HYDROLOGIC ASSESSMENT OF A SMALL MARITIME HYDROMETRIC MONITORING NETWORK USING A PROCESS-GUIDED CATCHMENT CLASSIFICATION SYSTEM

by

Lindsay Johnston

Submitted in partial fulfilment of the requirements
for the degree of Master of Applied Science

at

Dalhousie University
Halifax, Nova Scotia
August 2020

Table of Contents

List of Tables

List of Figures

Abstract

The National Hydrometric Program, operated by the Water Survey of Canada, is the primary source of surface water quantity data in Canada. The network is cost-shared between the federal and provincial governments, and decisions relating to station placement are made according to both federal and provincial interests. Nova Scotia is a small maritime province in Atlantic Canada with a relatively diverse climate and geology for its small size. The Nova Scotia hydrometric monitoring network currently consists of 31 stations. The overall objective of this study was to determine how well the current network captures the level of hydrologic variability expected in the province using a regional catchment classification scheme. To accomplish this, we developed a combined inductive-deductive catchment classification system and applied it to the province's active monitoring network and 246 ungauged major watersheds. Initially, hydrologic signatures were used to quantify the catchment function of 47 long-term gauged catchments and to cluster similarly behaving catchments. We identified five generalized flow classes and then attempted to replicate this classification using a deductive-based decision tree framework with physiographic and meteorological explanatory variables. The validated decision tree was used to classify the active hydrometric network and 246 major watersheds in the province. The network was assessed to determine how well it covered the expected hydrologic variability in the major watersheds across the province. The results indicated that current active hydrometric monitoring network does not adequately capture the range of hydrologic variability within this region; three of the five primary flow regimes observed in the province are not well represented by the active network. The analysis also illustrated that the stations in the active monitoring network are not evenly distributed among the five generalized flow classes seen in Nova Scotia. The decision tree proved to be a useful tool for understanding the current network's coverage and could also be easily applied by practitioners to identify appropriate donor catchments for ungauged watersheds.

List of Abbreviations

CNHN – Canadian National Hydrometric Network

DEM – Digital Elevation Model

ECCC – Environment and Climate Change Canada

FDC – Flow duration curve

GEV – Generalized extreme value

HBN – Hydrologic benchmark network

HL – Hydrologic landscape

HLR – Hydrologic landscape region

IAHS – International Association of Hydrological Sciences

NHP – National Hydrometric Program

NSE – Nova Scotia Environment

PCA – Principal component analysis

PUB – Predictions in Ungauged Basins

RHBN – Reference Hydrometric Basin Network

UKBN – United Kingdom Benchmark Network

USGS – United States Geological Survey

WMO – World Meteorological Organization

WSC – Water Survey of Canada

Acknowledgements

I want to thank my supervisory committee, who remained calm and supportive as I explored the hydrology of my favourite place (at my own pace). I am seriously indebted to Rob Jamieson who provided ample leeway for me to investigate this topic and was always there to listen to and deconstruct or encourage my “revelations”. Rob has unparalleled patience and seems unaffected when I spend half a meeting talking about sandwiches. Mark Greenwood, external scholar extraordinaire, thanks for being such an enthusiastic cheerleader in both research and life. Barret Kurylyk, your judgement free insight on water below the ground was necessary and appreciated, as I am usually content to ignore it.

Rick, thank you for begrudgingly letting me into your boat to disturb your peace and trying, but largely failing, to impart your wisdom upon me. You are my favourite senior to disagree with! Dewey, you are an excellent teacher, I sometimes like to think about what this thesis would look like if you did not exist. Aud, thanks for dealing with my chaos and being my forever office mate, I can only hope to bug you for another 22 years.

I am very grateful for a research group that has simultaneously acted as a sounding board and enabled me when finding new and interesting ways to procrastinate: Audrey Hiscock, Amy Jackson, Tessa Bermarija, Jenny Hayward, Kiley Daley, Meggie Letman, Baillie Holmes, Yannan Huang, Aidan Van Heyst, and Nicole Bell, without you to enjoy life with, I may have graduated on schedule.

I give my parents enormous credit for supporting me in a pursuit that probably felt largely abstract to them. Many thanks to my sister, Mary-Dan, for being fresh eyes when I forget how to write and to Effie for just being there.

Thank you to the Natural Sciences and Engineering Research Council of Canada and Nova Scotia Environment for providing funding for this research. Tarra Chartrand and June Ferguson both provided much needed administrative support, I have felt very lost these last six months without you.

Chapter 1 Introduction

1.1 PROJECT CONTEXT

The National Hydrometric Network (NHN) serves as an invaluable resource for a wide audience of users across Canada (e.g. for hydro-electric generation, agricultural use, and infrastructure planning). Hydrometric networks allow us to track the spatial and temporal availability of water resources (World Meteorological Organization, 2008). In less populated regions like the Canadian province of Nova Scotia, a trade-off must be made between hydrometric station density and available resources. Nova Scotia has abundant surface water resources but limited economic resources. Use of these economic resources may be especially difficult to justify for monitoring purposes, which can be viewed as nonessential by the general population (see Walker, 2000). As a resource-based economy, Nova Scotia is reliant on its freshwater resources, with many of its sectors requiring some level of hydrometric data. Reliable streamflow data is required for agricultural and industrial use, flood mitigation, forestry management, hydroelectric projects, fisheries management, wastewater treatment, oil and gas operations, and transportation planning. There are economic costs and social implications associated with both over- and under-design of infrastructure resulting from inaccurate water level and discharge predictions. Additionally, stations provide data which contributes to the scientific body of knowledge for hydrology and climate change, as well as many other peripheral fields (Mlynowski et al., 2011).

In 2008, the World Meteorological Organization (WMO) established guidelines for minimum hydrometric density by physiographic region (WMO, 2008; Table 1.1).

Table 1.1 Physiographic region and WMO minimum recommended area in km^2 per station

Physiographic Unit	**Streamflow**
Coastal	2750
Mountains	1000
Interior plains	1875
Hilly/undulating	1875
Small islands	300
Urban areas	–
Polar/arid	20000

Coulibaly et al. (2013) evaluated the Canadian National Hydrometric Network (CNHN) according to these guidelines and found only 12% of the nation's area met the minimum acceptable density. In 2010, an audit of the National Hydrometric Program (Government of Canada, 2010) found that of the countries compared (similarly developed nations), Canada had the highest Total Actual Renewable Water Resource and the lowest station density per water volume. A properly reviewed network could provide value to the government, ensuring evidence-based justification for keeping stations active, and an understanding of which stations are lower priority if a decision needs to be made quickly due to budget cuts. In the province of Nova Scotia, the hydrometric monitoring network has been assessed on four separate occasions between 1967 and 1985. Since the 1985 review, the network has shrunk by 22 stations, and has not been formally re-evaluated. In order to understand how well the province's hydrometric network is performing, we first have to characterize the dominant hydrologic regimes in the region and level of variability that can be expected. Catchment classification, which refers to the grouping of similarly behaving catchments, is a useful tool to organize and understand hydrologic variability.

1.2 SCOPE AND OBJECTIVES

The overall objective of this thesis is to provide information and tools to support optimization of the Nova Scotia hydrometric monitoring network. The specific objectives of the thesis are:

- Develop a catchment classification system for gauged Nova Scotia watersheds that can be extended to ungauged watersheds,
- Apply the system across all Nova Scotia watersheds and evaluate the representativeness of the current hydrometric monitoring network.

1.3 THESIS STRUCTURE

This thesis is made up of five chapters, the first two chapters are introductory, followed by two manuscript chapters, and a concluding chapter. Chapter 1 outlines the context of why an efficient hydrometric network is important and the objectives the research will address. Chapter 2 provides a brief literature review on hydrometric networks and a tool to organize hydrologic variability through catchment classification.

In Chapter 3, we develop a catchment classification system for Nova Scotia watersheds. In Chapter 4, we apply the previously developed catchment classification to all secondary watersheds and the current hydrometric network to assess its representativeness. Chapter 5 links the two manuscripts together, with conclusions and recommendations for future work.

