Environmental Modelling and Software 135 (2021) 104927

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Mainstems: A logical data model implementing *mainstem* and *drainage basin* feature types based on WaterML2 Part 3: HY Features concepts

David Blodgett [a,*], J. Michael Johnson [b], Mark Sondheim [c], Michael Wieczorek [a], Nels Frazier [d]

[a] U.S. Geological Survey, 12201 Sunrise Valley Dr, Reston, VA, 20192, USA
[b] University of California, Santa Barbara Department of Geography 1832 Ellison Hall UC Santa Barbara Santa Barbara, CA, 93106-4060, USA
[c] Natural Resources Canada 506 Burnside Rd W, Victoria, BC V8Z 1M5, Canada
[d] National Oceanic and Atmospheric Administration (NOAA) Affiliate, CyberData Technologies, Office of Water Prediction, National Weather Service, National Water Center 205 Hackberry Ln, Tuscaloosa, AL, 35401, USA

ARTICLE INFO

Keywords:
Hydrography
Mainstem
Drainage basin
Logical data model
Hydrologic location

ABSTRACT

The Mainstems data model implements the *catchment* and *flowpath* concepts from WaterML2 Part 3: Surface Hydrology Features (HY_Features) for persistent, cross-scale, identification of hydrologic features. The data model itself provides a focused and lightweight method to describe hydrologic networks with minimum but sufficient information. The design is intended to provide a model for data integration that can be used for network navigation and persistent hydrologic indexing (hydrographic addressing) functionality. Mainstems is designed to provide long-term stability with minimal maintenance requirements. The data model is not meant to advance hydrologic process representation or uniquely represent geomorphic characteristics. The principle assumption in Mainstems is that all *drainage basins* have one - and only one - *headwater* source area and a single *mainstem* that flows to a single *outlet*. Using these base feature types, (*headwater*, *outlet*, *mainstem*, and *drainage basin*) a nested set of *drainage basins* - and the associated dendritic network of *mainstems* - can be identified.

1. Introduction

Petts (1996) describes *drainage basin*, as a basic hydrologic unit of the landscape that encompasses a “cascading system” of connected “hillslope and channel subsystems”. As early as the 1500s, it was recognized that rain accumulates in catchment areas to form rivers and springs (Biswas, 1970). In the 1600s, Edme Marriotte mapped the Seine River upstream of Paris including its drainage basin boundary, flowpath network, and “source” (Dooge 1959). Marriotte’s work included calculations of “catchment area”, mean annual rainfall, and total annual rainfall volume. Since these early notions of quantitative hydrology, a fundamental relationship between a *drainage basin* area and a predominant *mainstem* that transports water from a “source” (*headwater*) to an *outlet* has been recognized.

Nearly 400 years later, the familiar and intuitive properties of these landscape mechanics have resulted in a general vagueness about how hydrologic features are described, referenced, and understood. In the computer age, where knowledge must be expressed explicitly for machine interpretation and learning, this vagueness limits the integration and long-term compilation of knowledge (i.e., progress) in hydrographic, hydrologic, hydrodynamic, or hydrometric sciences (collectively hydroscience). Such a challenge is not newly recognized: Abbot (1993) pointed to the need to encapsulate knowledge (i.e., adding nuance and context to data) to further hydroscience integration. Archfield et al. 2015, called for a machine-independent information sharing infrastructure to enable communication and integration across hydroscience domains, and in, Jiang et al. 2019 described the potential value in using a knowledge-based approach to catalog and re-use specific application context for integrated hydroscience modeling. These examples illustrate the need for—and continuing lack of—focus in

Abbreviations: Common Hydrology Features, (CHyF); Geospatial Fabric, (GF); Hydrologic Unit, (HU); National Hydrography Dataset, (NHD); National Hydrography Dataset Plus Version 1, (NHDPlusV1); National Hydrography Dataset Plus Version 2, (NHDPlusV2); National Hydro Network, (NHN); Network Linked Data Index, (NLDI); NHDPlusV2 catchment ID, (COMID); Points of Interest, (POIs); River Reach File 1, (RF1); WaterML2 Part 3, Surface Hydrology Features, (HY_Features); Watershed Boundary Dataset, (WBD).

* Corresponding author.
E-mail addresses: dblodgett@usgs.gov (D. Blodgett), jmj00@ucsb.edu (J.M. Johnson), mark.sondheim@canada.ca (M. Sondheim), mewieczo@usgs.gov (M. Wieczorek), nels.frazier@noaa.gov (N. Frazier).

https://doi.org/10.1016/j.envsoft.2020.104927
Accepted 9 November 2020
Available online 13 November 2020
1364-8152/Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

hydroscience knowledge representation.

To start addressing the problem of vagueness in surface hydrology, the first international standard for surface water features—WaterML2 Part 3: Surface Hydrology Features (HY_Features)—was published by the Open Geospatial Consortium in 2018 (See Box 1). This standard describes building blocks for formalizing imprecise terminology in support of data integration and scientific reproducibility in the hydrosciences. (Atkinson, 2012; Blodgett and Dornblut, 2018). While HY_Features defines concepts for a common abstraction, it does not specify implementation logic. In practice, integration and reproducibility is achieved only through use of a common delineation methodology, a specific dataset, or a particular modeling software. This current state of the hydroscience field is so diverse that reproducibility and interoperability built on these foundations can be impossible without expert interpretation (Hutton, 2016).

As a purely conceptual model, HY_Features was intended to be adopted in use-specific logical and/or physical data models as described by Brodaric et al., 2018 (Box 2). The challenge with modern hydrologic applications is that many physical data models were developed before the existence of a common conceptual model. With the introduction of HY_Features, there is an opportunity to create explicit, logical models to use as a tool for unifying existing physical data models.

Toward this end, this research presents the Mainstems logical data model (shortened to Mainstems), which is built on the HY_Features concepts. The goal of Mainstems is to specify a minimum set of location, linear, and areal feature types needed to define and integrate disparate datasets using features that can be uniquely and easily identified. Mainstems is especially useful for integrating spatially precise but inexact feature geometries meant to represent networks of rivers, catchments, and locations related to observation and prediction of hydrologic phenomena. More broadly, Mainstems is intended to be implemented in multiple physical data models in support of hydroscience data integration.

1.1. Detailed problem statement and research question

All hydroscience physical (applied) data models describe hydrology through a set of concepts, logic, and constraints that are useful to specific applications and scales. Users wanting to apply the data outside of those contexts must modify - or integrate - their data in a way that respects both the geographic and hydrologic representation of the original data. In this research, *data integration* refers to the alignment of spatial features ***and*** the conflation of properties to establish matching features by hydrologic function.

These coupled needs prompt a number of questions with respect to what type of data should and should not be integrated, and the nuances that need to be considered prior to doing so:

1. Is hydrology represented in the same way in each dataset?
2. Are geometries in different datasets meant to represent the same feature?
3. Is the scale of feature types compatible?
4. Does the original intent of a dataset align with the needs of the new application?

Without common concepts, the only way to integrate datasets is through spatial proximity or alignment. Spatial proximity is a challenging integrator when using datasets that map coordinates with varying levels of accuracy (or coarseness of scale), but with high levels of precision (coordinates are exact points on the Earth).

Alternatively, the use of well-defined concepts and logic could allow data to be aligned by the features' purpose, intention, or identity. That is, that features in the various data sources would be aligned based on the "real-world" features they represent and not just their 'nearness' in an arbitrary scale.

A secondary—and perhaps more relatable—challenge that limits data integration, is that the spatial coverage and attribute richness of most authoritative datasets come at the cost of complexity and data volume. For instance, the National Hydrography Dataset Plus Version 2 (NHDPlusV2) (McKay, 2015) is a multiple gigabyte database with hundreds of inter-related attributes across millions of features. While necessary for some applications, data complexity and volume can pose a major barrier to use. Therefore, to provide a functional, integrative data model built on consistent concepts, the Mainstems model must be able to satisfy both the lack of a shared conceptual model and the complexity and volume of existing datasets.

