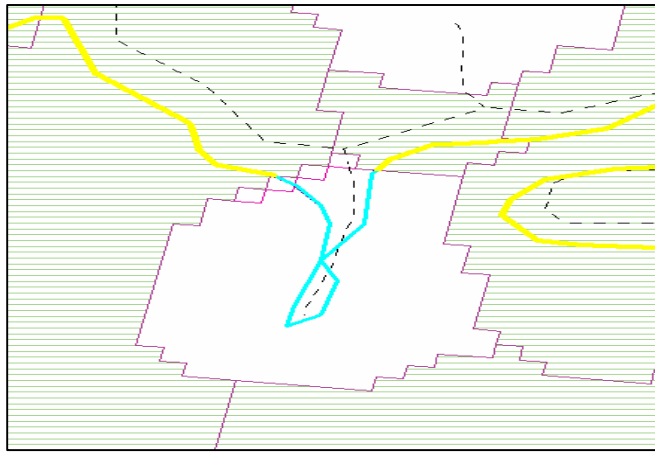
**Figure 19:** This figure shows a linear Assessment Unit that follows a stream bank of a two-dimensional stream. Two catchments that intersect the shoreline Assessment Unit (catchments 22730779 and 22720775, circled in yellow) are excluded from the output file because those catchments are on a different level path, associated with tributaries flowing into the stream. The catchment associated with the stream centerline (catchment 22720779) is included instead for connectivity.



**Figure 20:** The Linear-to-Catchment process may miss downstream segments of an Event (Assessment Unit) when it follows a shoreline of a two-dimensional stream or river. On the downstream end, the two dimensional stream is part of the main level path, which flows into catchment 22730781, but the Assessment Unit for the stream bank flows into catchments 22730783 and 22720783, instead, and those catchments are associated with a tributary so they fall on a different level path. Since that level path is not between the maximum and minimum Hydrosequence number, the process does not pick up the catchments that intersect the Assessment Unit.

**Issue 2:** Event follows edge of a wide river and loops out around a tributary and the catchment(s) for the tributary is not included. Figure 21 shows an example.

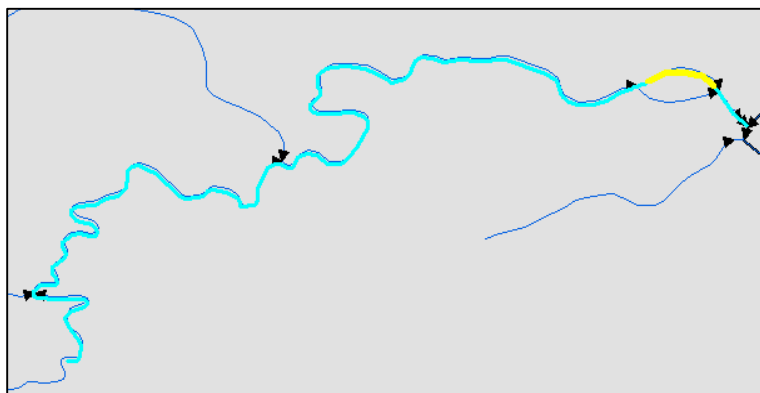
Solution: No solution identified. There is no way to automatically identify this issue as a partial failure.



**Figure 21:** If a linear Assessment Unit (Event) follows the shoreline of a two-dimensional stream and loops out around a tributary before rejoining the main stream, the Linear-to-Catchment process may miss catchments related to the tributary. The tributary is on a different level path, so the catchment is excluded from the output file. The dotted line is the NHD stream centerline. The solid line is the state Assessment Unit. The polygons are the catchments—the shaded catchments were included in the output and the white catchments were excluded. The catchment for the portion of the Assessment Unit in blue is excluded from the output file.

**Issue 3:** Event follows minor path of a divergence and connects back into the LPI. There is no foolproof way to identify this or deal with it. The section that is added during this step will follow the main path of the divergence while the event may follow the minor path of the divergence.

Solution: No solution identified. There is no systematic way to automatically identify this issue as a partial failure.



**Figure 22:** If a linear Assessment Unit (Event) includes a stream divergence, the Linear-to-Catchment process selects the catchments along the path that is designated as the main path in the NHD. In this example, the state Assessment Unit follows the divergent path shown in yellow, which is designated in NHD as the minor path. The process chooses the catchment associated with the main path (the thin blue line below the yellow highlighted line), so the process would have selected the wrong catchment in this example.

## **Area Waterbody Event NHDPlusV2 (Waterbody-to-Catchment) Specification**

### **Specification for Linking State Area Waterbody events such as Lakes, Ponds, or Reservoirs to NHDPlusV2 Catchments**

Note: Specification is based on review of “Lakes, Sounds, Harbors” area events from Alabama, Colorado, Florida, Idaho, and Kentucky.

Note: All areas should be computed using USGS Albers projection. The specification document includes technical steps for the process. *Additional information or explanation regarding a step is provided in italics.*

Input Properties:

Input:

Temp folder path (input form)  
Name/Location of Catchments shapefile (input form)  
Name/Location of input event shapefile (input form)  
Location for output dbf named <inputfilename>\_2Catch (input form)  
Name/Location for output failed shapefile (input form)  
PctXX  
PctYY

Development environment:

.NET  
Microsoft SQL Server 2012 Localdb  
ArcGIS 10.1  
Windows 7 virtual machine