Chapter 2 Literature Review

2.1 CATCHMENTS AND THEIR PROPERTIES

In hydrology, catchments are the primary spatial unit of measure and also act as a practical management unit for decision-makers. A catchment is spatially defined as the contributing area (or volume) that drains to a common point such as a outlet or surface depression. The unique physical properties of this contributing area influence the partitioning, storage, and release of water (Wagener et al., 2007). Precipitation is collected by a catchment, which acts as an integrator taking this input and generating a signal (discharge) based on the landscape and antecedent properties. These landscape properties can be broken into five general categories: topographic, geologic, soil, land cover, and climatic.

Topographic (or morphometric) attributes describe the physical features of the catchment. Catchment area, slope, and elevation characteristics are commonly used to characterize the terrain. For example, to describe elevation, Boscarello et al. (2016) used 5th and 95th percentile elevation, while Ssegane et al. (2012a) computed the mean, median, maximum, minimum, and standard deviation of elevations. Another example is the relief ratio, which Sawicz et al. (2011) calculated by $[Elev_{median} - Elev_{min}]/[Elev_{max} - Elev_{min}]$. Attributes describing the drainage network are also widely used. Chiverton et al. (2014) computed drainage path slope, longest drainage path length, and mean drainage path length, while Lane et al. (2017) computed several metrics related to Strahler stream order (mean, maximum, and percentage of stream length that is first-order). Post and Jakeman (1996) calculated drainage density which is given by the ratio of total stream length to catchment area. Lastly, there are several metrics defining watershed shape. Jin et al. (2017) computed the length, width, form factor, elongation ratio, and circularity ratio. Lane et al. (2017) used a metric watershed compactness, which is given by the ratio of catchment area to the square of perimeter.

Geologic attributes describe the makeup of the surficial and bedrock geology of the catchment. Pyne et al. (2016) calculated the percentage of the catchment underlain by sedimentary, ultramafic, and volcanic bedrock, as well as the percentage of surficial rock composed of mountain glacial deposits. Laize and Hannah (2010) categorized both bedrock and surficial deposits as high, moderate (mixed), and low permeability. Sanborn and Bledsoe (2006) computed the average area-weighted minimum depth to bedrock.

Soils attributes typically describe soil properties controlling infiltration. Fractions of sand, silt, and clay have been used to characterize catchment soils (e.g. Addor et al., 2017; Beck et al., 2015), while some studies incorporate only percent sand (Sawicz et al., 2011). Alternatively, Allchin (2015) used eight classes of soil drainage: very poor, poor, imperfect, moderate, good, rapid, very rapid, or N/A. More generally, soils can be classified as high, medium, and low permeability (Boscarello et al., 2016; Yadav et al., 2007). Several studies computed more specific properties, such as field capacity, root depth, and saturated hydraulic conductivity, but these are limited by data availability (Ley et al., 2011; Boscarello et al., 2016; Sawicz et al., 2011, Addor et al., 2017).

Land cover attributes characterize important permeability and storage properties of the catchment. A heavily urbanized catchment will have a different response to precipitation input than a pristine, forested catchment. Kuentz et al. (2017) selected ten land cover variables (water, glacier, urban, forest, agriculture, pasture, wetland, open with vegetation, open without vegetation, irrigated) and computed the percentage of catchment area covered by each. Sanborn and Bledsoe (2006) computed the ratio of forest to urban area. Several studies have used vegetation metrics, such as the normalized difference vegetation index and leaf area index (Beck et al., 2015; Addor et al., 2017).

Climate metrics should characterize the meteorological forcing acting upon each catchment. Precipitation is supplied to the catchment and partitioned into different flow paths, stored, then released as streamflow, groundwater flow, and evapotranspiration (Waegner et al., 2007). Climate metrics can give insight into the amount of water supplied to the catchment, type of storage (e.g. as snow), and quantity released through specific pathways (i.e. evapotranspiration). Average monthly or seasonal temperature and precipitation are commonly used in the literature to characterize the climate, e.g. average June precipitation, average fall temperature (Ssegane et al., 2012b; Caratti et al., 2004). Precipitation may also be partitioned into rainfall and snowfall, as done by Sanborn and Bledsoe (2006). Toth et al. (2013) computed areal precipitation estimates using Thiessen-polygons, while Ley et al. (2011) calculated mean long-term potential evapotranspiration using raster data sourced from the Hydrological Atlas of Germany for their catchment analysis. Precipitation and evapotranspiration can be combined to give an indicator of water availability. For example, Ssegane et al. (2012b) computed an annual dryness index (ADI), given by the ratio of mean annual precipitation to mean annual evapotranspiration. Other studies incorporate radiation, e.g. mean annual or summer solar radiation (Kennard et al., 2010a; Caratti et al., 2004).

2.2 HYDROMETRIC NETWORKS

Hydrometry refers to the measurement of water, from the Greek *hydor* (meaning water) and *metron* (meaning measure). This can encompass measuring water inputs through rainfall, or loss through evaporation, or as it most often describes, flow in open channels. In this thesis, *hydrometric* will refer to measuring flow in open channels. Although early hydrometric networks were generally purpose built, e.g. for dam construction and operation (Nemec and Askew, 1986), networks have since evolved to meet modern objectives (e.g. regionalization, climate change detection). A variety of approaches have been applied for efficient network design and review (e.g. statistical-based methods, information-theory, user survey, physiographic components, sampling strategies, and hybrid; Mishra and Coulibaly, 2009). Stations can have specific intended functions and network design, and review can evaluate a station's fitness-for-purpose for its desired function, e.g. infrastructure design and operation, hydrologic regionalization, or detection of climate trends. Hydrologic regionalization refers to the classification of catchments into groups, allowing for extrapolation of information from gauged donor to ungauged receptor catchments.

2.3 REFERENCE NETWORKS

Reference, or benchmark, networks are often a subset of a larger network and typically include "pristine" catchments (< 10% affected area) with stable land use (Whitfield et al., 2012). In the United States, Langbein and Hoyt (1959) as cited in

Whitfield et al. (2012) recommended a hydrologic benchmark network (HBN) to record natural hydrological change and provide a baseline for future studies. The United States Geological Survey (USGS) HBN was created less than ten years later (Cobb and Biesecker, 1971, as cited in Whitfield et al., 2012). Bradford and Marsh (2003) defined a network of benchmark catchments in the United Kingdom, with the objective of maintaining catchments suitable for detecting climate-driven trends. The UK Benchmark Network (UKBN) was designed using the following criteria (Bradford and Marsh, 2003):

- *Relatively natural flow regimes;*
- *Good and consistent hydrometric data quality;*
- *Relatively long records (ideally > 25 years);*
- *Representative of UK hydroclimatic conditions with good geographical coverage.*

In Canada, select stations are designated as part of the Reference Hydrometric Basin Network (RHBN) and are mandated to be maintained for the use in the "*detection, monitoring, and assessment of climate change*" (Brimley et al., 1999). The inclusion of a reference network in a hydrometric program should be accounted for in network design.

2.4 DECLINE IN HYDROMETRIC NETWORK DENSITY

Researchers worldwide have lamented the loss of hydrometric stations (Lanfear and Hirsch, 1999 as cited in Vörösmarty, 2002; Shiklomanov, Lammers, & Vörösmarty, 2002). Frequently, cuts to station operations coincide with economic downturn. The recent decline comes at a particularly worrisome time, given intensifying climate change and its impact on the hydrological cycle (Huntington, 2006). The impact of insufficient and declining hydrometric networks has been compounded by anthropogenic changes to the climate and landscape, which has led to the initiation of the International Association of Hydrological Sciences (IAHS) Decade on Predictions in Ungauged Basins (PUB) (Sivapalan et al., 2003). One of the PUB objectives was to raise awareness of the value of data, particularly hydrological variables, and stress the impact of station closures on prediction uncertainty. Pearson (1998) detailed the re-design of New Zealand's national hydrometric network in the 1990s, pointing out the alarming worldwide trend of declining hydrometric sites. According to Spence et al. (2007), a reduced network may increase uncertainty associated with streamflow estimates, potentially resulting in the over-design of structures.