The aim of this research is to establish whether a relatively lightweight logical data model (i.e., Mainstems), grounded in a conceptual standard (i.e., HY_Features), can meet these goals. If feasible, concept-based hydrologic feature integration will allow relative hydrologic location to be an empowering technique for structuring shared knowledge within the hydrosciences. In this vein, testing of the Mainstems model focuses on three key benchmarks to measure success:

Concepts: Can the HY_Features conceptual model be used with varied historical datasets?

Functionality: Can a general, lightweight logical data model (Mainstems) provide the required functionality for indexing and network operations over large, complex data?

Integration: Can hydrologic feature integration be used to better solve traditional geographic issues of scale and granularity than spatial integration alone?

To communicate the details of Mainstems, rigorous and precise terminology is required. Appendix A contains terms and definitions specific to HY_Features (*catchment, flowpath, nexus, hydrologic location)* and the Mainstems data model (*headwater, outlet, mainstem, drainage basin)*. These terms are *italicized* throughout this paper.

1.2. Data modeling and scope

The intended scope of the HY_Features and the Mainstems logical data model is broad to enable data integration. They can be applied to any hydroscience data, model, or workflow but do not introduce new techniques or methods. Specific physical data models—such as NHDPlusV2—or hydrology aspects of model interface specifications—such as the Community Surface Dynamics Modeling System (CSDMS) Basic Model Interface and Standard Names (Peckham, 2013, 2014)—should be viewed as implementations that Mainstems is compatible with. Mainstems is not intended to be a new or different hydrologic network coding system—such as Gravelius Order (Gravelius, 1914) or to introduce unique fluvial-geomorphic characteristics or representations; instead, it is intended to allow integration and representation of datasets that encode such systems.

Box 1
WaterML2 suite details

The WaterML2 suite aims to provide a comprehensive set of standards for the hydrology domain. It is designed around the observations and measurements data model (Cox, 2013) that recognizes observations as activities involving sampling a particular feature of interest. HY_Features provides hydrology-specific feature types for identifying features of interest.

Box 2
Logical Data Modeling.

Conceptual models provide unifying terms and associations to describe complex or abstract ideas in a precise and formal way. Logical models add constraints and extensions to conceptual models that meet specific needs and can be implemented in a variety of use-specific data models. This third tier is commonly referred to as a "physical" data model.

This three-layered paradigm of conceptual, logical, and physical data modeling has been used in other parts of WaterML2 (Brodaric, 2017) and allows applications (e.g., hydrologic model software) to implement HY_Features concepts in a way that is interoperable with data models employing the same conceptual model (Brodaric, 2018).

1.3. Case studies

Case studies used in the development and evaluation of Mainstems are presented here to illustrate the utility and functionality supported by the model. The following is a brief summary of the case studies and how they specifically evaluate Mainstems.

1) The first case study concerns an integrated hydrologic network indexing system based on the NHDPlus data model called the Network Linked Data Index. It shows how Mainstems can be used to provide continental-scale data indexing and network navigation functionality.

2) The second case study integrates linear and areal continental-scale datasets, the NHDPlus and Watershed Boundary Dataset (WBD), using a mainstems-based hydrologic network to match watershed outlet locations. This case study shows how Mainstems can be used as a tool in matching network locations by their place in an overall hydrologic network.

3) The third case study integrates a linear representation of a hydrologic network (flowpaths) with an aerial representation of hydrologic units (catchment boundaries). This example shows how Mainstems can be used to unify networks of polygonal units with networks of linear features across scale.

4) The fourth case study integrates a coarse network (the U.S, River Reach File known as RF1) with a more resolved network of flowpaths and catchments (The NHDPlus Version 2.1). This case study explores the functional limitations of Mainstems-based integration in headwater areas.

5) The final case study describes integration of a new graph-theoretic data model (the Common Hydrology Features known as CHyF) for representation of catchments and flowpaths with unique identifiers for mainstems and drainage basins. This case study shows how Mainstems-based identifiers can be embedded in a modern highly resolved network (graph) of hydrologic features.

These case studies are described in detail in section 3 of this paper. Data models of the datasets used for case studies are described in Appendix B.

2. Mainstems logical data model

2.1. Overview

In geography, location describes a relation, and not a property (Kuhn, 2012). All descriptions of location express a spatial relation between features to be located and chosen reference system (a region, a street network, coordinate axes).

Absolute location uses precise (if approximate) coordinates (e.g., latitude 37.5467, longitude −119.5678) to describe the finite location of a feature on a reference coordinate system (Goodchild, 1992; Herring, 2010). Typical spatial integration looks for features that share the same absolute location with some level of error. While absolute location can establish spatial proximity, there is an equal need for a relative referencing system that is meaningful to people and software that use hydroscience data (e.g., a specific gage is upstream of a specific dam). The typical way this is achieved in hydrographic data is with linear referencing such as that in NHDPlusV2. While this approach achieves a degree of relative referencing, it is limited to the scope of a single data model and/or a specific geometry. As a result, the utility of linear referencing for broad, cross dataset integration is limited.

An alternative representation of location describes one feature in relation to another (e.g., the school is right behind the post office). Referencing by relative location requires *contextual* knowledge of at least one of the places (e.g., the location of the post office). Context can also be driven by concepts, such as *homes are buildings in residential zonings and stores are buildings in commercial zonings*. This idea of concept driven relative referencing is how mainstems seek to match features by the hydrologic processes they serve.

Mainstems aspires to provide a nested, multi-resolution system to provide persistent features for relative referencing of hydroscience data to any hydroscience dataset. An application that illustrates Mainstems' intended use is integration of observed and simulated information. For example, consider a set of monitoring stations and a hydrologic model that predicts streamflow at unmonitored locations within a *drainage basin*. If we know the absolute location of the monitoring and prediction locations and what *mainstem* they are "on" (relative location), we know their *hydrologic location* and can draw meaningful comparisons between the two sources. While existing coordinate reference systems and linear referencing systems can be used to define absolute location, a multi-scale relative reference system for hydrologic networks does not exist.

To handle cross-scale issues within the concept of *catchment*, Mainstems emphasizes incremental *catchments*. Incremental *catchment* datasets exist both globally (Lehner, 2013; Yamazaki, 2019) and for specific countries (McKay, 2015; Sondheim, 2019; Bureau of Meteorology, 2012). These datasets discretize the landscape into one "incremental" *catchment* area per confluence-to-confluence *flowpath*. Using well-established elevation-derived hydrography processing methods (Dixon and Uddameri, 2015), a coverage of incremental *catchments* can be delineated from elevation and existing incremental *flowpaths*. An important feature of Mainstems is that incremental *catchments* can be seen as a "flattened" combination of nested *drainage basin* boundaries. A *mainstem flowpath* network and associated total *drainage basins* are embedded in an incremental *catchment* network. By defining *mainstem flowpaths* and identifying them across datasets, the sophistication of incremental catchment data models can be handled simultaneously with a persistent and minimal network of *mainstem flowpaths* connecting *headwater* and *outlet* locations. Fig. 1 illustrates the *mainstem flowpath, drainage basin, headwater,* and relationship with incremental *catchments.*

2.2. Headwaters, flowpath and hydrologic location

A *headwater* is the region where flow coalesces and starts to form a flowing body of water (Montgomery and Dietrich, 1988, 1992; Wohl, 2018). Regardless of physical reality or geomorphic theory, in a given hydroscience dataset, the *headwater* is the area upstream of an observed or predicted flowing river. As the nesting level of incremental *drainage basins* increases, and their sizes decrease, the spatial characteristics of a