#### **A. Manual Step: Load state’s waterbody events into the standard area Event feature class**

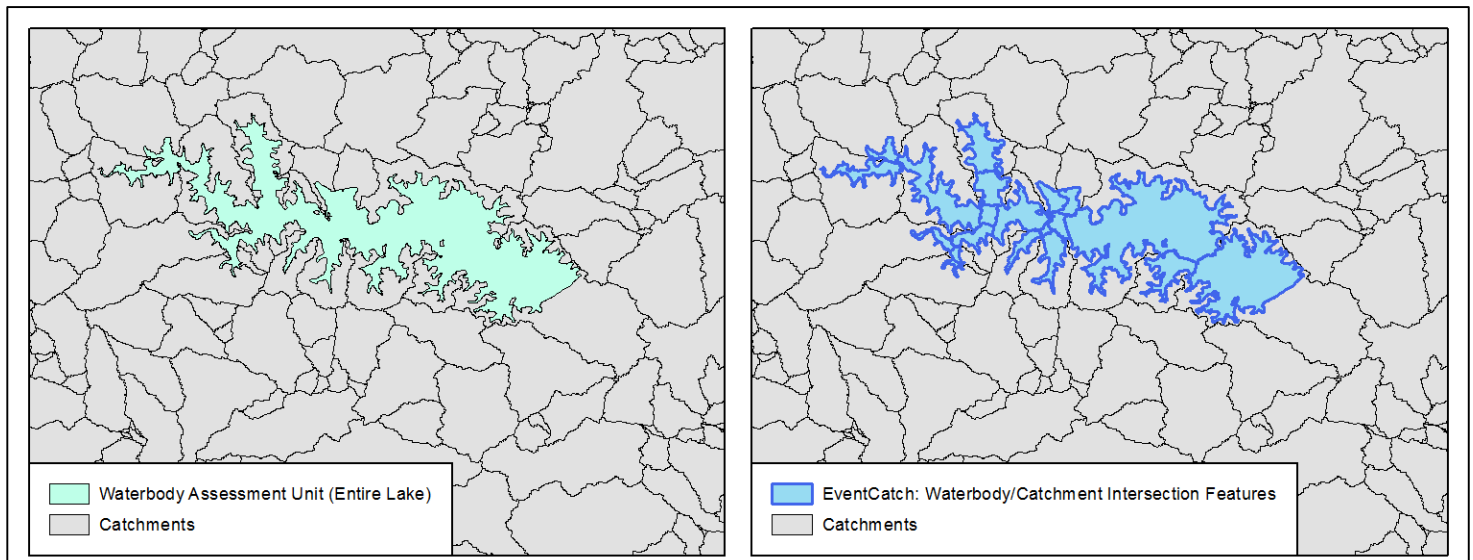
**(NHDAreaEvent Feature Class) format** associating corresponding fields as needed. This step places the state data into a common table format that can then be run through the automated process. Add the following fields. Some fields will be populated later in the process:

1. EventID [Text(100)] – If STOrigID is unique, populate EventID with STOrigID, else create unique values for EventID.
2. STOrigID [Text(100)] – State Original ID: populate with state’s event id.
3. EvtArea [Double] – area of event in square meters. Field will be populated in Step B.
4. EvtCatArea [Double] – Area of event/catchment part in square meters. Field will be populated in Step C.3.
5. EvtCatPct [Double] – Percent of event in an Event/Catchment piece. Field will be populated in Step C.4.
6. CatEvtPct [Double] – Percent of Catchment in an Event/Catchment piece. Field will be populated in Step C.5.
7. CatArea [Double] – Area of catchment in square meters. This field is added to the Catchment feature class. Field will be populated in Step C.1.

#### **B. Compute EvtArea.**

C. Implement the following steps for all waterbody features:

1. In Catchment feature class, compute area of catchments: record that in the **CatArea** field in the same units as “EvtArea”.
2. Perform “Identity” of Events with Catchments creating **EventCatch** feature class. Figure 23 shows an example.



**Figure 23: After the "Identity" function in Step C.2 of the Waterbody-to-Catchment process, the area Waterbody Assessment Unit (Entire Lake) in the left image is broken into pieces corresponding to the portions of the Assessment Unit that fall into each catchment (creating EventCatch) in the right image.**

3. Compute area of each EventCatch feature (i.e. Event/Catchment pieces): - record that in the **EvtCatArea** field.
4. Compute % of event in an Event/Catchment piece:  
**EvtCatPct = EvtCatArea\*100/EvtArea.**
5. Compute % of Catchment in an Event/Catchment piece:  
**CatEvtPct = EvtCatArea\*100/CatArea.**

6. **Keep features in EventCatch (i.e. Event/Catchment pieces) where**  
a. **EvtCatPct >= xx% or CatEvtPct >= yy%**

**Recommendation: xx = 1% and yy = 50%.** Figure 24 (on page 46) illustrates this concept.

*EvtCatPct (xx%) >= 1%*

*This is the portion of the Assessment Unit (Event) that falls within the catchment. If it is greater than or equal to 1% of the entire size of the Assessment Unit, then the catchment is included in the output. The small threshold of 1% was used so that long slender polygons such as two-dimensional streams could be corresponded to catchments without losing connecting pieces.*

*CatEvtPct (yy%) >= 50%*

*This is the portion of the catchment that is covered by the piece of the Assessment Unit (Event) that falls within the catchment. If the piece of the Assessment Unit within that catchment covers greater than or equal to 50% of the catchment size, then the catchment is included in the output.*

Note: These percentages may need to be adjusted based on further testing.

**Or**

- b. **The catchment holds an NHDFlowline of type Artificial Path.**

The resulting catchments associated with the area waterbody Assessment Unit are shown in Figure 25, on the next page.

### **Recommended Enhancements:**

One issue was found for the Waterbody-to-Catchment process, and was addressed by adding an additional criterion under Step C.6, to retain catchments associated with Artificial Paths. However, the artificial path criterion may not work with new reservoirs in state datasets that are not coded with Artificial Paths in NHDPlus.

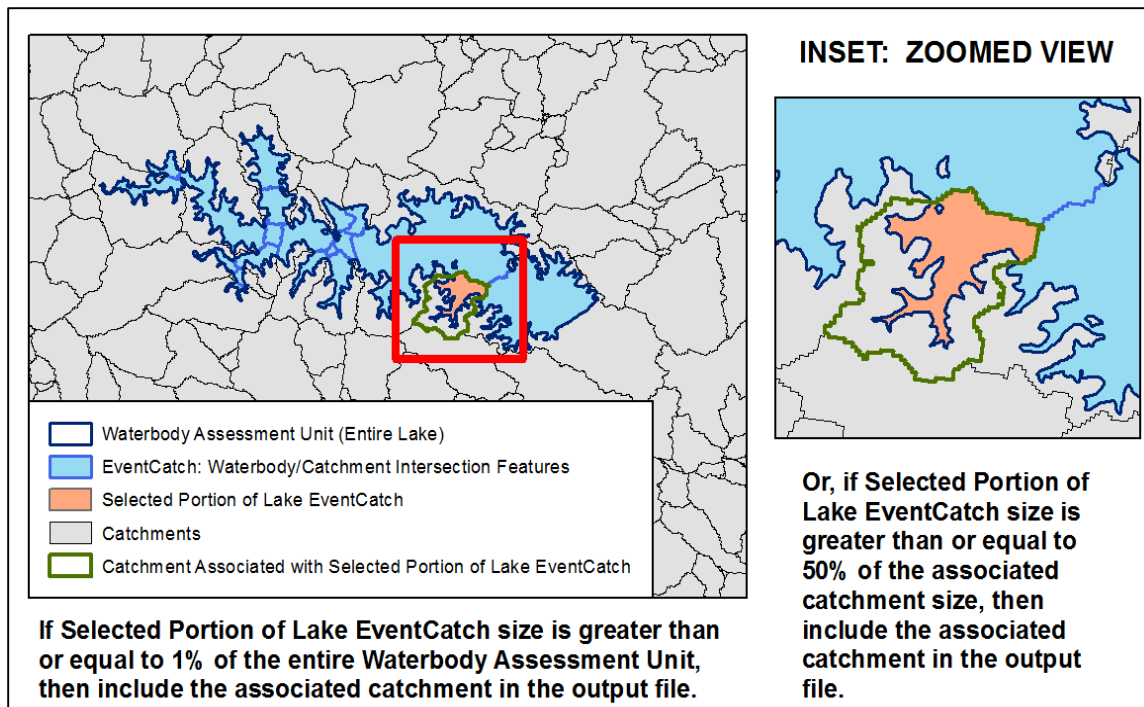


Figure 24: For an area waterbody event such as a lake, a catchment is included in the output file when any of the following are true: 1) the portion of the Assessment Unit (Event) that falls into the catchment is at least 1% of the overall Assessment Unit size (as shown in the left image); or 2) the portion of the Assessment Unit in the catchment covers at least 50% of the catchment size (as shown in the right image); or 3) the catchment containing part of the Assessment Unit is associated with an artificial path from the NHD Flowline (not shown).