2.5 CANADA'S NATIONAL HYDROMETRIC NETWORK

A hydrometric network has operated in Canada in some capacity since the late nineteenth century. Its present-day iteration, the National Hydrometric Network, has become the principal source of surface water quantity information in Canada. Hydrometric monitoring began humbly in Canada in 1894, when the Department of the Interior began taking hydrometric measurements for irrigation planning in Alberta and Saskatchewan (Government of Canada, 2009). A formal approach was initiated in 1908 to measure the country's water resources (Canada and Canada, 2007). As demand increased, the network expanded, reaching all provinces and territories by 1922 and 1944, respectively (Scott et al., 1999). In the past, network density has correlated strongly with economic prosperity, experiencing a boom between 1910 and 1930, with a six-fold increase in active stations, which was later reduced by nearly a third during the Great Depression, only to double in size following World War II (Scott et al., 1999). The establishment of the International Hydrological Decade (1965 – 1974) saw a rapid increase in stations, from 1250 in 1960 to 3000 by 1974. In 1975, existing cost-sharing and data collection structures between the federal and provincial/territorial governments were formalized with the *Federal/Provincial Cost-Share Agreements on Water Quantity Surveys* (Scott et al., 1999). Fisheries, navigation, federal lands, interjurisdictional waters, and relations with foreign governments fall under federal jurisdiction. Matters including those related to property, civil rights, and the management and sale of public lands are under the jurisdiction of provincial government (Government of Canada, 2009). The network peaked in 1984, with over 3400 active stations, but was greatly reduced as a cost-saving measure in the 1990s (Scott et al., 1999). This was discussed by Pilon et al. (1996) who described how reductions of the 1990s were brought on by increased budget pressures presented by the 1995 federal budget. The national hydrometric program we know today is operated by Water Survey of Canada (within the National Hydrologic Services, Environment and Climate Change Canada [ECCC]). The agency is responsible for the collection, interpretation, and dissemination of water resource data and information.

2.5.1 NOVA SCOTIA HYDROMETRIC NETWORK

Hydrometric records have been collected in Nova Scotia since May of 1915 (Environment Canada and NS DOE, 1985). At that time, the primary purpose was the initiation of hydroelectric development in Nova Scotia. Since its establishment, the network has been assessed by various parties on four separate occasions. The first review, “An Assessment of the Hydrometric Network of the Province of Nova Scotia, Present and Future”, was completed in 1967 by departmental staff at Energy, Mines, and Resources. Then, in 1970, Ingledow & Associates produced a report on the hydrometric networks of the Atlantic provinces, entitled “Hydrometric Network Plan for the Provinces of Newfoundland, New Brunswick, Nova Scotia, and Prince Edward Island”. Ingledow et al. (1970) divided the Atlantic provinces into hydrologic zones. Nova Scotia was portioned in three zones: 2B – Southwestern Nova Scotia, 2C – Northeastern Nova Scotia (excluding Cape Breton Highlands), 3 – Cape Breton Highlands (Figure 2.1).

Figure 2.1 Hydrologic zones in Nova Scotia proposed by Ingledow (1970)

The Ingledow report recommended the addition of a minimum of five stations to the existing 39 station network:

- One station located on Bear River (or alternative) to allow correlation with a discontinued long-term gauge;
- Three stations near the Minas Basin, tentatively Kennetcook River, Economy River near Economy (or Great Village River near Great Village), and Apple River at East Apple River;
- The reactivation of Stewiacke at Upper Stewiacke or establishment of a station on the boundary of Zone 2C (e.g. Tangier at Tangier).

Of these suggestions, only one has been implemented; in 1993 a station was installed at Great Village River near Scrabble Hill.

In 1984, the Inland Waters Directorate, Water Resources Branch, Hydrology Division, of Environment Canada completed a report on the spatial distribution of stations by drainage area, entitled "An Analysis of the Hydrometric Networks of the Atlantic Region". The most recent, and likely most comprehensive review was conducted jointly by Environment Canada and Nova Scotia Department of Environment in 1985 to assess streamflow, meteorological, and groundwater monitoring networks. The objective was to determine the adequacy of these networks for providing data for the management of Nova Scotia's water resources. By 1985, the network had expanded to 53 active stations, up from 39 in 1970. In terms of density, there was roughly one station for every 1000 square kilometres. The 1985 review found little potential for data transfer between stations, but provided several recommendations relevant to this project which are given below:

- Distribute long-term stations between the three hydrologic zones determined by Ingledow et al. (1970);

- Establish low-flow partial record stations in ungauged watersheds, co-located with groundwater monitoring wells;
- Perform full analysis of stream gauging program;
- Undertake study to define relationships between stations within a hydrologic zone;
- Adopt WMO guidelines;
- Prioritize establishing stations in areas determined to be deficient in previous network reviews;
- Undertake studies to identify stations monitoring flow below the confluence of streams from physically distinct catchments, then establish stations on the identified streams (nested stations);
- 1) monitor smaller catchments throughout the southern and western portion of zone 2B; 2) monitor small and medium sized "basins" in the western portion of zone 2C;
  3) monitor smaller catchments in zone 3;
- Assess degree of regulation for long-term regulated stations and determine if station should be maintained;
- Increase monitoring of small coastal catchments; and
- Investigate the need for a sub-network of manual stations.

Although these recommendations are largely still relevant today, since 1985 the network has shrunk to 31 active stations, translating to roughly one station per 1800 square kilometres. Existing deficiencies may have been amplified and new gaps created. Few of the recommendations in the 1985 report have been implemented. The first recommendation is concerning the distribution of long-term stations amongst the province's hydrologic zones, which were delineated primarily by precipitation. The present-day configuration includes two RHBN stations in the Cape Breton Highlands, four in Northeastern Nova Scotia, and five in Southwestern Nova Scotia. The two Cape Breton Highlands RHBN stations gauge similarly sized drainage areas; there is another long-term station on the Cheticamp River, but the river is supplemented with additional flows. Evidently, this recommendation has not been applied to the Cape Breton Highlands. Another recommendation suggests identifying physiographically diversified watersheds that could be good candidates for nested stations. Three such pairings exist in the current configuration, all located in the Southwestern zone: Annapolis River at Wilmot and Lawrencetown, Mersey River below George Lake and Mill Falls, and Sackville River at Bedford and Little Sackville River at Middle Sackville. The 1985 review also suggests an increase in monitoring of small and medium catchments throughout all zones. Located in the interior highlands, the smallest gauged zone 3 catchment is 123 km^2. Smaller and coastal catchment flow regimes are not adequately captured in zone 3.

2.6 RATIONALIZATION AND REVIEW

In recent years, researchers have attempted to review and rationalize hydrometric networks. Rationalization efforts may focus on improving efficiency by removing stations with redundant information (e.g. Burn and Goulter, 1991) or determining the minimum required station density for an area (Karasseff, 1986). Mishra and Coulibaly (2009) summarized six standard methods for hydrometric network review and design: 1) statistical-based methods, 2) information-theory, 3) user survey, 4) physiographic components, 5) sampling strategies, and 6) hybrid.

Many authors have applied one or more of these techniques in national network reviews. For example, Black et al. (1995) completed a comprehensive review of the Northern Ireland hydrometric network, including a theoretical assessment of station representativeness, a data use survey, use of spatial data to transfer hydrological parameters, a density comparison with other jurisdictions, and lastly a data quality review. In the UK, Hannaford et al. (2013) evaluated the hydrometric network in terms of suitability for regionalization. Their method allowed for the identification of high and low value stations for regionalization -- specifically, stations which could enhance regionalization with upgrades to improve data quality, and redundant stations that did not contribute to regionalization and could potentially be removed without serious information loss. The authors recognized that a qualified person's judgement is required in this approach, as some purpose-built stations do not benefit regionalization, but have other value.

Information theory, developed by Shannon (1948), refers to the measurement of information content in a dataset. Several researchers have applied information theory techniques to hydrometric network analysis, but these methods may be difficult for decision-makers to understand and replicate (Coulibaly et al., 2010; Li, Singh, & Mishra, 2012; Alfonso, 2013; Boisvert et al., 2017). The entropy concept is a popular branch of information theory increasingly applied to hydrometric network design. Entropy refers to the randomness or lack of predictability of a system. In Shannon's information theory, system entropy will decrease given more information. For example, Li et al. (2012) used an entropy theory-based criterion to maximize the transinformation between gauging stations, while minimizing duplicate information.

In addition to performing network review for a primary objective (e.g. coverage of physiographic components), researchers have considered secondary factors that should be included in hydrometric network design and review. According to Mishra and Coulibaly (2014), seasonality should be incorporated into network design, as network efficiency is season dependent. It has been observed that networks designed using streamflow signatures tend to emphasize headwaters, while incorporating indicators of alteration emphasized downstream and disturbed regions (Leach et al., 2015).