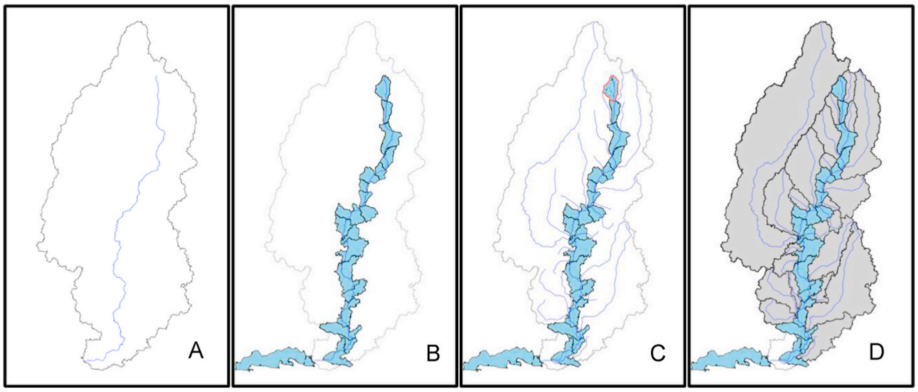

Fig. 1. Illustration of how mainstems and drainage basins integrate with an incremental catchment data model. A) One mainstem flowpath (blue line) and its drainage basin (grey outline). B) Incremental catchments for the mainstem shaded in light blue. C) Headwater catchment (outlined in red) and tributaries contributing to the mainstem (blue lines outside of incremental catchments). D) Drainage basins for each tributary shaded in grey and outlined in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

headwater location must be known with increasing precision. The Mainstems data model builds on the understanding that the spatial representation of a *headwater* can vary across datasets. The characteristics of *headwater* areas upstream of observable *flowpaths* or shorelines are not explored in this research but are required for a complete hydrographic data model and provide avenues for future work. *Headwater* areas often contain isolated *drainage basins* that do not connect into nearby dendritic systems through surface channels. In these instances, it is practical to include them as parts of larger downgradient drainage basins. By focusing on sufficiently coarse spatial resolution—*drainage basins* large enough to have an established flowpath at their outlet, what Schumm (1977) termed the "transfer" and "storage" zones—Mainstems avoids the complexities of local hydrology yet maintains wide applicability and stability.

Mainstems assumes dendritic connections in the downstream direction meaning divergences are treated as new *headwater* locations. If divergences are identified as *hydrologic locations*, capable applications of mainstems could route flow through divergent parts of the network. The dendritic assumption helps reduce complexity, promotes stability, and improves reliability of references and integrations that use the logical data model.

2.3. *Flowpath, catchment, and hydro nexus*

The downstream end of a mainstem flowpath is a hydrologic location of type outlet. Outlets can be referenced to the next downstream flowpath. Each outlet location has one, and only one, associated headwater location. If a given dataset includes incremental catchments, an outlet location could be a realization of a *hydrologic nexus (HY_HydroNexus)* connecting contributing and receiving catchments.

HY_Features describes a *HY_HydroNexus* as the interface between two catchments. If implemented in a hydrologic model, a *nexus* occurs when an upstream model domain is coupled to a downstream domain. In a hydrographic dataset, a *nexus* feature is implemented with attributes describing what catchments are upstream and/or downstream of a given feature. HY_Features allows *nexus locations* to be "realized" as different kinds of hydrologic features and/or more complex classes such as modeling software interfaces. While other *nexus realizations* can be

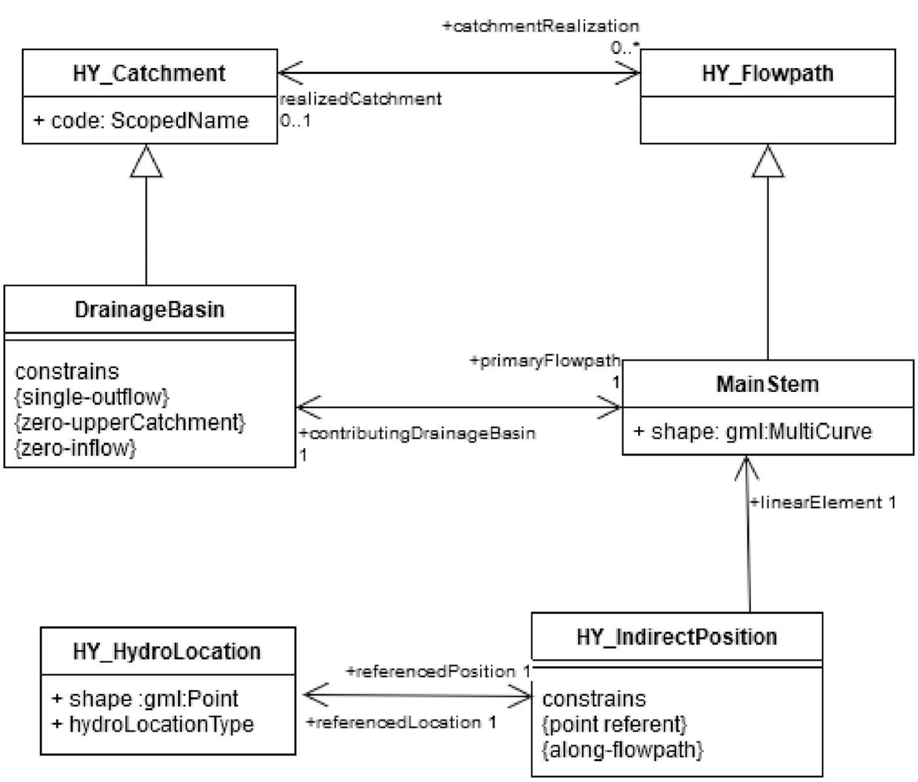

Fig. 2. UML class diagram showing two subclasses (lines with open triangle arrows) and one relation (lines with open arrow heads) introduced in the Mainstems logical data model. Hydrolocation and indirect position from HY_Features are included for context. As subclasses, drainage basin and mainstem inherit realizedCatchment and catchmentRealization associations from catchment and flowpath parents. In addition to the HY_Features class associations, more specific associations, contributingDrainageBasin and primaryFlowpath are introduced.

specified, only *hydrologic locations* with explicit point geometries are specified by HY_Features.

Mainstems does not require that *nexuses* between *catchments* be included but it is compatible with systems that require *nexus* interfaces between *catchments* and their *flowpaths* such as the NHDPlusV2 or hydrologic models that explicitly represent processes at confluences. See Fig. 2 for a formal Unified Markup Language (UML) representation of the Mainstems extension of HY_Features classes and relations.

3. Case studies

The following case studies focus on the relationship between *mainstem flowpaths* and *hydrologic locations* while testing the complete Mainstems data model. Dataset-specific feature types and attributes are denoted with **bold text** and defined in detail in Appendix B.

3.1. Network Linked Data Index

The Network Linked Data Index (NLDI) (https://waterdata.usgs.gov/blog/nldi-intro/; Fig. 3) provides navigation functions over the flowline network, and returns indexed *hydrologic locations* found along the navigation. The NLDI system works with the complex continental-scale NHDPlusV2 dataset and any number of datasets indexed to the NHDPlusV2. The NLDI system is a test and demonstration of Mainstems ability to provide intuitive functions for network navigation and hydrologic data discovery. NLDI indexing and data retrieval resolve hydrologic locations to predefined, spatially indexed NHDPlusV2 catchment IDs (COMID). All navigation requests start at a known COMID and each COMID is associated with a unique **Level Path ID** (*mainstem* ID), and **Hydro Sequence** (upstream-downstream sort order) that can be used for flowline navigation.

The upstream *mainstem* is defined by all features associated with the same **Level Path (*mainstem*) ID** with a **Hydro Sequence** greater than the starting location. Downstream *mainstem* and upstream with tributaries use similar logic but recursively follow relationships between *mainstems*, their tributaries, and the larger mainstems to which they contribute. The NLDI also supports the retrieval of drainage basin boundaries for any NHDPlusV2 catchment outlet. Given this functionality, the NLDI system can resolve the mainstem, hydrographic network, drainage basin, and linked hydrologic locations for any NHDPlusV2 COMID or any hydrologic location that has been indexed to the NHDPlusV2 network.