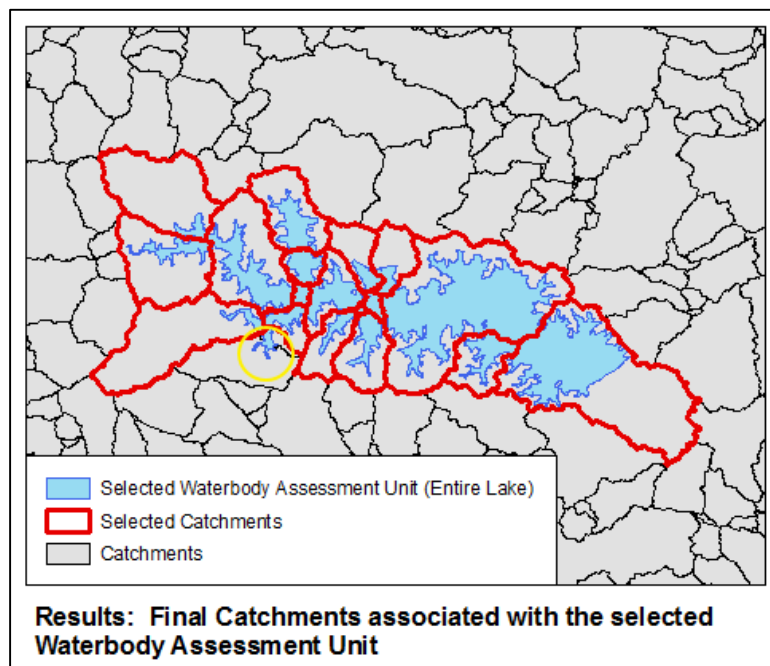


Figure 25: This figure shows the catchments associated with the example area waterbody Assessment Unit (Event) using the area Waterbody-to-Catchment process. The waterbody crosses into two catchments in the lower left (circled in yellow) that are not included because the portion of the waterbody in each of those catchments is lower than 1% of the overall assessment unit size, and the portion of the waterbody covers less than 50% of each of those catchments. The artificial path rule does not apply here because this was a new reservoir that was not included in NHDPlus, so the catchments here are based on the stream that was present before the reservoir was created.

## **HUC-Like Event NHDPlusV2 (HUC-to-Catchment) Specification**

### **Specification for Linking State Events that Represent HUCs or Similar Land Based Polygons (HUC-Like) Events to NHDPlusV2 Catchments**

Input Properties:

Input:

Temp folder path (input form)  
Name/Location of Catchments shapefile (input form)  
Name/Location of input event shapefile (input form)  
Location for output dbf named <inputfilename>\_2Catch (input form)  
Name/Location for output failed shapefile (input form)  
PctXX  
PctYY

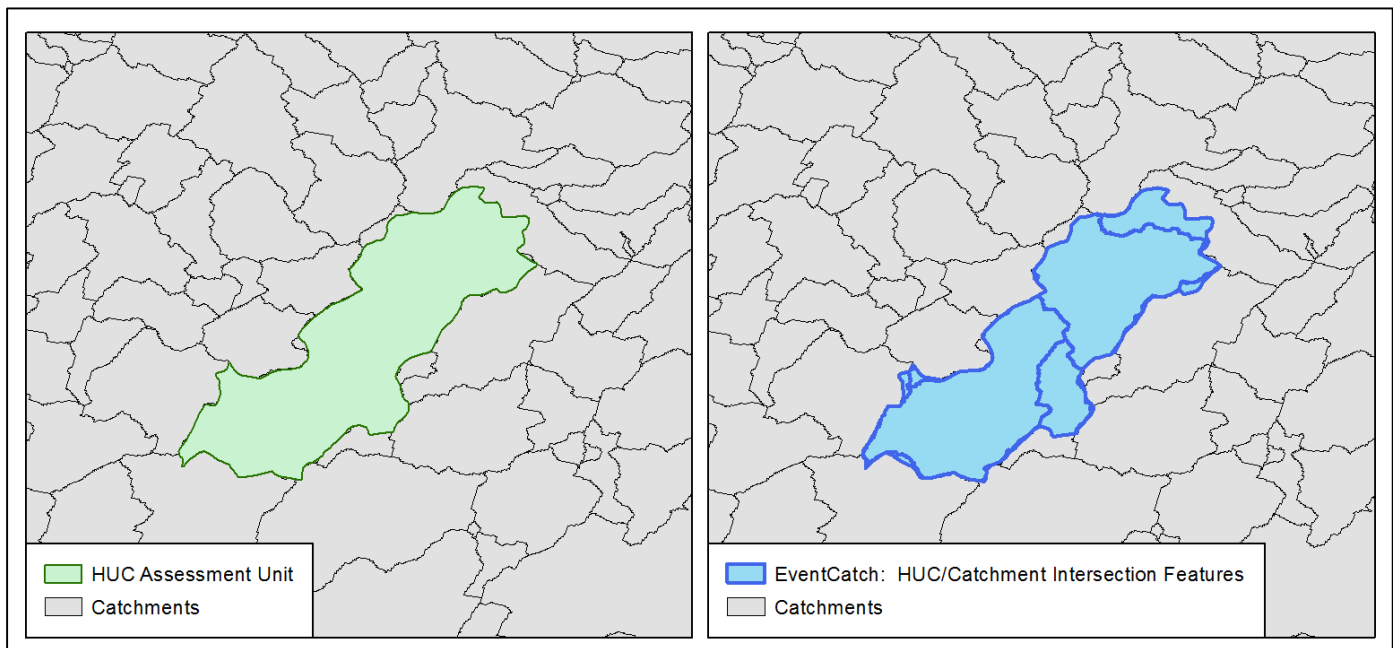
Development environment:

.NET  
Microsoft SQL Server 2012 Localdb  
ArcGIS 10.1  
Windows 7 virtual machine

Note: Specification is based on review of HUC-like events from New Jersey (HUC16) and Utah (HUC10 and HUC12) and Alabama.