2.6.1 CANADIAN EXAMPLES

Burn and Goulter (1991) developed a network rationalization approach and applied it to stations in the Pembina River Basin in Manitoba. Stations were clustered by hydrologic characteristics, and one representative station from each cluster was retained. The authors rationalized the network for three cases: a) similarity in extremes only; b) similarity in seasonal values (annual flows) only; c) similarity in extremes and seasonal values weighted equally. A station was retained for each cluster based on the following criteria: 1) overall similarity with other stations in the cluster; 2) number and types of uses of station; 3) unique users; 4) record length; 5) data quality; and 6) drainage area. Mlynowski et al. (2011) evaluated hydrometric density in the Canadian pan-Arctic, recommending the addition of stations on smaller rivers. Boisvert et al. (2017) attempted to rationalize the New Brunswick hydrometric network by clustering stations by their generalized extreme value (GEV) shape parameter for the peak annual flow series, then applying an entropy method to calculate shared information between stations. Philip and McLaughlin (2018) evaluated hydrometric station density in Northern Ontario and found that none of the *interior plains* (which most of Northern Ontario is designated as) watersheds satisfied World Meteorological Organization (WMO) guidelines.

2.7 CATCHMENT CLASSIFICATION

Catchment classification refers to the grouping of similarly behaving catchments for organization and data transfer. In limnology, lakes can be classified in terms of productivity (using trophic status) or thermal stratification (using Hutchinson and Löeffler's 1956 scheme). Yet in hydrology, a universally accepted classification system has not yet been realized for the discipline's primary unit of measure, the catchment. Sivapalan (2006) describes the catchment as "a self-organizing system, whose form, drainage network, ground and channel slopes, channel hydraulic geometries, soils, and vegetation, are all a result of adaptive, ecological, geomorphic or landforming processes". Classification is based on the premise that physically analogous catchments should respond similarly to precipitation input, and the dominant process by which runoff is generated are similar, i.e. hydrologic similarity. A catchment classification system would enable straightforward comparison of catchments and could permit the extension of data in locations with limited available data (McDonnell & Woods, 2004).

The motivation behind the development of a common system is most simply explained by Grigg (1965; 1967) as cited in Sawicz (2013), who gave three primary advantages of a classification system:

- To give names to things;
- To permit transfer of information; and
- To permit the development of generalizations.

A comprehensive catchment classification system would provide researchers with means to organize the immense variability in time, space, and process inherent in hydrologic systems around the world (McDonnell & Woods, 2004).

Several researchers have made arguments for their preferred theoretical basis for classification, i.e. what metrics similarity should be judged on. McDonnel and Woods (2004) suggested classifying measures of fluxes, storages, and response times. On the other hand, Wagener et al. (2007) argued that classification should include static characteristics of the catchment's 'form' and 'forcing' (physical and climatic, respectively), alongside dynamic response characteristics of 'function' (hydrologic). The literature suggests catchment similarity metrics should be chosen on a regional or situational basis (Ali, Tetzlaff, Soulsby, McDonnell, & Capell, 2012), which is at odds with the demand for a common system. The area of hydrological classification encompasses a large range of techniques to characterize hydrologic properties. The two broad classification approaches distinguished in Olden, Kennard, & Pusey (2012), (the application of deductive and inductive reasoning) are discussed below.

2.7.1 INDUCTIVE APPROACHES

Inductive classification approaches use direct measures of hydrology as similarity metrics, e.g. streamflow signatures (Olden et al., 2012). Unfortunately, hydrometric networks are not always robust enough to capture all watersheds in a jurisdiction (due to limited spatial coverage and quality of discharge data) and can thus be limiting. For example, Nova Scotia is divided into three principal drainage basins: portion of the mainland draining into the Bay of Fundy or Northumberland Strait, portion of the mainland draining into the Atlantic Ocean, and Cape Breton Island. These three principal drainage basins are further divided into 44 primary watersheds (generalized drainage basins that do not meet the definition of a watershed), and again into 246 secondary watersheds. Secondary watersheds refer to the major watersheds draining the hinterland of the province and terminating at a single point where the main stem meets the ocean. The province has only 31 active gauging stations to represent these 246 secondary watersheds and thousands of tertiary and shoreline direct watersheds. It is unlikely that the full range of flow regimes present in the province can be captured using an inductive approach alone. Much of the application of inductive approaches is done in the context of ecohydrology.

Kennard et al. (2010a) completed the first continental-scale classification of flow regime in Australia using an inductive approach, forming 12 distinct hydrologic classes. The authors found that some hydrologic classes were distributed non-contiguously across the country, emphasizing that caution should be exercised when transferring information from gauged to ungauged catchments, regardless of spatial proximity. Similarly, from an ecohydrological standpoint, Monk et al. (2011) identified six distinct hydrologic classes using agglomerative hierarchical clustering with data from 172 RHBN stations across Canada. Peñas et al. (2016) carried out three inductive classifications based on flow metrics derived from daily, monthly, and modelled monthly flow series. Due to strong correlation between flow metrics, the variables were first reduced using principal component analysis (PCA). Daily and monthly flow metrics were clustered using the Partitioning Around Medoids (PAM) algorithm. The authors found that classification based on monthly flows performed comparably to that of daily flows. However, monthly flows did not perform as well for discriminating low flow conditions and may result in the loss of important hydroecological information. The inductive classifications were evaluated against a fourth expert-driven classification, achieved using rules developed by a qualified person, rather than through impartial multivariate statistical methods. The daily and monthly inductive classifications were found to substantially outperform the expert-driven classification.

2.7.2 DEDUCTIVE APPROACHES

Deductive classification approaches are frequently applied when measured or modelled data is either unavailable or unreliable, e.g. in the case of a sparse, poorly maintained hydrometric network (Olden et al., 2012). This approach relies on indirect environmental surrogates for hydrology, e.g. metrics describing elevation or hydraulic conductivity. Deductive classification benefits from the availability of high-quality hydrologically applicable datasets, such as climate, topography, soils, geology, and land use. Olden et al. (2012) divides deductive reasoning approaches into three groups:

- Environmental regionalization;

- Hydrologic regionalization; and
- Environmental classification.

Environmental regionalization refers to an approach that relies on existing maps and spatial data in place of flow data. The result is a simple spatial representation of hydrologic similarity, where regions are considered uniform at a specific scale, with regards to designated environmental characteristics (Olden et al., 2012). *Hydrologic regionalization* refers to an approach that attempts to extrapolate information from gauged to ungauged regions (Olden et al., 2012). This is achieved by relating catchment characteristics to hydrologic indices within groups that are thought to have comparable hydrologic response (Olden et al., 2012).

Environmental classification or environmental domain analysis (Mackey, Berry, & Brown, 2008) refers to an approach that classifies catchments based on physical and climatic attributes known to influence streamflow. This approach can be employed in regions lacking a comprehensive hydrometric network and is often geographically independent. However, it can be limited, in that some properties of catchment response are inadequately represented by environmental surrogates due to the coarse resolution of existing data (Olden et al., 2012).

Winter (2001) first introduced the concept of hydrologic landscapes, made up of various fundamental hydrologic landscape units (FHLUs). A FHLU is defined by a) "its land-surface form of an upland adjacent to a lowland separated by an intervening steeper slope", b) "its geologic framework", and c) "its climatic setting". Wolock et al. (2004) demonstrated the hydrologic landscape concept, grouping 43,931 small watersheds (~200 km^2) into 20 non-contiguous hydrologic landscape regions (HLRs). Building on these techniques, Wigington et al. (2013) developed a hydrologic landscape classification approach for Oregon, USA. Instead of FHLUs, the authors delineated *assessment units* (headwater catchments or areas draining directly into streams) with a minimum threshold of 25 km^2; an area small enough to ensure reasonably consistent physiographic features and large enough for meaningful terrain indices. Then, metrics of climate, aquifer, soil, and terrain were computed for all *assessment units* and 157 unique hydrologic landscapes (HL) were formed by integrating the indices. For example, HLs that are semiarid, winter maximum surplus, with high and moderate aquifer permeability account for almost one-third of Oregon's land area. Since single assessment units were unlikely to be gauged, the authors could not directly compare specific hydrologic landscapes to annual hydrographs. Instead, the distribution of HLs within larger gauged watersheds was compared to indices describing the shape of the 30-year mean annual hydrograph of each watershed. Leibowitz et al. (2016) revised the approach of Wigington et al. (2013), suggesting their approach is more applicable, given its flexibility. This approach was also applied to Bristol Bay, Alaska by Todd et al. (2017), who were able to show that watersheds with similar HL distributions had similar runoff-based classifications.