This case study illustrates the utility of mainstems for indexing and discovering hydrologic locations. When a location is indexed, the association is recorded as a relationship with a COMID. Since the COMID has a unique mainstem (Level Path ID), it defines the mainstem that the hydrologic location is “on”. If the identifiers for flowpaths and catchments of other relevant hydrographic datasets are added to the NLDI as linked hydrologic locations, this approach opens the possibility to interface hydrography and hydrologic location data from any integrated dataset. The challenge is how to identify the mainstem in NHDPlusV2 (therefore NLDI) with mainstem features from other hydrologic representations.

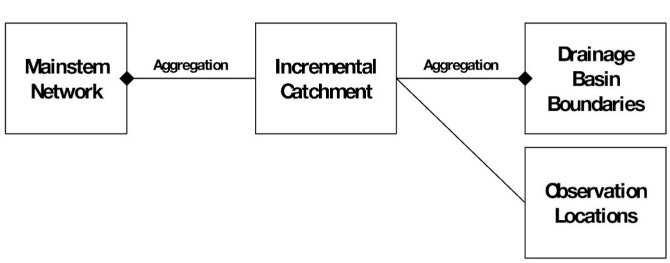

Fig. 3. Simplified UML class diagram of the NLDI as a network of mainstems that aggregate incremental catchments that are linked to observations. Aggregation of incremental catchment flowpaths and boundaries can be mainstem-based or drainage basin -based, respectively.

3.2. Associating the National Hydrologic Model Geospatial Fabric to twelve digit hydrologic units

This case study tests Mainstems ability to integrate network locations (outlet points) referenced to different large-scale and complex hydrographic networks. The case study focuses on integrating the National Hydrologic Model Geospatial Fabric (GF) (Viger, 2014) and outlets of the Watershed Boundary Dataset **Twelve Digit Hydrologic Units (HU12s)** (U.S. Geological Survey, 2019; Price, 2018). The association between these outlets was needed for a daily water balance modeling study.

The GF includes approximately 100,000 *catchments* (called **Hydrologic Response Units (HRUs)**. The *hydrologic nexuses* of GF HRUs (referred to as **Points of Interest (POIs)**) were defined based on the *hydrologic location* of stream gages, water quality sample sites, and modeling criteria like maximum flow distance. **HU12** outlets were established after publication of the GF, so they were not included as GF POIs.

A complication that makes this an especially useful test case for Mainstems, is that the GF is based on NHDPlusV1 while the HU12 outlets are based on NHDPlusV2, meaning the underlying networks are not identical. To reconcile this, both hydrologic and spatial associations were established.

Where substantial upstream drainage area and flow exist, both the HU12 and GF datasets had representations of equivalent mainstems. Fortunately, a near complete mapping from NHDPlusV1 to NHDPlusV2 is available, so creating one here was not required. With a known mapping, HU12 outlets and GF POIs could be placed relative to each other along mainstems and linear distance and drainage area differences could be established. With this information, network matches (hydrologic associations) were established, and drainage-area ratios were created to adjust total flow from the GF POI on the same mainstem and with the nearest drainage area to each HU12 outlet. In headwater areas with insufficient flow or drainage area to use this approach, spatial intersection relations were used instead. The need for spatial intersection for some HU12 outlets illustrates an important limitation of Mainstems. In *headwater* areas, there may be areas that have ambiguous or poorly established *flowpaths*. This is not a limiting factor of the logical model (and rather a practical limit of the data) as a *mainstem flowpath* is defined as a feature where flow can be observed and/or predictions of flow can be made. A *mainstem flowpath* should not be expected to extend further upstream than flowing water is regularly observed.

3.3. NHDPlusV2 mapping to twelve digit hydrologic units

This case study integrates two hydrographic datasets that use different incremental catchments. These include the NHDPlusV2, which contains linear flowpaths and polygon catchment boundaries and the WBD HU12s, which contains only polygon catchment boundaries.

Integrating these datasets requires determining the most representative NHDPlusV2 flowpath for each WBD HU12 catchment. This task is a good test of Mainstems because it deals with mainstem collections of polygonal HU12s that correspond to mainstem collections of linear NHDPlusv2 flowlines. This ability to identify collections of linear *or* polygonal features that make up the same *mainstem* as shown in Fig. 4 is a core function provided by Mainstems.

The boundaries of WBD **HU12s** were hand-drawn using a variety of base maps resulting in potential disagreement with different scale versions of the digitally derived NHD. To associate NHDPlusV2 flowlines and HU12 units, two sources of evidence were used: (1) the collection of HU12s found via spatial intersection with a mainstem of NHDPlusV2 flowlines, and (2) the headwater to outlet connectivity of HU12s.

The processing required a dendritic tree of HU12s and removal of

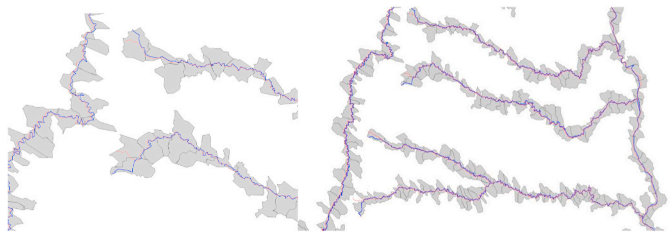

Fig. 4. Mainstems from NHDPlus V2 NHD flowlines (blue), WBD HU12s (grey), and RF1 segments (red). Left panel shows close-up highlighting how headwater representation may be different. The right panel shows that, as mainstem collections, the three datasets generally agree.

non-dendritic and coastal NHD **flowlines**. Three phases of matching and two cleanup steps were performed.

1) First, NHDPlusV2 flowlines were intersected with the HU12 units and the most downstream (largest) intersecting *mainstem* for each HU12 was found.
2) *Mainstem* collections of **HU12s** were identified by tracing the hydrologic connectivity attributes from *headwater* to *outlet* **HU12s.**
3) Results of part one and two were compared, noting where the spatial and hydrologic matches are different.
4) **HU12s** present in set 1 but not set 2 were assigned to the next smallest mainstem contained in set 1.
5) **HU12s** that were in set 2 but not set 1 were flagged as potential errors and assigned the largest non-intersecting mainstem in set 2.

This case study illustrates how Mainstems can help identify hydrologically similar collections of *flowpaths* and *catchment* boundaries from disparate datasets. This process is particularly useful near confluences in datasets that do not have perfectly aligned geometries (Fig. 5). In these cases, it was helpful to identify a single set of **HU12s** for each *mainstem* through a process of elimination starting with the most dominant (longest) *mainstem* in the network. This ensures that large *mainstems* have a complete set of **HU12s;** smaller *mainstems* do not get matched to river bottom **HU12s** they cross; and that every **HU12** that intersects multiple **flowlines** gets attributed to the largest river it intersects.

A challenge when matching mainstems across datasets is the identification of headwater locations. In the case of HU12s and NHDPlusV2 flowlines, the top of a flowline sometimes crosses over a drainage basin boundary such that – if the upper extent of the flowline were used for matching – the wrong headwater HU12 would be identified. In areas of low relief, especially where ditches have been created, it is not always clear where the true boundary begins. To manage these situations, the outlet of a first order NHDPlusV2 catchment (typically much smaller than a HU12) was used for headwater matching. This issue is included here to further illustrate that *headwater* areas must be treated with care when implementing the Mainstems model.

3.4. NHDPlusV2 mapping to River Reach File 1

To further exercise the Mainstems data model, our fourth case study integrates an early digital hydrographic data model, the River Reach File 1 (RF1) (Nolan, 2003; Brakebill, 2011) with the more modern NHDPlusV2. In contrast to the NHDPlusV2 with over 2.7 million catchments for the conterminous United States, RF1 has less than 70, 000. The problem addressed here is to identify the collection of RF1 segments (the RF1 name for flowpaths) that correspond to the mainstems of the NHDPlusV2. This is a good test to explore limitations of Mainstems where datasets have different headwater flowpath density and geometry precision.