- A. **Manual Step: Load state events into the area Event feature class** associating corresponding fields as needed. This step places the state data into a common table format that can then be run through the automated process. Event feature class contains:
1. EventID [Text(100)] – If STOrigID is unique, populate EventID with STOrigID, else create unique values for EventID.
  2. STOrigID [Text(100)] – Populate with state’s event id.
  3. EvtArea [Double] – Area of event in square meters. Field will be populated in Step B.
  4. CatArea [Double] – Area of catchment in square meters. This field is added to the Catchment feature class. Field will be populated in Step C.
  5. EvtCatArea [Double] – Area of the event/catchment part in square meters. Field will be added in Step E and populated in Step F.
  6. EvtCatPct [Double] – Percent of event in an event/catchment piece. Field will be added in Step E and populated in Step G.
  7. CatEvtPct [Double] – Percent of catchment in an event/catchment piece. Field will be added in Step E and populated in Step H.
- B. **Compute area of events** and record that in the **EvtArea** field.
- C. **In Catchment feature class, compute area of catchments** and record that in **CatArea** field in same units as EvtArea.

D. Perform “Identity” of Events with Catchments, creating **EventCatch** feature class. Figure 26 provides an example.



**Figure 26:** After the "Identity" function in Step D of the HUC-to-Catchment process, the HUC Assessment Unit (Event) in the left image is broken into pieces corresponding to the portions of the Assessment Unit that fall into each catchment (creating EventCatch) in the right image. This example uses a HUC14 from New Jersey.

E. **Add fields:** EvtCatArea, EvtCatPct, CatEvtPct, all of field type Double.

F. **Compute area** of each Event/Catchment pieces, and record in EventCatch.EvtCatArea.

G. **Compute percent of event** in an Event/Catchment piece:

$$\text{EventCatch.EvtCatPct} = \text{EventCatch.EvtCatArea} * 100 / \text{EventCatch.EvtArea}.$$

H. **Compute percent of catchment** in an Event/Catchment piece:

$$\text{EventCatch.CatEvtPct} = \text{EventCatch.EvtCatArea} * 100 / \text{EventCatch.CatArea}.$$



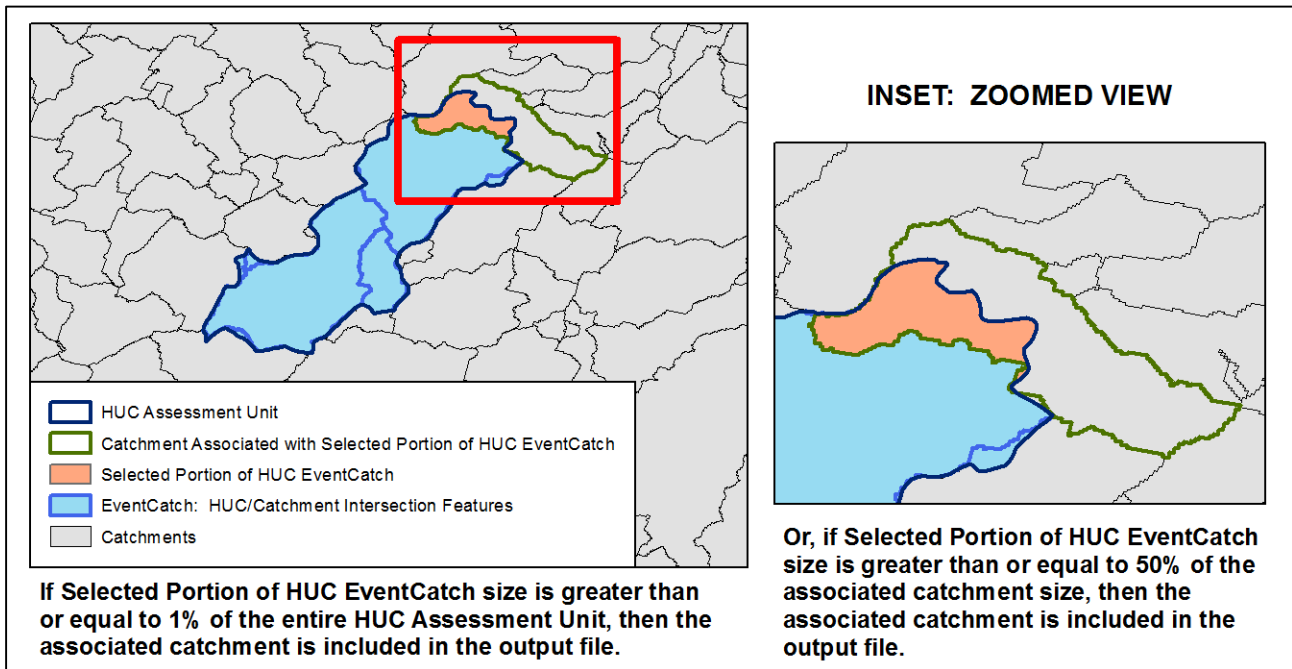


Figure 27: For a land-based area Assessment Unit (Event) such as a watershed or Hydrologic Unit (HUC), a catchment is included in the output file when either of the following are true: 1) the portion of the Assessment Unit that falls into the catchment is at least 1% of the overall Assessment Unit size (EvtCatPct) as shown in the left image; or 2) the portion of the Assessment Unit in the catchment covers at least 50% of the catchment size (CatEvtPct) as shown in the right image.

I. **Keep Event/Catchment pieces where**

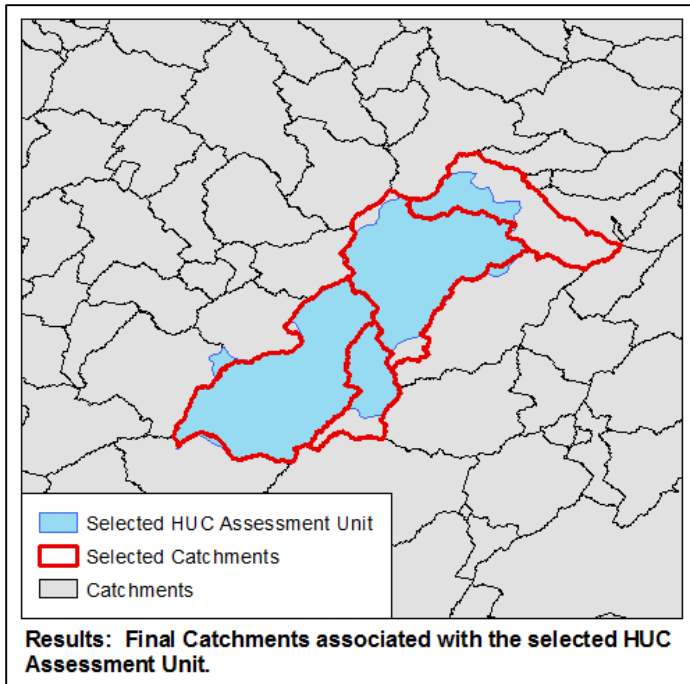
EventCatch.EvtCatPct >= PCTxx% or EventCatch.CatEvtPct >= PCTyy%