Laizé's (2004) novel representative catchment index (RCI) relies on a similar premise to hydrologic landscapes. However, instead of using catchment-based *assessment units*, RCI aggregates data on a much smaller-scale, individual cells. Raster datasets are selected to capture hydrologic response and resampled at the chosen cell size to reduce computational burden (e.g. 200 m × 200 m). The rasters are then reclassed according to user objectives, with hydrologically significant classes (e.g. land covers producing similar runoff are grouped together as one class). Finally, the rasters are overlain, resulting in an xdimensional array (where x is the number of spatial datasets) with unique combinations of hydrologically significant classes. The distribution of these unique combinations (like a HL) within a given watershed is compared to that of the reference area, resulting in a similarity score expressed as a percentage. This approach could be easily applied by non-scientific personnel, as it is easily interpreted (Laize, 2004).

2.7.3 MULTIVARIATE STATISTICAL METHODS FOR CLASSIFICATION

Numerous multivariate methods have been applied for catchment classification. Complexity ranges from relatively simple multiple regression, to popular hierarchical, k-means, or Bayesian cluster analysis, to more complicated and data intensive methods such as, random forest (Brown, Lester, Versace, Fawcett, Laurenson, 2014), classification and regression trees (CART; Flores, Bledsoe, Cuhaciyan, Wohl, 2006) and self-organizing maps (SOMs; Wallner, Haberlandt & Dietrich, 2013). The latter generally require much larger datasets. In addition to methods used to group catchments, many studies apply ordination approaches, such as PCA, to their initial explanatory and response variables. This allows a reduction in dimensionality and variable redundancy, by condensing many similarity metrics into few principal components which are linear combinations of the original metrics (Toth et al. 2013). Ssegane et al. (2012a) examined variable selection methods, comparing five causal selection methods to stepwise regression and PCA. They found variables selected by stepwise regression had high predictive potential but lacked accuracy. The use of more than one selection method was recommended to improve reliability. The authors noted that neither stepwise regression nor PCA are meant to determine the underlying relationship between the response and explanatory variables. In a follow up study, Ssegane et al. (2012b) found a single causal selection method was able to isolate variables unique to each of the three eco-regions examined and maintain classification performance.

2.7.4 STREAMFLOW SIGNATURES

Streamflow signatures (also called hydrologic indices, streamflow metrics, dynamic response characteristics, etc.) are statistics derived from observed or modelled hydrological time series (e.g. streamflow, precipitation, evaporation) that describe hydrological function (McMillan et al., 2017). A robust streamflow signature suite should describe all parts of the annual hydrograph, as well as finer-scale elements not necessarily apparent from a visual inspection of the hydrograph: (1) magnitude, (2) frequency, (3) duration, (4) timing, and (5) rate of change of certain water conditions on an annual basis (Richter et al., 1996). Significant research has focused on the use of streamflow signatures for similarity analysis in ecohydrology (e.g. Olden and Poff, 2003; Monk et al., 2007; Kennard et al., 2010b; Pyne et al., 2017). Olden and Poff (2003) reviewed 171 streamflow signatures with the objective of determining a suite that adequately described flow regimes by the five constituents given by Richter et al. (1996). Many other studies have similar goals (e.g. Yadav et al., 2007; Detenbeck, 2005). The five fundamental characteristics describing flow regime must represent the full range of flow conditions, e.g. the magnitude of average and low flows or the timing of peaks flows. Patil and Stieglitz (2011) observed hydrologic similarity was dependent on flow conditions; similarity was better preserved at higher flow conditions than low flow conditions.

There does not appear to be a clear consensus on which streamflow signatures or equivalent hydrologic similarity measures should be applied in this context. McMillian, Westerberg and Branger (2017) suggested a more rigorous approach to signature selection. They argued signature selection justification applied in past studies have fallen into five categories:

- Subset of previous study;
- Capture broad range of information;
- Nonredundant signatures;
- Signatures related to hydrological function; or
- No reason given.

The authors instead propose the following guidelines for selection:

- Identifiability – uncertainty should be small compared to the values seen across the dataset;
- Robustness – signatures should be robust against data collection design, i.e. a different installation type or length of time should not yield different results;
- Consistency – signatures should not be sensitive to extraneous influences, e.g. normalize signature by catchment area;
- Representativeness – signatures should be representative of catchment-scale processes; and

- Discriminatory power – functionally similar catchments should have similar signature values.

Keeping these points in mind, one should consider the number and breadth of signatures required to fully characterize flow regime, while avoiding redundancy. This number varies considerably throughout the literature; Monk et al. (2011) used 32 ecologically important signatures to group 211 RHBN rivers, while other studies (Razavi and Coulibaly, 2013; Ssegane et al., 2012a) used a subset of six signatures presented by Sawicz et al. (2011). The Sawicz subset was selected from a larger group given in Yadav et al. (2007): runoff ratio, slope of flow duration curve (FDC), baseflow index, streamflow elasticity, snow day ratio, and rising limb density. Sawicz et al. (2011) stressed the importance of reducing redundancies and ensuring signatures have an interpretable link to catchment function. The Richards-Baker Flashiness Index (RBI), which quantifies changes in flow relative to total flow (Baker et al., 2004), could be correlated with catchment shape or imperviousness. In a similar manner, baseflow index (BFI) is the long-term ratio of baseflow to streamflow, which may correspond with surficial and bedrock geology and catchment size. Ensuring non-redundant signatures might come at a later stage of the analysis; an initial suite of signatures can be chosen and further refined using a method like PCA. Table 2.1 is a non-exhaustive list of common streamflow signatures used for classification in the literature.

Table 2.1. Streamflow signatures in literature

Streamflow Signature	**Flow Condition**	**Regime Characteristic**	**Published Example**
Flow duration curve (FDC) percentiles (e.g. 5th, 50th, 95th)	All	Magnitude	Patil and Stieglitz (2010)
FDC slope (e.g. 0.33 – 0.66)	-	Rate of change	Boscarello et al. (2016)
Mean annual flow	Average	Magnitude	Lane et al. (2017)
Baseflow index	Low	Magnitude	Beck et al. (2015)
Standard deviation	-	Magnitude	Toth (2013)
Coefficient of variation	-	Magnitude	Kennard et al. (2010b)
Runoff coefficient	-	Magnitude	Carrillo et al. (2011)
Streamflow elasticity	-	Rate of change	Sawicz et al. (2011)
Mean half-flow date	-	Timing	Knoben et al. (2018)
Reversals	-	Rate of change	Oueslati et al. (2015)
High flow pulses	High	Frequency	Sanborn and Bledsoe (2006)
Low flow pulses	Low	Frequency	Sanborn and Bledsoe (2006)
High flow duration	High	Duration	Lane et al. (2017)
Low flow duration	Low	Duration	Lane et al. (2017)
Seasonal mean flow (e.g. spring)	-	Magnitude	Harrigan et al. (2017)

Some authors have avoided using summary indices, instead examining the temporal dependence structure of a catchment. In this context, temporal dependence refers to the likeness of streamflow on a given day to the previous days. Chiverton et al. (2015) represented the temporal dependence structure of streamflow using semi-variograms. Likewise, Toth et al. (2013) used two metrics characterizing the correlation structure of streamflow: lag-1 autocorrelation coefficient, and correlation scaling exponent. Another method is to describe catchments by parameters derived from fitting specific flow metrics to probability distributions; Boisvert et al. (2017) used the generalized extreme value (GEV) distribution shape parameter for annual maximum flows.