The process of integrating RF1 segments and NHDPlusV2 flowlines started by matching the mid-point along a headwater RF1 segment to a NHDPlusV2 catchment. Given that the NHDPlusV2 network contains much smaller headwater catchments than the headwater RF1 segments, the top of RF1 segments can fall in a headwater catchment that are not the headwater catchment of the NHDPlusV2-indicated mainstem. To work around this, an algorithm using a largest-river-first process of elimination, similar to that implemented in previous case studies, was used to find the representative mainstem for each NHDPlusV2 catchment matched to a headwater RF1 segment.

Once the representative NHDPlusV2 mainstem was established for each RF1 headwater segment, the same largest-river-first process of elimination was used to match collections of RF1 segments to NHDPlusV2 mainstems. This process worked largely as expected (Fig. 4). An expected issue regarding the upstream limit of Mainstems is shown in Fig. 6. For a dataset, such as RF1, that resolves *headwater catchments at coarser resolutions than* the dataset being integrated, *headwater flowpaths* may diverge from the finer-scale *mainstems*. This should be viewed as a limitation of the resolution of hydrographic datasets used.

3.5. Defining mainstems and drainage basins with CHyF

This case study questions whether *mainstems* and associated *drainage basins* can be determined programmatically using a test area in the Richelieu River Valley in southern Quebec, Canada. The challenge here is that the CHyF data model and service implementation uses computation (rather than pre-calculated attributes) as much as possible, while maintaining hydrologically consistent and correct relationships between *flowpaths* and *catchments*. This case study tested whether Mainstems is compatible with basic network operations using a graph theoretical technical baseline. That is, if the geometries associated with existing and calculated attributes can be used to determine *mainstems* and associated *drainage basins*.

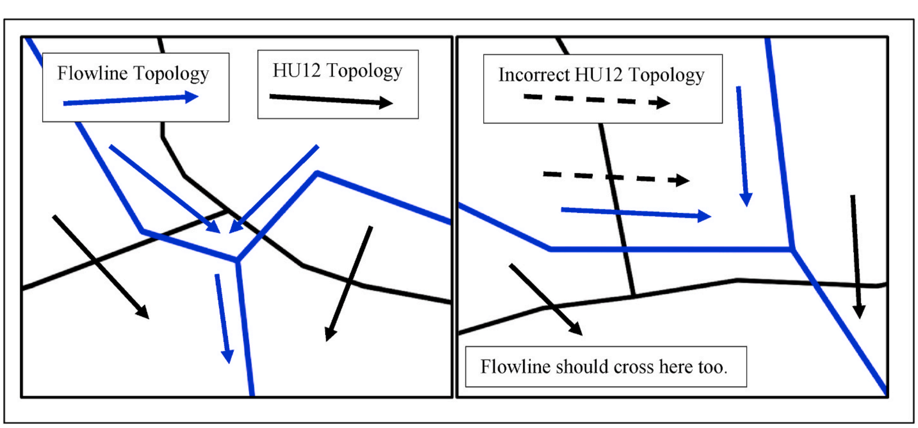

Fig. 5. Flowline geometry and topology in blue and HU12 geometry and topology in solid black. The ideal case is shown on the left where flowline and HU12 geometry agree. As shown on the right, when the geometry does not agree, topological attributes are sometimes "hydrologically incorrect" but in agreement with the geometry and other times correct but in disagreement with the geometry. This ambiguity between geometric topology and attribute topology can be identified and solved using mainstem logic. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

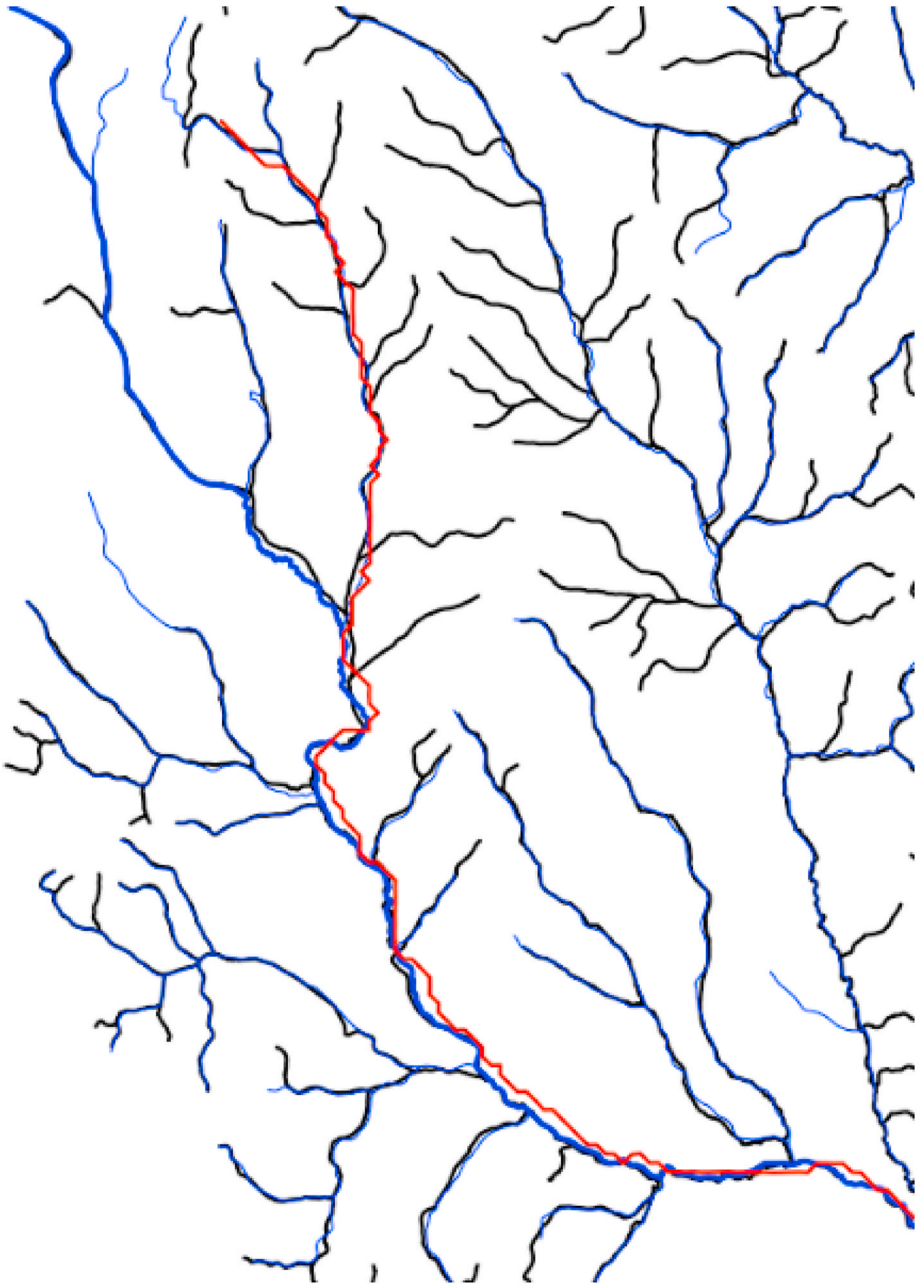

Fig. 6. NHDPlusV2 (in blue) and NHDPlus High Resolution (NHDPlusHR) (in black) compared to one RF1 headwater segment (in red). NHDPlusV2 levelpath matched to RF1 headwater shown as thick blue line. Notice that the red RF1 segment is a single headwater (has no tributaries). Any integration with RF1 upstream of the outlet of this headwater segment may not be represented well with Mainstems. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Input data consisted of **elementary** (incremental) *flowpaths* and *catchments*, as well as the boundary for the area of interest. The *flowpaths* have various attributes, including a primary name and rank, designating whether the flowpath represents a primary or secondary flow. When data are first read into a CHyF repository, a graph is created and attributes including length, Strahler Order (stream order), Horton Order, and Hack/Gravelius Order (Gravelious, 1914; Hack, 1957) (stream level) are populated.