Current recommendation: PCTxx = 1% and PCTyy = 50%

*These thresholds were selected to match the thresholds used for the area Waterbody-to-Catchment process. The difference between the HUC and Waterbody-to-Catchment protocols is that the HUC process does not reference Artificial Paths. So the HUC-to-Catchment process only relies on the threshold percentages.*

**Percentages may need to be adjusted based on further testing.**

Figure 27, on the next page, illustrates this concept. The resulting catchments associated with the example HUC Assessment Unit are shown in Figure 28.



**Figure 28:** This figure shows the catchments associated with the example HUC Assessment Unit (Event) using the area HUC-to-Catchment process.

**There are multiple types of HUC-to-Catchment situations, and the quality of the results may vary.**

Case 1: This occurs where one HUC Assessment Unit overlaps with multiple catchments. This is the most common case. The example used in the figures above illustrates this case.

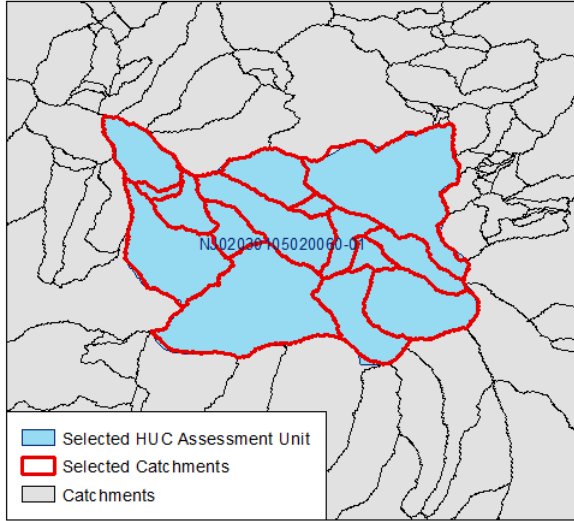
Case 2: A second case occurs where there is a large catchment, and one or more smaller HUC Assessment Units. Figure 32 (on page 51) illustrates this case.

**Recommended Enhancements:**

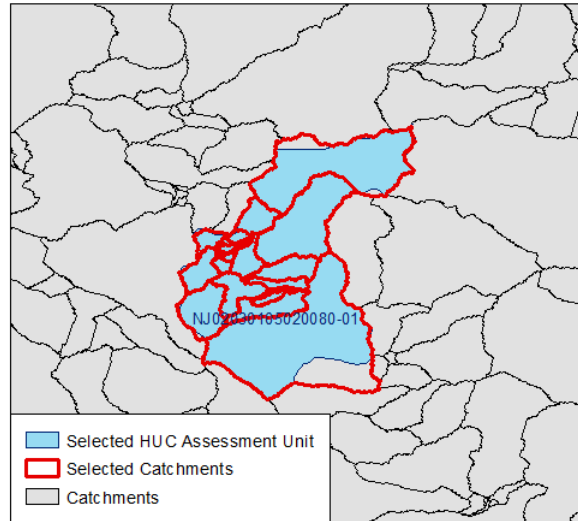
The tolerance level on the threshold percentages used in Step I of the HUC-to-Catchment process needs to be further explored to determine whether adjustments may improve the results.

*Additional HUC Examples*

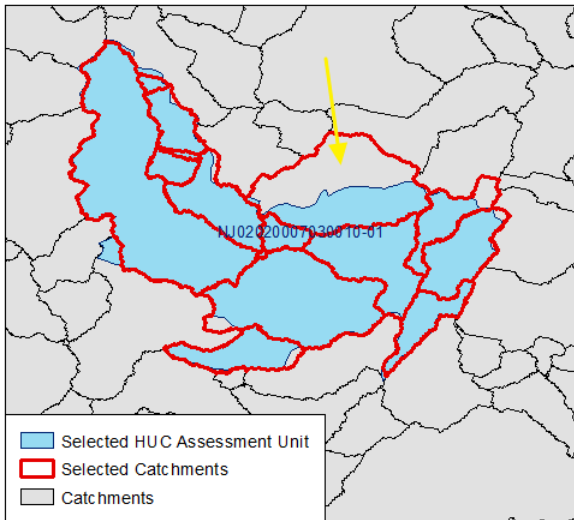
The HUC-to-Catchment process works well for HUC12s as shown in Figure 10 (on page 29), because the Watershed Boundary Dataset HUC12s were used in the initial creation of NHDPlus catchments. However, for HUCs that are less standardized such as HUC14s or other state-defined boundary polygons, the HUC-to-Catchment process may be less accurate. The examples illustrated in Figure 26 through Figure 34 represent the correspondence between HUC14s from New Jersey and catchments. For some HUC14s, the HUC-to-Catchment process works fairly well, as illustrated in Figure 28, above, as well as Figure 29 and Figure 30, below. In most cases, HUCs are larger than catchments and cover multiple catchments, as shown in the previous examples. While Figure 31 is less ideal than the previous examples, the results are still good, since the HUC covers about half of the catchment with the yellow arrow. Figure 32 shows a case where a catchment is larger than a small HUC. However, there are some areas where the process needs to be refined in order to improve the results, as shown in Figure 33 and Figure 34 (on page 52).



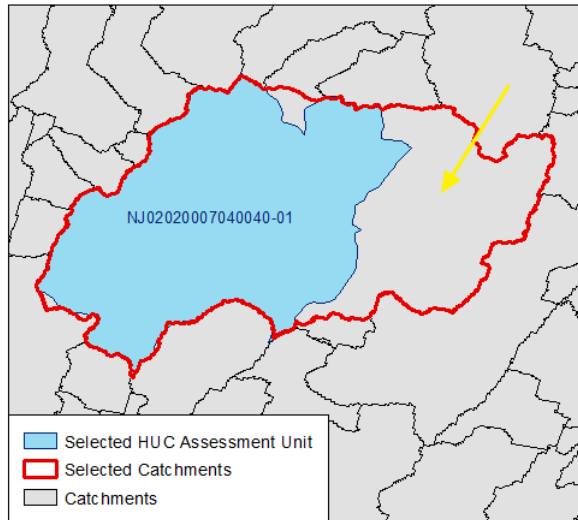
**Figure 29:** This figure shows a HUC14 Assessment Unit that corresponded well with the catchments, aligning almost perfectly.