An equally important consideration, independent of signature selection, is the length of record used to derive signatures. Researchers have attempted to find a compromise between maximizing the number of study stations and maintaining adequate record length; longer records are more likely to capture extreme events and limit uncertainty, making them more valuable to users. Beck et al. (2015) used stations with a minimum record length of ten years but did not require consecutive years. Pyne et al. (2017) used stations with a minimum record length of five years to derive signatures that predominantly described average flow conditions. Detenbeck et al. (2005) used a very short record length of only one growing season but acknowledged this would not be appropriate for calculating extreme flow metrics. A common minimum record length is 15 years inside of a specified temporal window (e.g. 1980 – 2010, 30 years) (Brown et al., 2014; Lane et al., 2017). Locally, Nova Scotia Environment’s (NSE) Guide to Surface Water Withdrawal Approvals requires at least twenty years of data from within the last 30 years for hydrological assessment for calculation of flow signatures. Kennard et al. (2010b) concluded that a record length of fifteen years was suitable for hydrologic similarity work and ideally, discharge records should have at least a 50% overlap within the specified temporal window. Liermann et al. (2011) confirmed the findings of Kennard et al. (2010b) that a 15-year record length provides an acceptable estimate of streamflow signatures, with each additional year providing minor improvements in bias, precision, and accuracy.

The final consideration in signature derivation is streamflow normalization. Streamflow quantity is governed primarily by watershed area; thus, discharge should be normalized to give streamflow per unit area or depth over the watershed, enabling comparison between catchments of varying sizes. This can be achieved by dividing streamflow by drainage area (e.g. Ssegane et al. 2012b), or through dividing streamflow by mean or median flow (e.g. Yadav et al., 2007). However, using mean or median flow to normalize streamflow could obscure differences between catchment responses.

2.7.5 CATCHMENT ATTRIBUTES AND CLIMATE METRICS

The deductive reasoning approach relies on environmental surrogates for describing the catchment's *form* and *forcing* (catchment attributes and climate metrics, respectively). Catchment attributes may describe topography, geology, soils, or land cover, while climate metrics describe meteorological input and influence on the catchment. Allchin (2015) examined criteria for selecting appropriate spatial datasets for this purpose: datasets should be publicly available, of good quality, consistent, have credible provenance, and adequate spatio-temporal extent and resolution. Choice of catchment attributes and climate metrics are mainly constrained by available data.

Catchment attributes and climate metrics are often used to judge catchment similarity. As discussed previously, the hydrologic landscapes concept was applied to the United States by Wolock et al. (2004). Land-surface form attributes were derived from a 1-km resolution DEM (relief, total % flatland, % upland, % lowland). Geologic texture was characterized by soil and bedrock permeability. Soil permeability was approximated by percent sand, while bedrock permeability was characterized by lithologic class (not a principal aquifer, sandstone, consolidates sand, basalt/other volcanic rocks, sandstone/carbonate rocks, unconsolidated sand/gravel, and carbonate rock). Climate was characterized by the difference between mean annual precipitation (MAP) and potential evapotranspiration (PET). Wigington et al. (2013) built on Wolock's previous work, developing a modified hydrologic landscape approach for Oregon, USA. Their descriptors belonged to the same categories: climate, aquifer, terrain, and soil. For climate, the authors selected both a moisture and seasonality index. The moisture index classifies the assessment units from very wet to arid. The seasonality index assigns *assessment units* to the season with the most available water (fall/winter, summer, or spring). For aquifer characteristics, the authors created three permeability classes: i) low (median $K \leq 1.5$ m/day), ii) moderate (median $K > 1.5$ and ≤ 3 m/day), and iii) high (median $K > 3$ m/day) and assigned each unit the dominant class. Similarly, soil permeability within the top 10 cm was defined as high, moderate, and low. For terrain, the authors computed two metrics, relief and total percentage of flatland (defined as cells with <1% slope), then combined them to classify the terrain as mountain (flatland < 10% and relief > 300 m), flat (flatland >50%), or transitional (all remaining). Another study by Leibowitz et al. (2016) revised Wigington et al.'s (2013) approach, expanding it to the Pacific Northwest. The authors changed some parameters (e.g. permeability classes) for easier application across the country. Todd et al. (2017) modified the hydrologic landscapes approach once again, for a watershed in southwest Alaska, adding attributes relevant to the region's unique hydrologic features: large lakes and glaciers. The authors extracted waterbodies ≥ 5 km^2 and "ice mass" features from the National Hydrography Dataset (NHD) and computed lake and glacial influence indices: i) no influence (<5% coverage); ii) low influence (5 – 25% coverage); iii) high influence (>25% coverage).

For ecohydrology applications, Snelder and Biggs (2002) developed the River Environment Classification (REC) system for New Zealand Rivers. The REC uses a top-down approach to classify rivers in terms of ecological patterns and function. The six levels of the hierarchy, in order of dominance, are climate, source of flow, geology, land cover, network position, and valley landform. Climate is the dominant control, consisting of six classes describing temperature and water availability (e.g. warm extremely wet vs. cool dry). Source of flow is categorized as mountain, hill, low elevation, and lake. The geology level gives the dominant catchment geology: alluvium, hard sedimentary, soft sedimentary, volcanic basic, volcanic acidic, and plutonic. Land cover is divided into eight classes: bare, indigenous forest, pasture, tussock, scrub, exotic forest, wetland, and urban. Network position gives the stream order of the river of interest (low, middle, high). Lastly, valley landform describes the valley slope as high (> 0.04), medium (0.02 – 0.04), and low (> 0.02).

Complex climate metrics may be unreliable or difficult to calculate for Nova Scotia, given its marine influence and relative scarcity of long-term weather stations. The spatial representativeness of Nova Scotia weather stations is tied to the major spatial climate-forcing factors. Daly (2006) examined the effects of these factors (elevation, terrain-induced transitions, cold air drainage, and coastal zones) at difference spatial scales and found terrain and coastal effects to be greatest at scales < 10 km and lowest when scales > 100 km. The author recommended regions with significant coastal and terrain effects be mapped by experts as they likely cannot be handled using simple methods like ordinary kriging and inverse distance weighting. Gridded climate datasets are a popular vehicle for including variables like precipitation and temperature, e.g. PRISM datasets for the United States. In Canada, McKenney et al. (2011) generated climate models for North America using ANUSPLIN, which fits thin-plate smoothing spline to weather station data. While ANUSPLIN accounts for dependencies on elevation, predictors of coastal gradients (e.g. distance to water bodies) are not incorporated (McKenney et al., 2011). Long-term mean surfaces for temperature, precipitation, and additional bioclimatic variables are available.

Chapter 3 A Process-Guided Decision-Making Tool for Hybrid Catchment Classification

ABSTRACT

Even within relatively small geographic regions, catchments can display a considerable amount of hydrologic variability. Regional catchment classification allows for the quantification and organization of this variability and facilitates extrapolation of information from gauged to ungauged catchments. The primary objective of this study was to produce a decision-making tree for classifying ungauged catchments in a hydrologically diverse region (Nova Scotia, Canada). Initially, an inductive methodology was used to classify the hydrologic regime of 47 gauged catchments using 16 streamflow signatures describing the magnitude, frequency, duration, and rate of change of flows. Principal component analysis and partitioning around medoids cluster analysis were used to identify nine hydrologic clusters. These clusters were merged to form five generalized flow classes: highland baseflow, highland flashy, NS flashy, lake-influenced baseflow, and mainland transition.

Descriptors characterizing the geology, land-use, storage, and meteorological forcing on the catchments were then assembled for the 47 study catchments, and a classification and regression tree analysis was used to identify dominant factors and potential breakpoints for deductive classification. A deductive decision tree was manually constructed using soil drainage, large lake influence, hypsometric integral, the fraction of catchment with glaciofluvial sediments and drumlins, lake and wetland influence, and relief ratio as splits

in the decision tree. When tested on a sub-set of study catchments not used to construct the decision tree, 79% of the catchments were classified correctly. The classification framework used in this study relied on a combination of deductive and inductive approaches, but was guided by theoretical reasoning and judgement due to data limitations. The approach could be transferable for other regions with limited hydrometric networks.