In CHyF, a collection of *flowpaths* designated as a *mainstem* has several key characteristics: all **elementary** *flowpaths* represent primary flows, have the same name, and are assigned the same Gravelius Order value. Additionally, the *mainstem* of a *drainage basin* can be specified as meeting a minimum length or drainage area criterion, both of which can readily be determined. Length is calculated by summing the **elementary** *flowpath* lengths along the *mainstem*, and area drained by summing of the areas of the upstream **elementary** *catchments* as assessed by graph navigation. A geometric union of these **elementary** *catchments* gives a polygon representing the associated *drainage basin* in the same way as the NLDI.

The process described above works, but it has two caveats that illustrate nuances related to Mainstems. First, it depends on Gravelius Order, which has been calculated using a longest path with geographic name approach to reach the *headwater*; this assumes that Gravelius Order, as defined, corresponds to the *mainstem*. Second, the results are not necessarily as expected due to divergent flows. CHyF allows for secondary diverted flows, with the result that one large drainage can leak into another if one or more secondary flows connect them. To circumvent this, as is implied by Mainstems, the navigation algorithms could be adjusted to ignore secondary flows or to treat them as primary flows, as described below.

For Canada, with no equivalent of the level of detail provided by **HU12** in the United States, the automated technique described above is worth investigating further. Provided the input data are of sufficient quality, it may be practical for major *mainstems* and *drainage basins* to be determined relatively quickly across much of the country.

3.6. Case study summary

The collection of case studies presented show the utility of Mainstems to integrate point-representations of hydrologic locations, linear representations of flowpaths, and polygon representations of catchment boundaries. The first, uses Mainstems as the primary index in a network navigation and linked-data discovery utility. The second demonstrates how mainstems can be used to associate *catchment* outlets using *mainstem* matching. The third shows how Mainstems can integrate *mainstem* collections of linear *flowpaths* and polygon *catchment* boundaries. The fourth provides an important test and demonstration of how the Mainstems data model behaves in *headwater* areas. The final case study illustrates how the Mainstems model can extend to datasets from other countries, and those that are implemented using graph concepts. While generally positive, the results if these case studies illustrate some important limitations of Mainstems.

The first is that *headwater* is a scale-dependent location and each hydrographic dataset considered has one upstream-most *catchment* along a given path. The entire upstream-most feature could be considered a representation of the *headwater* of the *mainstem* in question. As a result, anything upstream of the *outlet* of a *headwater* catchment area (itself a small *drainage basin*) should not be expected to match from one dataset to another. This is illustrated in Fig. 6 where the red line is one RF1 **segment** and the blue line is the NHDPlusV2 Level Path with the same outlet. The RF1 segment is a headwater path that diverges from the NHDPlusV2. The data model has value in that it provides a stable reference for a *mainstem*, but for the upper extents of a *mainstem*, the precision is limited by the resolution of the datasets in question. Practically, this means the Mainstems data model should be used with caution in *headwaters*.

Divergences present some interesting issues for Mainstems. Practically, there were three types of divergences encountered in the case studies: 1) anthropogenic divergences, 2) hydrologic divergences, and 3) complex *flowpaths*. Anthropogenic divergences are not directly considered by Mainstems, other than treating them as *hydrologic locations*. Hydrologic divergences form new *headwaters* and a specific data model for locating headwaters of divergences along their source *flowpath* could be the subject of future work. Complex *flowpaths*, such as a river encircling an island, may be represented by a single line in one dataset and several segments in another. In this case, all *flowpaths* could be considered part of the *mainstem* as there is no independent drainage basin resolved in Mainstems. The distinction between hydrologic divergences and complex flowpaths is dependent on the resolution of the smallest drainage basins in a given implementation of Mainstems. As a result, we suggest the Mainstems model be used on basins of sufficient size to have an established *flowpath*.

In general, the case studies demonstrate that using a single *mainstem* that flows from a *headwater* source area to the outlet of a drainage basin has great utility. For persistent identification of rivers in hydrographic addressing, Mainstems can provide a minimum yet sufficient set of easily identified linear paths through a hydrologic network. For integration of *hydrologic locations* and linear and aerial representations of *catchment* networks, Mainstems provides a data model and set of assumptions that aid processing.

4. Considerations and future work

The Australian Hydrologic Geospatial Fabric (Commonwealth of Australia, Bureau of Meteorology, 2012) introduces the concept of “contracted nodes” as permanent *hydrologic locations* that are intended to persist through space and time. Mainstems does not include this concept explicitly; however, the *nexuses* of persistent *mainstem flowpaths* (confluences) are implicit “contracted nodes”. A given implementation of Mainstems could create identifiers for these mainstem confluence locations and other important *hydrologic locations* in the spirit of “contracted nodes”. As an implementation issue, CHyF allows for these major *nexuses* to be identified as “contracted nodes”.

Mainstems assumes that all *drainage basins* have one *headwater* and one *outlet*. This assumption is only valid for *drainage basins* that have a *mainstem* that terminates at its *outlet*. Areas with no *flowpaths*, either upstream of a first order *flowpath* or along the shore of a waterbody, so-called zero-order catchments (Dietrich, 1987), are not accounted for in Mainstems. Such areas must be encompassed in a *drainage basin* that includes an identifiable *mainstem* in order to be included in the hydrologic landscape. This is not an issue for *flowpath*-oriented hydrologic integration and addressing but will need to be accounted for in future work aimed at integration of surface water, groundwater, and especially waterbodies. The upstream-most location of a *mainstem* is not a single point; in reality, there is some *drainage basin* without *flowpaths* upstream of the top of the initiation of flowing water. These locations might be a wetland, a spring, a glacier, or a field. In each case, there are unique nuances of hydrogeomorphology that dictate how and when a *mainstem* can be said to exist. Exploration of data models to support these kinds of headwater areas to more concretely establish the characteristics and geometry of headwaters is left for future work.

If implemented for all rivers from large to small, the mainstem - drainage basin paradigm implies a collection of incremental catchments that cover the landscape. It does not address the specific logical data model for implementing them. When disaggregating the landscape into a collection of incremental catchments, these catchme*nts* correspond to a directed, acyclic graph, with each *catchment* draining into another. The associated *flowpaths* form a separate directed, acyclic graph that can be combined with the first such that *catchments* and *flowpaths* are represented in the same graph. Each *drainage basin* and *mainstem* can be defined as the union of elements represented in the graph. A wide array of *hydrologic locations* such as tributary confluences too small to be part of the *mainstems* network, stream gages, dams, and outlets of major waterbodies can all be related to locations on the graph. Thus, a specific data model for incremental *catchments* and associated *flowpaths* can be built on graph theoretical constructs.

CHyF does exactly this by providing a logical data model defined as a profile of HY_Features in the context of graph theory. This approach has two strong benefits, provided that the data are topologically clean, with correct *flowpath-catchment* relationships at *nexus*es. The first benefit is that assigning identifiers to graph elements and building the directed acyclic graph using them is a fast process. This limits the importance of maintaining identifiers, especially for smaller streams and *catchments*, since point and linear referencing along the *flowpath* network can be based on (XY) coordinates. The second benefit pertains to the fast navigation through the graph that can be implemented, which can be combined with a new coverage union operation (Davis, 2019) to quickly create *drainage basins* or other large *catchments* from incremental *catchments*. CHyF recognizes the importance of *mainstems* and *drainage basi*ns, supports them in its data model, and will be adding services to support their use.

4.1. Conclusion

This research documents the design and testing of the Mainstems data model. The case studies presented in this paper demonstrate that:

- HY_Features concepts, which are the basis for Mainstems, are compatible with varied historical datasets,
- that the Mainstems data model supports both hydrologic indexing (hydrographic addressing) and network navigation functions, and
- that Mainstems-based logic supports integration of hydrographic data in ways that spatial-integration alone cannot.