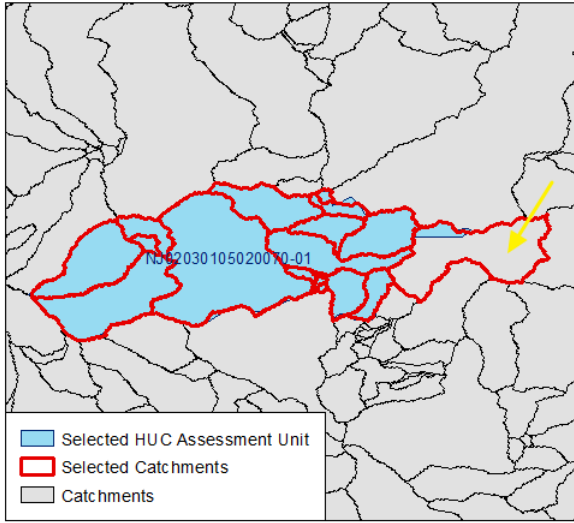
**Figure 30:** This figure shows another HUC14 Assessment Unit that corresponded well with the catchments. The alignment is not as close as with the previous figure, but the catchments still show a good representation of the HUC14.



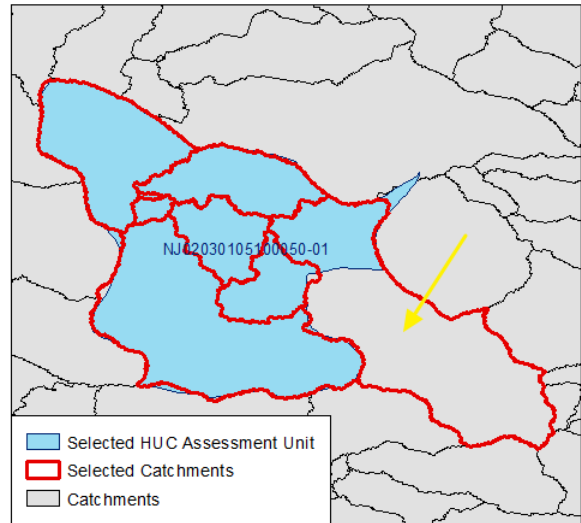
**Figure 31:** This figure shows a HUC14 Assessment Unit that corresponded fairly well with the catchments. Visually, it may seem incorrect that the catchment with the yellow arrow is included in the output. However, that catchment is about half covered by the HUC Assessment Unit. That catchment is also included because the portion of the Assessment Unit within the catchment is more than 1% of the total Assessment Unit size. While that catchment only needs to meet one of the criteria, it meets both threshold criteria. Changes to the threshold rules may not change the results in cases similar to this.



**Figure 32:** While most catchments are smaller than HUCs, this figure shows a catchment that is larger than a HUC14.



**Figure 33:** This figure shows a HUC14 example that corresponded well except for one catchment. The catchment with the yellow arrow seems to be extra and unnecessary, visually. This catchment is included because the piece of the HUC Assessment Unit within that catchment is at least 1% of the area of the entire HUC Assessment Unit. Changes to the thresholds or logic rules may improve the results in cases similar to this.



**Figure 34:** This figure shows another HUC14 example that corresponded well except for one catchment. This is another example where the catchment with the yellow arrow seems to be extra and unnecessary, visually. This catchment is included because the piece of the HUC Assessment Unit within that catchment is at least 1% of the area of the entire HUC Assessment Unit. Changes to the thresholds or logic rules may improve the results in cases similar to this.

**Potential refinements for the HUC-to-Catchment process include:**

- 1) Increasing the threshold for the percent of the Assessment Unit piece within the catchment (perhaps from 1% to 10%),
- 2) Changing the logic rules so that both specifications must be met (the piece must be at least 1% of the total Assessment Unit size AND cover at least 50% of the catchment to be included).

## Appendix D: EPA Implementation Considerations

As mentioned in Section 9, Implementation Considerations, the EPA Office of Water geospatial software and database infrastructure will need to be examined in light of any new requirements presented by Catchment-based indexing, including the retention of state-submitted geospatial data for display in conjunction with their Catchment-based representation. This appendix discusses some specific items that should be considered as part of that process.

One approach would be to load all of the state GIS coverages as custom events in the Reach Address Database (RAD) component of Watershed Assessment Tracking and Environmental Results (WATERS). This change in approach will require that EPA's geospatial infrastructure undergo some modifications in order to accommodate this new approach.

Some items that will need to be considered by the RAD team include:

1. What would the table structure be for 'catchment' events? One approach could be to use the same event table structure as exists for the linear events, but not include the 'from' and 'to' measure columns.
2. How would catchment events be included in existing WATERS services?
3. What are the additional considerations in displaying non-NHD events? In many cases these events will be at a higher resolution than the NHDPlus. All of EPA's mapping applications are optimized to the medium resolution NHDPlus. How will these applications need to adjust to accommodate a mixed resolution?
4. How does EPA explain mixed resolution data, especially across state boundaries?
5. We need an approach for including off-shore events in programmatic measures.

The RAD team should discuss these issues in consultation with the Georeferencing Pilot team prior to any implementation of the new simplified catchment approach.

In addition to these issues in the RAD, some Data-related considerations include addressing the current lack of NHDPlus data for Alaska, and accommodating near coastal assessment information where catchment do not exist.

1. EPA is currently working with Alaska, for which NHDPlus catchments are not yet developed, and discussions are underway to determine an interim solution.
2. Coastal catchments only reflect the landscape that flows to a coastline segment, and would not accurately reflect an off-shore event. Since most of these events are area events, they should not be associated with a catchment, but rather the area being represented by the event would be used for any programmatic measures purposes. Additional discussion about how to best capture these types of events for programmatic measures purposes should be part of the programmatic measures discussion that EPA is having with the states in 2014 as part of the Water Quality Framework ATTAINS Redesign workgroups.
3. The simplified catchment approach that was tested under this Pilot did not test a methodology for point events; however, a simple intersect between point events and catchments could be performed with fairly little difficulty, and this intersect would not require any of the additional logic that has been applied for the linear or areal processes.

To implement the simplified Catchment-based indexing approach, there are several options that EPA could pursue. Most of the process can be automated with manual steps at the beginning of the process to get the data in the right format, and manual steps at the end of the process to QA/QC the output and address any events that did not migrate properly. For the first implementation of this approach, this semi-automated process should be run by EPA and EPA contract staff. The potential exists to develop the process as a service that states could run independent of EPA, but it would be difficult to implement this at the initial implementation outset. For the initial implementation, there are a few options that could be explored:

1. Use Esri ArcGIS: The simplified catchment methodology was developed and tested using ArcGIS functionality. EPA would need to finalize this code and make it production ready, and add in the ability for a user to QA/QC the output.
2. Use spatial database services: Oracle Spatial was used to develop and test the other Pilot options described in this report. PostGIS (a component of PostgreSQL, an open-source database) could also be used to develop this capability.