3.1 INTRODUCTION

The concept of hydrologic catchment classification refers to grouping similarly behaving catchments together (Sivapalan, 2006). Hydrologic catchment classification can support a variety of outcomes, but one of the most common objectives is to enable the transfer of information from gauged to ungauged catchments. Often, geographic neighbours are thought to be the most appropriate surrogate for data transfer; however, in certain regions, this assumption may not be correct, and practical approaches for incorporating other factors to guide data transfer decisions are lacking. The characterization and classification of streamflow regimes can be very impactful for jurisdictions that may have abundant surface water resources, but relatively limited financial resources with which to monitor and understand them. However, a common classification system has not yet been realized for the discipline's main spatial unit, the catchment. Realistically, a universal classification system may be out of reach due to the level of hydrologic variability observed globally, and even nationally, and therefore regional-level classifications may be a more practical objective (Ali et al., 2012). A catchment classification system that could be applied to an entire region (gauged and ungauged catchments) would enable the straightforward comparison of catchments and could permit the extension of data to locations with limited available data (McDonnell & Woods, 2004).

Wagener et al. (2007) described the catchment in terms of its form, forcing, and function characteristics, where form describes static physical characteristics and forcing describes meteorological inputs to the catchment. Together, these elements influence the catchment's function, or hydrologic regime. A collection of catchments can be partitioned into classes using these characteristics to judge similarity. Olden, Kennard, and Pusey (2012) separate the two broad catchment classification approaches into those based on inductive or deductive reasoning. An inductive classification approach uses direct measures of hydrology (function), such as streamflow signatures, to define similarity. In contrast, a deductive classification approach relies on indirect environmental surrogates (form and forcing), such as physical catchment descriptors. Deductive classification is based on the premise that physically analogous catchments respond similarly to precipitation input and that the dominant processes by which streamflow is generated are similar. The deductive approach is especially useful in situations where the monitoring network is not extensive. These approaches can be paired together as a hybrid approach, in which inductively derived classes are related to environmental variables (Olden, Kennard, and Pusey, 2012).

Regional inductive classifications have been performed under a variety of hydroclimatic conditions, with classification objectives related to catchment-, operational-, and eco-hydrology outcomes (e.g. Kennard et al., 2010a; Monk et al., 2011; Sawicz et al., 2011; Toth, 2013; Peñas et al., 2016). The original deductive classification concept is Winter's

(2001) *hydrologic landscapes*, which has been extended across regions of the United States (Wolock et al., 2004; Wigington et al., 2013; Patil et al., 2014; Leibowitz et al., 2016; Todd et al., 2017). Paired or hybrid approaches that apply both inductive and deductive classifications are becoming increasingly common (Ley et al., 2011; Auerbach et al., 2016; Boscarello et al., 2016; Kuentz et al., 2017), with varying levels of success. Hybrid approaches are supported by inductive classification, while still allowing the classification to be extended to ungauged regions. Clustering is a popular multivariate classification method for catchments (Sawicz et al., 2011; Wolfe et al., 2019), but more complex methods such as self-organizing maps (Carrillo et al., 2013; Toth 2013) and random forests (Carlisle et al., 2010) have also been applied.

The objectives of this study were to: (i) develop a novel regional catchment classification system using both inductive and deductive approaches guided by theory; and (ii) apply this classification to form the basis for a decision-making tree to allow regulators and practicing water resources professionals to identify appropriate donor catchments for ungauged watersheds. We test our approach for watersheds within a Canadian province (Nova Scotia) characterized by a limited hydrologic monitoring network and diverse geology and climate.

3.2 STUDY REGION

Located in Atlantic Canada (Figure 3.1A), the province of Nova Scotia has a varied topography and climate, resulting in relatively diverse hydrologic regimes for its small size. With a land area of 52,942 km^2, Nova Scotia is Canada's second smallest province. The mainland, connected to the adjacent province of New Brunswick by the Chignecto Isthmus (a 23-km long strip of land), accounts for most of the land area reaching a maximum elevation of ~350 m in the Cobequid Mountains. Cape Breton Island located to the northeast of the province covers ~10,000 km^2, roughly 20% of the province's footprint, and has a peak elevation of just over 530 m (Nova Scotia Department of Natural Resources, Mineral Resource Branch, 2003). The bedrock geology of Nova Scotia is diverse, containing large areas of evaporitic, metamorphic, plutonic, volcanic, and sedimentary rocks (Figure 3.1B).

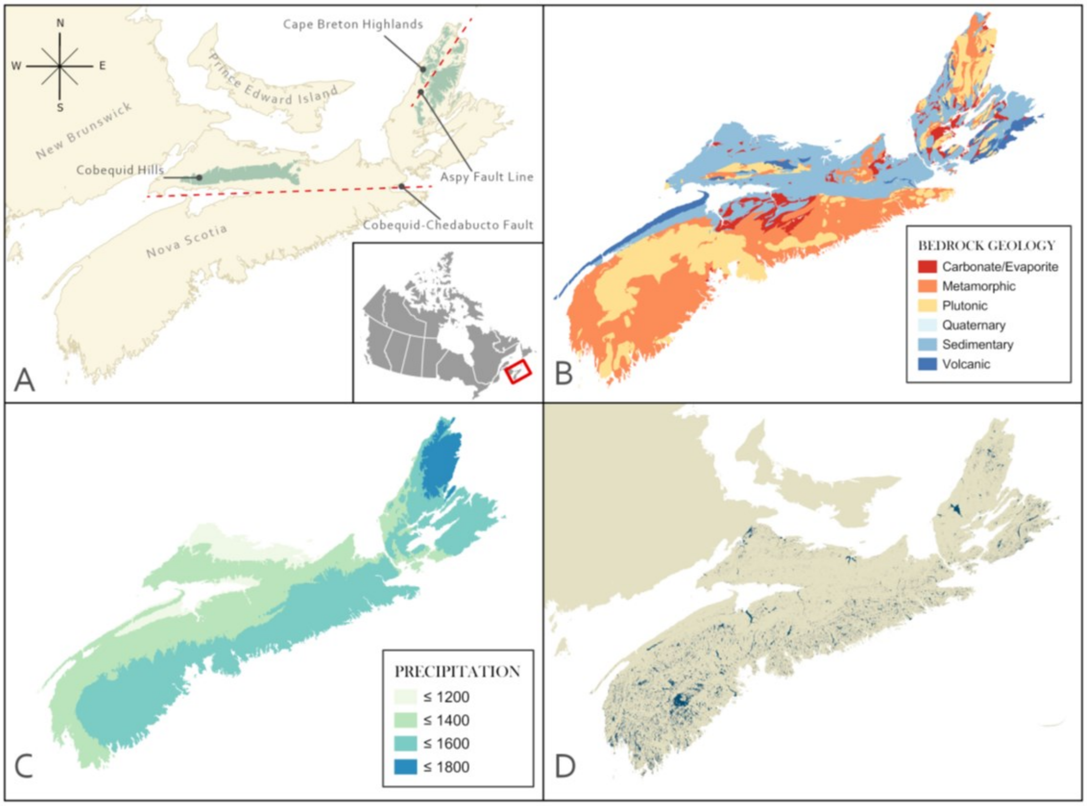

Figure 3.1. A) Map showing Nova Scotia's geographic position and key topographic points of interest; B) A simplified bedrock geology map of Nova Scotia, illustrating the complex geological setting; C) Variability of annual precipitation (mm) across Nova Scotia (McKenney et al., n.d.), D) Map of surface-water distribution

Nova Scotia is characterized by a humid, modified-continental climate, with higher precipitation input than loss through evaporation, owing to mild summers, frequent coastal fog, and moderation by the surrounding ocean (Nova Scotia Museum of Natural History, 1996). Mean daily air temperatures rarely exceed 20 °C in the summer or fall below -8 °C in the winter. Climate varies across the province, with higher precipitation in the Cape Breton Highlands and lower precipitation on parts of the mainland (Figure 3.1C). All of Nova Scotia is located within 70 km of the ocean, resulting in a strong maritime influence.

Nova Scotia has abundant surface freshwater resources in the form of lakes, rivers, and wetlands (Figure 3.1D). Nova Scotia lakes are largely the product of glaciation (Pleistocene), carved out by receding glaciers and/or created as a result of till deposition. Other types of lakes, including oxbow, levee, solution, Barachois, beaver dammed, and manmade lakes, are less dominant. Streams range in size from first-order headwater tributaries to larger seventh-order rivers. Nova Scotia wetlands are predominantly peatlands, with the highest density in the southwestern region of the province (T8.3 Freshwater Wetlands, Natural History of NS).