The case studies illustrate important limitations of Mainstems in *headwater* regions and how the dendritic assumption of the model can accommodate divergences as *hydrologic locations*. Our research has shown that Mainstems provides a useful approach to support persistent identification through a minimum, yet sufficient, set of networked hydrologic features.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Acknowledgements

All U.S. Geological Survey work documented in this paper was funded by the U.S. Geological Survey Water Mission Area Water Availability and Use Science Program. The ongoing development of the Common Hydrology Features model is being funded by the Centre for Mapping and Earth Observation (CCMEO), Natural Resources Canada and supported by both that agency and the National Hydrological Services, Environment and Climate Change Canada. USGS personnel have also provided substantial support and encouragement to CHyF developments through ongoing interaction with CCMEO.

Appendix A. : Terminology and Concepts

Note: the source of the term is denoted as (HY_Features) or (Mainstems).

7.1.1 Areal Features

Catchment (HY_Features): HY_Features defines the *catchment* concept as a holistic geomorphic feature type that represents all the aspects of hydrology draining to an identified outlet. *Catchments* are typically organized in a hydrologic cascade that fully partitions the landscape, only breaking at major waterbodies. A *catchment* receives flow from zero or one inflows and contributes flow to zero or one *outlets*.

Drainage Basin (Mainstems): Like the *catchment* feature type, a *drainage basin* is also a holistic feature; it is defined as the total upstream area draining to an *outlet*. It is comparable to a *catchment* with no inflows and a single *outlet*. *Drainage basins* can be thought of as a total accumulated or total upstream *catchment*. While having no formal description in HY_Features, *drainage basin* is a commonly used term (Petts, 1996; Schumm, 1977). A *drainage basin* can be described with a pair of locations: 1) the *headwtater* area with no discernible *flowpaths* where flow initiates and 2) the *outlet* where flow enters a larger river or waterbody. A single *mainstem flowpath* connects a *drainage basin's headwater* to its *outlet*.

7.1.2 Location Features

Hydrologic Nexus (HY_Features): A *hydrologic nexus*, shortened to *nexus*, represents the interface along a *flowpath* between two or more *catchments*. One or more *catchments* contribute flow to a *nexus,* which contributes flow to one or more other *catchments*. Conceptually, a *nexus* can be defined anywhere on the landscape, but datasets typically establish them at *confluences* along a *flowpath* or other important network locations such as stream gaging stations.

Hydrologic Location (HY_Features): Any location that can be thought to be “on” a hydrologic network. A hydrologic location may or may not be coincident with a hydrologic nexus. Any geospatial representation of a hydrologic nexus is inherently a hydrologic location. Hydrologic locations are commonly associated with multiple hydrographic datasets and/or hydrologic models providing points of integration.

Outlet (HY_Features): An *outlet* defines the most downstream location in a *catchment* where water exits into the ocean or the next downstream waterbody or *catchment*. HY_Features formalizes the type of *hydrologic location* as a *catchmentOutlet*, which is functionally where flow exits a *catchment* and a *nexus* is where multiple catchments flow, mix and/or split to contribute to one or more downstream *catchments*. Practically, every *catchment* drains to a single outlet *hydrologic nexus,* which could be represented by a *catchmentOutlet hydrologic location*.

Headwater (Mainstems): A *headwater* is scale-dependent and represents the most upstream location where water can theoretically exist in a *drainage basin*, typically on a *drainage basin* boundary. Given that the definition of a *flowpath* does not necessitate the existence of water, *headwater* can be imagined as a point where an extended *flowpath* touches a *drainage basin* boundary. Much in the same way location theory explains how the location of economic activities can be determined on a broad level such as a region or metropolitan area, or on a narrow one such as a zone, neighborhood, city block, or an individual site, the idea of *headwaters* can be defined by a *drainage basin* boundary, hillslope or a specific point in space.

7.1.3 Linear Features

Flowpath (HY_Features): A *flowpath* represents a one-dimensional idealized path that water follows through a *catchment*. In the case of a *catchment* with no inlet, a *flowpath* might extend along a main path to the *catchment* boundary.

Mainstem (Mainstems): The *mainstem* concept extends and constraints the concept of a *flowpath* by designating a single path from a *headwater* source to an *outlet* as the primary water feature used to traverse the network through a *drainage basin*. In other words, a *mainstem* is a linear realization or backbone of a *drainage basin*.

Appendix B. Hydrographic Data Sources

9.1.1 River Reach File 1

The RF1 was first introduced by the U.S. Environmental Protection Agency in 1985 (Horn, 1994). It is a linear network of major rivers in the continental United States (Dewald, 2015). The version of the RF1 used here is the most recent version, which has been quality controlled for network connectivity and value-added attributes such as drainage area and time of travel (Nolan, 2003). The RF1 network includes only linear representations of rivers, which are referred to as segments. Some non-dendritic connections are present, but all connections have an identified primary path in the network.

9.1.2 Watershed Boundary Dataset

The WBD (U.S. Geological Survey, 2013; U.S. Geological Survey, 2019) is a collection of nested basin boundaries based on a hierarchical drainage system of hydrologic units for the United States. Major river basins are given two-digit codes and progressively smaller hydrologic units are identified by 4-, 6-, 8-, 10-, and 12-digit hydrologic unit codes (shortened to HU2 through HU12). The WBD has developed over time with improvements in elevation data used for delineation and knowledge of hydrologic characteristics. Two versions of the WBD are used: a static snapshot of the HU12s that was used in the production of the NHDPlusV2 (Moore and Dewald, 2016) and the current release, which includes the best available geometry and attributes at the time of access (USGS, 2018). An attribute, “toHU”, provides routing capabilities and dendritic hydrologic unit connectivity; it is used extensively to identify the network of WBD HU12s in this paper.

9.1.3 National Hydrography Dataset Plus

The NHDPlus has two static versions (V1 and V2) that share the same data model (Moore and Dewald, 2016; McKay, 2015; Bondelid 2010). NHDPlusV2 is used almost exclusively here. The NHDPlusV2 provides an integration of linear river geometry, polygonal catchment boundary geometry, gridded elevation data, and "value-added" attributes such as landscape characteristics and flow estimates. NHDPlusV2 is created with a catchment delineation algorithm where pre-existing river flowlines and hydrologic units are used to modify elevation data in preparation for flow direction/accumulation processing. NHDPlusV2 catchment boundary geometries are created for each incremental (typically confluence to confluence) NHDPlusV2 flowline segment. Numerous value-added attributes have been created for each set of catchments. The value-added attributes associated with network connectivity and navigation are of particular importance here.

9.1.4 National Hydro Network

The NHN model was first published as a Canadian standard in 2004. NHN pertains specifically to features associated with the hydrographic network and does not include descriptions of catchments or their boundaries. Although linear referencing is described in the model, it has not been implemented. As well, hierarchical and containment relationships are not an explicit part of the model. One of the main drivers behind the NHN has been to produce a national coverage suitable for network analysis. The NHN data are grouped into 1382 large drainage areas that cover Canada. These correspond to the Water Survey of Canada Sub-Sub-Drainage Areas. The NHN has been created from existing topographic data ranging in scale from 1:50,000 to 1:20,000. The data are categorized into four completeness levels, CL1 through CL4; these levels pertain respectively to network topology, waterbody differentiation, data continuity, and toponymy (Belzile, 2008). United States-Canada cross-border hydro harmonization efforts under the International Joint Commission have been based on the NHDPlus and the NHN.

9.1.5 Common Hydrology Features

CHyF is a recent data model developed by Natural Resources Canada that implements catchment boundaries for incremental linear river geometries. CHyF incorporates the concepts and capabilities of graph theory applied to hydrologic elements in an analogous fashion to routing and navigation through a road network and uses the concepts of HY_Features. Catchments and flowpaths can be defined at different levels of granularity. CHyF supports the functional scope of the NHDPlusV2 while allowing for regular updates to underlying data and maintaining stability for integrated applications. CHyF specifies a HY_Features profile and includes some extensions required to implement graph-theoretic functionality. The latter involves the definition of what is called a hygraph, a graph structure made up by features referred to as elementary catchments, elementary flowpaths, and hydronodes. Version 0.9 of CHyF is being finalized at the time of this writing and supports the mainstem and drainage basin concepts described here (Sondheim and Hodgson, 2019).