## Appendix E: Information on Initial Pilot Options

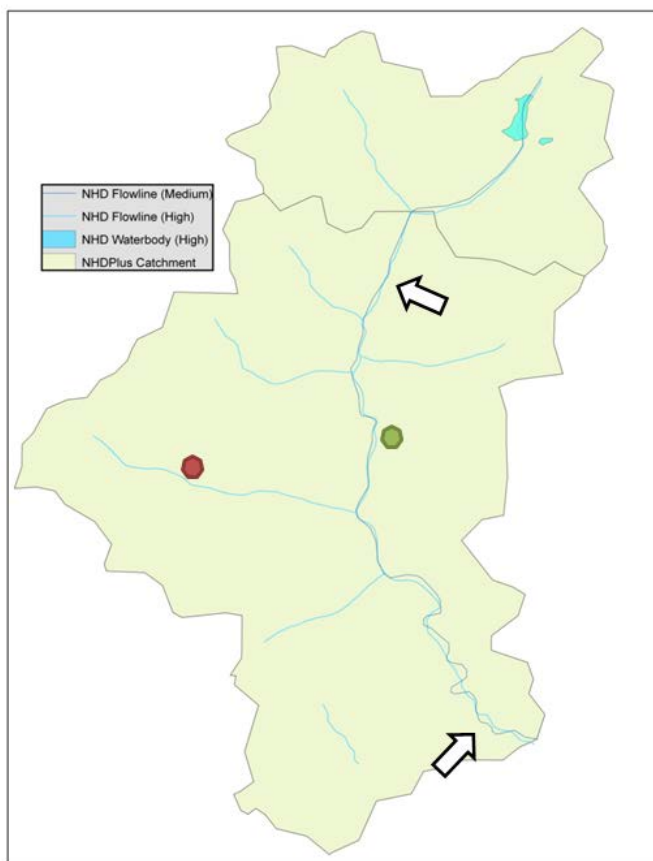
The Pilot originally began with three options: Point-based indexing, Watershed Polygon-based indexing, and Feature-based indexing. Those options are described briefly in Section 6, Initial Set of Pilot Options Explored. It was determined that none of those options were feasible to implement beyond the Pilot. However, details of those options can be found here.

### Point-based Indexing Details

Point-based indexing looked at Latitude and Longitude coordinates that represent the state's monitoring locations, which may be used to help make a water quality assessment decision for the relevant Assessment Unit. For this option, Pilot Workgroup states were asked to provide a table of information, rather than geospatial data. The idea was that states could provide tabular information as part of their water quality assessment attribute data, instead of submitting a separate geospatial data file. In order for this option to work, states needed to designate the target resolution (high resolution or medium resolution) and provide enough information in the table to allow EPA to use automated tools to associate the geographic coordinates with the appropriate hydrographic feature. The table needed to include columns containing the following information:

1. Latitude and Longitude Coordinates representing the monitoring location
2. Direction to navigate (Upstream, Downstream, Both Upstream and Downstream, Do not navigate)
3. Distance to navigate Upstream (if applicable)
4. Distance to navigate Downstream (if applicable)
5. The Assessment Unit ID (or Listing ID)

The target resolution was important because high resolution NHD contains more tributaries and small streams than medium resolution NHDPlus. One of the NHDPlus tools allows for upstream and downstream navigation. Navigating upstream and downstream, events on tributaries may or may not be of interest to users. For example, if a point is associated with a tributary in high resolution NHD that is not visible in medium resolution, a user may not want navigation tools to discover that point when navigating downstream on the main stem. Alternatively, a user may want to discover events on tributaries when starting from a downstream point and navigating upstream to all of the waters that eventually flow to that downstream point. Figure 35 provides an example.



**Figure 35:** With Point-based indexing, it is important to know whether a point is associated with a tributary on high resolution NHD or the mainstem on medium resolution NHDPlus. If a user was navigating downstream from the upper arrow, he would likely only want to discover the green point. If a user was navigating upstream from the lower arrow, he would likely want to discover both the green point and the red point because both streams flow to the lower arrow.

## **Watershed Polygon-based Indexing Details**

Watershed polygon-based indexing looked at polygons representing watershed boundaries that are used as Assessment Units by some states. In many cases, these refer to Hydrologic Unit Code boundaries at the 12-digit level (HUC12s), but it can refer to a state-defined watershed polygon framework as well. Initially, EPA planned to use the watershed polygon boundary to extract the underlying hydrology from NHDPlus, so that states could submit just the watershed boundaries instead of requiring states to georeference the linear streams within those watershed boundaries. For this option, states that use watershed boundaries as Assessment Units were asked to submit the watershed boundary polygons and designate the target resolution (high resolution or medium resolution).

## **Feature-based Indexing Details**

Feature-based indexing looked at Assessment Units that were based on linear features such as rivers and streams or area waterbody features such as lakes, ponds or reservoirs. For this option, states were asked to submit geospatial files representing geographic extents of the state's water quality assessment decisions and impaired waters listing decisions, and designate the target resolution (high resolution or medium resolution).

Feature-based indexing aimed to link high resolution and medium resolution state-submitted features to individual NHD features in NHDPlus. Since state geospatial data could reference high resolution NHD which does not have the additional hydrologic attributes that are part of NHDPlus and used by various EPA programs and services, one pre-processing step involved establishing a correspondence between the high resolution NHD snapshot and the medium resolution NHDPlus snapshot. More specifically, Feature-based indexing took advantage of the common content (reach codes, network relationships, stream names, etc.) shared by high resolution NHD and medium resolution NHD. These common NHD fields help establish a correspondence between EPA's high resolution NHD snapshot and the medium resolution NHD snapshot in NHDPlus. Any state-submitted high resolution geospatial data was first overlaid on EPA's high resolution NHD snapshot and then related to medium resolution NHD using this High-Resolution-to-Medium-Resolution NHD correspondence. This approach included attempting to identify the subset of high resolution features that corresponded to each medium resolution feature. State-submitted medium resolution geospatial data would still be directly overlaid on EPA's medium resolution NHD (in NHDPlus).

## **Elimination and Refinement of Options**

Before the technical part of the Pilot was initiated for the suggested inputs and outputs, the team did preliminary screening of input and output options to see if any of them should be eliminated. As part of the preliminary screening, the states participating in the Pilot workgroup were asked to fill out a set of questions that helped to understand how their Level of Effort (LOE) and willingness to submit data types may change with each input option. Appendix F: Questions to Pilot Workgroup Regarding Initial Pilot Options includes a copy of questions provided to the states. Table 3 below provides the summary of the responses.

The responses from Pilot states regarding their state's LOE and willingness to submit data for the different options demonstrated that even though there might be an initial increased LOE associated with submitting High resolution NHD features, the majority of the states are interested in this option. Based on the responses from Pilot workgroup members, the team chose to eliminate the Point-based indexing option from the Pilot. The responses also indicated that, if EPA had to require states to submit only one type of data input (such as points or polygons) states' willingness to submit the data would dramatically decrease. It was determined that to maximize state data submissions and to minimize state burden, EPA should accept any data format the state is using for their own purposes. Therefore, the Pilot focused on using a mix of data formats and scales that states submit to EPA. Since majority of the states submit features, with a few submitting watershed polygons, the automation needed to accept high resolution or medium resolution inputs of those types. There is already an automated process available to georeference points for those states that do choose to submit points, though it does not draw the extents of Assessment Units from attribute data as originally intended under the Point-based indexing option. Previous



projects have used that existing process for batch indexing points or Latitude and Longitude coordinates to NHDPlus, so no further work was done using points for this Pilot.

**Table 3: Summary of Pilot Workgroup Participants' Willingness to Submit Data for the Original Pilot Options and Estimate of State Level of Effort (LOE) Required to Format and Submit Data for each Option**

Pilot Option (Format)	Number of Pilot States Willing to Submit in Format	Estimated LOE Change From the Most Recent Reporting Cycle Submission to Future Data Submission*	
		First Reporting Cycle	Subsequent Reporting Cycle
<b>High Resolution NHD Features</b>	10 out of 11 states would submit	Average <u>increase by 65% **</u> (range -60% to 300%)	Average <u>increase by 15% **</u> (range -80% to 233%)
<b>Medium Resolution NHD Features</b>	3 out of 11 states would submit	Average <u>increase by 274%</u> (range 47% to 500%)	Average <u>increase by 104%</u> (range -5% to 300%)
<b>Non-NHD Features</b>	3 out of 11 states would submit	Average <u>decrease by 27%</u> (range -76% to 0%)	Average <u>decrease by 53%</u> (range -88% to -5%)
<b>High Resolution Points</b>	2 out of 11 states would submit	Average <u>increase by 67%</u> (range 33% to 100%)	Average <u>increase by 17%</u> (range -67% to 100%)
<b>Medium Resolution Points</b>	2 out of 11 states would submit	Average <u>increase by 298%</u> (range -5% to 600%)	Average <u>increase by 198%</u> (range -5% to 400%)
<b>Polygons</b>	5 out of 11 states would submit	Average <u>increase by 4%</u> (range -76% to 100%)	Average <u>decrease by 12%</u> (range -88% to 100%)

\* Average is calculated only for those states that will submit data for the Pilot option.

\*\*Increase for high resolution Features could be related to states initially switching their georeferencing from medium resolution to high resolution.

During the Feature-based indexing, the step to establish the correspondence between the high resolution and medium resolution snapshots required a high level of processing time. As a result, it took about one week of computer processing time to process one 8-digit HUC using the Feature-based indexing option. That processing time was determined to be too long to implement Feature-based indexing beyond the Pilot. So, EPA looked at simplifying the process to georeference to the NHDPlus catchments rather than to the underlying hydrologic features. That refinement was rolled into the Catchment-based indexing approach.

## Appendix F: Questions to Pilot Workgroup Regarding Initial Pilot Options

The Pilot Workgroup participants were asked to consider the original Pilot options, and respond to questions regarding their willingness to submit data for each Pilot option and estimate the Level of Effort (LOE) required for them to format and submit data for each option. This Appendix details the questions that were posed to the Pilot Workgroup. The responses from the Pilot Workgroup participants, which are summarized in Table 3 (on page 57), affected the Pilot design.

### INTRODUCTION

When thinking about the responses to the questions below, please consider all steps that involve geospatial data processing and review done for the purpose of submitting the geospatial data to EPA. You may use the worksheet below and optionally fill in your estimates of hours for each individual step in the process.

WORKSHEET (Optional): This information is intended to assist states when responding to the questions below.

The steps to consider include, but are not limited to:

1. Generate or process geospatial data for submittal to EPA. [\* - Please consider only what it took to put data into the format required for EPA submission. Some states create geospatial data for their own needs and give the data to EPA as is – in this case changing it to an EPA format has not been done and the estimate of effort should be minimal.] \_\_\_\_\_
2. Perform QA/QC on the geospatial data for EPA submittal. \_\_\_\_\_
3. Make edits to the geospatial data based on the EPA Regional review of the 303(d) list submittal. \_\_\_\_\_
4. Submit the geospatial data to EPA. \_\_\_\_\_
5. Provide additional information if there are gaps based on the feedback received from the contractor after processing the geospatial data. \_\_\_\_\_
6. Any other steps related to processing, submitting, or editing geospatial data not mentioned in #1-5. \_\_\_\_\_

### QUESTIONS

- A. State: \_\_\_\_\_
- B. Based on your experience with the submission process for the most recently completed data cycle
  - a) Please estimate the total number of hours it took to perform steps 1 through 6 in the worksheet above. \_\_\_\_\_
  - b) Total hours aside, over what time period were steps 1-6 completed (days, weeks, months)? \_\_\_\_\_
  - c) Does your state typically carry out tasks in step #5 or is it usually done by your corresponding EPA Region? \_\_\_\_\_
  - d) If you included any additional steps in #6 in your estimate, what were they?  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

- e) Did the cycle you based your answers on involve (select as many as are applicable):
- First-time original development of the geospatial dataset in this cycle
  - Major changes in data format (e.g., re-segmentation, hi-res/medium res change)
  - Significant number of new waters added
  - Significant changes in waters attributes
  - Minor changes in waters attributes
  - Other (write in): \_\_\_\_\_

f) Which cycle year are you using for the answers in (a) – (e)? \_\_\_\_\_

C. Consider steps 1-6 in the worksheet above and provide your best judgment estimate of the hours to submit your data in the formats considered in the pilot\* (listed across in the table below) in the **FIRST** cycle year the change (if any) would take place and in **SUBSEQUENT** cycle years (e.g., 2<sup>nd</sup>, 3<sup>rd</sup> ). If you think your state would choose not to submit data in a certain format, please mark the relevant box with “X”.

		High Resolution NHD Features	Medium Resolution NHD Features	non-NHD Features	High Resolution Points	Medium Resolution Points	Polygons
Number of hours	<b>FIRST</b> cycle						
	<b>SUBSEQUENT</b> cycle						