References

Abbott, M.B., 1993. The electronic encapsulation of knowledge in hydraulics, hydrology and water resources. Adv. Water Resour. 16, 21–39. https://doi.org/10.1016/0309-1708(93)90027-D.

Archfield, S.A., Clark, M., Arheimer, B., Hay, L.E., McMillan, H., Kiang, J.E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W.H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., Over, T., 2015. Accelerating advances in continental domain hydrologic modeling. Water Resour. Res. 51, 10078–10091. https://doi.org/10.1002/2015WR017498.

Atkinson, R., Dornblut, I., Smith, D., 2012. An international standard conceptual model for sharing references to hydrologic features. J. Hydrol. 424–425, 24–36. https://doi.org/10.1016/j.jhydrol.2011.12.002.

Belzile, Y., 2008. National hydro network (NHN). Proc. Can. Hydrogr. Conf. Natl. Surv. Conf.

Biswas, A., 1970. History of Hydrology. North-Holland Publishing Company.

Blodgett, D., Dornblut, I., 2018. OGC® WaterML 2: Part 3 - Surface Hydrology Features (HY_Features) - Conceptual Model. http://www.opengis.net/doc/IS/hy-features/1.0.

Bondelid, T., Johnston, J., McKay, L., Moore, R., Rea, A., 2010. NHDPlus Version 1 (NHDPlusV1) User Guide 126.

Brakebill, J., Wolock, D., Terziotti, s., 2011. Digital hydrologic networks supporting applications related to spatially referenced regression modeling. J. Am. Water Resour. Assoc. 47, 916–932. https://doi.org/10.1111/j.1752-1688.2011.00578.x.

Brodaric, B., Boisvert, E., Letourneau, F., Lucido, J., Simsons, B., Dahlaus, P., Grellet, S., Chery, L., Kmoc, A., 2017. Open Geospatial Consortium OGC WaterML 2 : Part 4 – GroundWaterML 2. GWML2, p. 160.

Brodaric, B., 2018. Interoperability of representations. Int. Encycl. Geogr. 1–18 https://doi.org/10.1002/9781118786352.wbieg0894.pub2.

Brodaric, B., Boisvert, E., Dahlhaus, P., Grellet, S., Kmoch, A., Létourneau, F., Lucido, J., Simons, B., Wagner, B., 2018. The conceptual schema in geospatial data standard design with application to GroundWaterML2. Open Geospatial Data, Softw. Stand. 3 https://doi.org/10.1186/s40965-018-0058-3.

Commonwealth of Australia, Bureau of Meteorology, 2012. Australian Hydrological Geospatial Fabric (Geofabric) Australian Hydrological Geospatial Fabric (Geofabric) 1–4.

Cox, S., 2013. OGC Abstract Specification: Geographic Information — Observations and Measurements.

Davis, M., 2019. JTS Topology Suite, v1.17.0 (forthcoming) [WWW Document]. LocationTech. URL. https://github.com/locationtech/jts/.

Dewald, T.G., 2017. Making the Digital Water Flow the Evolution of Geospatial Surfacewater Frameworks.

Dietrich, W.E., Reneau, S., Wilson, C., 1987. Overview: "Zero order basins" and problems of drainage density, sediment transport and hillslope morphology, pp. 27–38. English.

Dixon, B., Uddameri, V., 2015. GIS and Geocomputation for Water Resource Science and Engineering, GIS and Geocomputation for Water Resource Science and Engineering. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9781118826171.

Dooge, J.C.L., 1959. Quantitative hydrology in the 17th century. La Houle Blanche 6, 799–807.

Goodchild, M.F., 1992. Geographical information science. Int. J. Geogr. Inf. Syst. 6, 31–45. https://doi.org/10.1080/02693799208901893.

Gravelius, H., 1914. Grundriβ der gesamten Gewässerkunde, Band 1: fluβkunde. Compend. Hydrol. 1, 265–278.

Hack, J.T., Seaton, F.A., Nolan, T.B., 1957. Studies of Longitudinal Stream Profiles in Virginia and Maryland. UNITED STATES DEPARTMENT OF THE INTERIOR.

Herring, J., 2010. OpenGIS® implementation standard for geographic information - simple feature access - Part 1: common architecture, 93. Open Geospatial Consortium, Inc.

Horn, C.R., 1994. U . S . Geological Survey Appendix A to Metadata for RF1 USEPA Reach File 1 Converted to Arc/INFO.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., Arheimer, B., 2016. Most computational hydrology is not reproducible, so is it really science? Water Resour. Res. 52, 7548–7555. https://doi.org/10.1002/2016WR019285.

Jiang, J., Zhu, A.-X., Qin, C.-Z., Liu, J., 2019. A knowledge-based method for the automatic determination of hydrological model structures. J. Hydroinf. 21, 1163–1178. https://doi.org/10.2166/hydro.2019.029.

Kuhn, W., 2012. Core concepts of spatial information for transdisciplinary research. Int. J. Geogr. Inf. Sci. 26 (12), 2267–2276.

Lehner, B., Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrol. Process. 27, 2171–2186. https://doi.org/10.1002/hyp.9740.

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., Rea, A., 2015. NHD Plus Version 2 : User Guide.

Montgomery, D.R., Dietrich, W.E., 1988. Where do channels begin? Repr. from Nat. 336, 232–234.

Montgomery, D.R., Dietrich, W.E., 1992. Channel initiation and the problem of landscape scale. Science 84 255, 826–830. https://doi.org/10.1126/science.255.5046.826.

Moore, R.B., Dewald, T.G., 2016. The road to NHDPlus — advancements in digital stream networks and associated catchments. J. Am. Water Resour. Assoc. 52, 890–900. https://doi.org/10.1111/1752-1688.12389.

Nolan, J.V., Brakebill, J.W., Alexander, R.B., Schwarz, G.E., 2002. Enhanced River Reach File 2.

Peckham, S.D., 2014. The CSDMS standard names: cross-domain naming conventions for describing process models, datasets and their associated variables. Proc. - 7th Int. Congr. Environ. Model. Softw. Bold Visions Environ. Model. iEMSs 1, 67–74.

Peckham, S.D., Hutton, E.W.H., Norris, B., 2013. A component-based approach to integrated modeling in the geosciences: the design of CSDMS. Comput. Geosci. 53, 3–12. https://doi.org/10.1016/j.cageo.2012.04.002.

Petts, G.E., 1996. The Fluvial Hydrosystems. Springer Netherlands, Dordrecht.

Price, C., Buto, S., 2018. 12-digit Hydrologic Unit Outlet (Pour) Points for the NHDPlus V2.1 WBD Snapshot. https://doi.org/10.5066/P9SOT2VG.

Schumm, S., 2003. The Fluvial System. Blackburn Press, Caldwell, N.J.

Sondheim, M., Hodgson, C., 2019. Common Hydrology Features (CHyF) v0.9 [WWW Document]. URL. https://github.com/NRCan/chyf/.

U.S. Geological Survey, 2013. U.S. Department of agriculture natural resources conservation service federal standards and procedures for the national watershed boundary dataset (WBD). U.S. Geol. Surv. Tech. Methods 11. A3 63.

Viger, R.J., Bock, A., 2014. GIS Features of the Geospatial Fabric for National Hydrologic Modeling. https://doi.org/10.5066/F7542KMD.

Wohl, E., 2018. The challenges of channel heads. Earth Sci. Rev. 185, 649–664. https://doi.org/10.1016/j.earscirev.2018.07.008.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P.D., Allen, G., Pavelsky, T., 2019. MERIT Hydro: a high-resolution global hydrography map based on latest topography datasets. Water Resour. Res. 2019WR024873. https://doi.org/10.1029/2019WR024873.

U.S. Geological Survey, 2019. (USGS) U.S. Department of Agriculture - Natural Resource Conservation Service (NRCS). U.S. Environmental Protection Agency (EPA). USGS National Watershed Boundary Dataset in FileGDB 10.1 format (published 20190628).