3.3 METHODOLOGY

An overview of the inductive-deductive procedure developed and applied in this study is provided in Figure 3.2. This approach was developed to characterize the relationship between catchment function and regional controls on hydrology and allow for the identification of appropriate donor catchments for ungauged catchments. The hydrologic regime of a group of gauged study catchments was first established using a suite of streamflow signatures, and similar hydrologic regimes were grouped together to form clusters. The form and forcing characteristics of these clusters were explored for notable features linked to hydrologic regimes that could distinguish catchments based on their respective clusters.

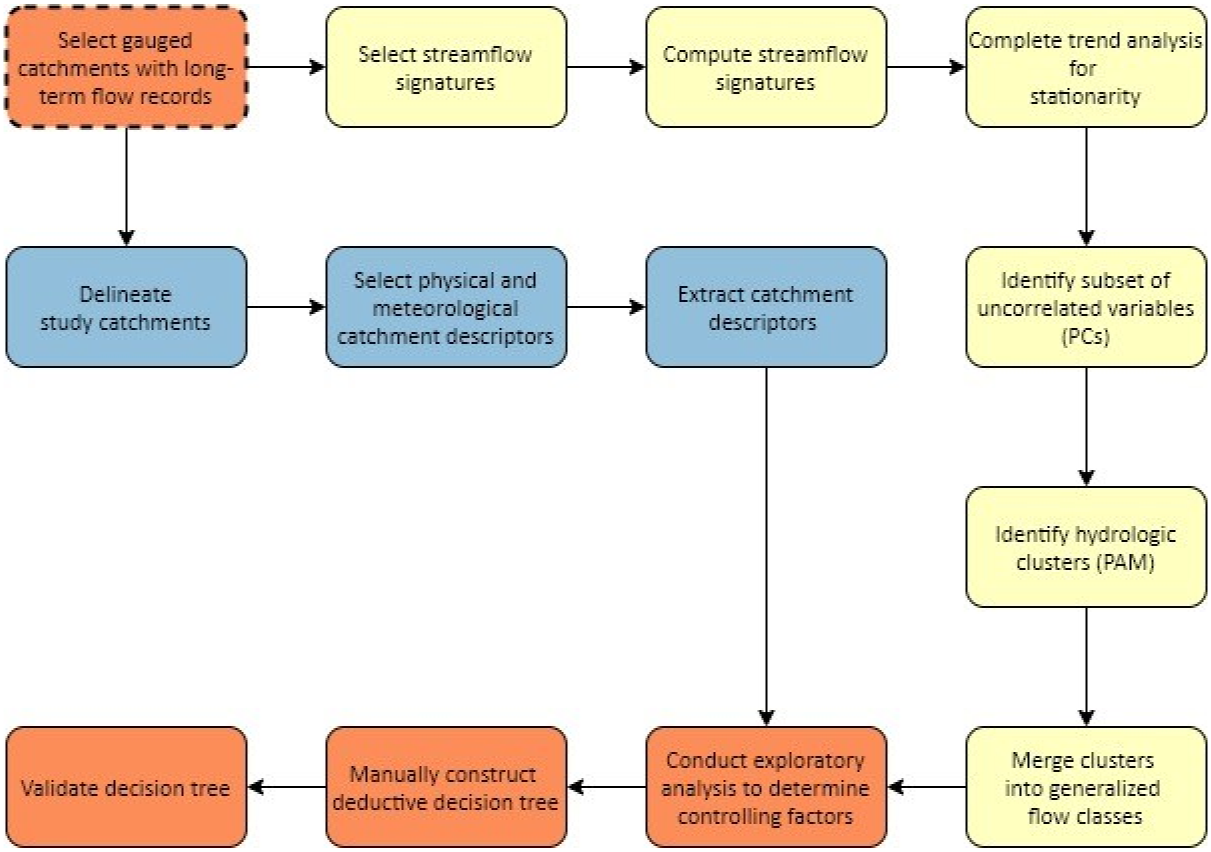

Figure 3.2. A step-by-step procedure for the inductive-deductive catchment classification used in this study. Actions relating to inductive classification are seen in yellow, actions relating to deductive classification are seen in blue, actions relating to both are seen in orange. PCs refers to principal components, and PAM refers to partitioning around medoids

3.3.1 Study Site Selection

The national hydrometric program managed by Water Survey of Canada is the primary source of standardized water resources data in Canada. The network has 140 historic streamflow records in Nova Scotia, beginning in 1915. Forty-seven catchments, ranging in size from 8 to 1350 km^2, were selected for analysis from the 140 available records, with consideration given to record length, regulation, and stationarity. Originally, the search was limited to catchments with a minimum record length of 15 years; however, because this yielded only 36 stations, the record length criterion was reduced to 8 years. Regulated catchments were not included, as major abstractions or water control structures could not be reliably accounted for. Catchments designated "regulated" by the Water Survey of Canada were examined to determine the likely degree of regulation and retained if it was found to be limited. Hydrologic stationarity was confirmed with a Mann-Kendall trend analysis. Mann-Kendall is a non-parametric test which detects monotonic (upward or downward) trends in the dataset (Mann, 1945; Kendall, 1975). The modified Mann-Kendall (Hamed and Rao, 1998) was applied to three streamflow signatures characterizing key components of the annual hydrograph, RBI, MAF, and JAS (see Table 3.1), computed on an annual basis at the 0.05 significance level. The test was implemented using the *Kendall* package (McLeod, 2011) within R statistical software (R Core Team, 2013).

3.3.2 Catchment Characterization

Function

Streamflow signatures (also known as hydrologic indices or streamflow metrics) refer to statistics derived from long-term hydrometric records. Average daily and annual instantaneous peak flows were extracted from the Hydat database (ECCC, 2019) using the tidyhydat package within the R software (Albers et al., 2017). The following suite of streamflow signatures was selected to characterize the hydrology of each of the 47 gauged catchments (Table 3.1).

Table 3.1 The suite of streamflow signatures calculated for each of the 47 gauged catchments

Signature	Description
Q10	Flow exceeded 10% of the time
Q50	Flow exceeded 50% of the time (median flow)
Q90	Flow exceeded 90% of the time
BFI	Baseflow index
RBI	Richard Baker Flashiness Index (Baker et al., 2004)
FH	Frequency of high flow events
FL	Frequency of low flow events
DH	Duration of high flow events
DL	Duration of low flow events
T50	Flow centroid
MAF	Mean annual flow
JAS	Mean summer flow (July, August, and September)
AMJ	Mean spring flow (April, May, and June)
EWF	Early winter flow
SQ10lp3	7-day low flow occurring once in ten years (Log Pearson III distribution)

Signature	Description
FLgum	20-year flood flow (Gumbel distribution)
FLlp3	20-year flood flow (Log Pearson III distribution)

Form and Forcing

Watersheds were delineated for each study hydrometric station in ArcMap 10.5 (ESRI, 2016) using Arc Hydro Tools. The 20-m resolution Nova Scotia Enhanced Digital Elevation Model (Nova Scotia Department of Natural Resources, Mineral Resource Branch, 2003) and a centreline version of the Nova Scotia Hydrographic Network obtained from Nova Scotia Environment were used as inputs (Williams, 2017). These catchment geometries were used as masks to extract catchment descriptors from spatial datasets using the *raster* (Hijmans, 2019) and *sf* packages (Pebesma, 2018) for R statistical software (R Core Team, 2013).

The selection of catchment descriptors was mainly limited by the availability and reliability of spatial datasets. Often, more suitable metrics exist, but they could not be calculated for the study region. The metrics selected to describe catchment form and forcing are given in Table 3.2.

.

Table 3.2. Physical and climatic catchment descriptors used to represent catchment topography, bedrock and surficial geology, surface storage, soils, landcover, and meteorological forcing

Attribute	**Description**	**Data Source**
Mean stream order	Mean Strahler stream order in catchment	Stream Orders – 1:10,000 (Nova Scotia Environment, 2017)
Max stream order	Maximum Strahler stream order in catchment	
Stream length	Total stream length in catchment	
Drainage density	Ratio stream length to catchment area	
Mean elevation	Mean elevation in catchment	Enhanced 20-m Digital Elevation Model (Nova Scotia Geomatics Centre, 2006)
10th percentile elevation	Elevation below which only 10% of the catchment sits	
90th percentile elevation	Elevation above which only 10% of the catchment sits	
Relief	Difference between maximum and minimum elevation	
Relief Ratio	$(Elev_{Median} - Elev_{Min})/(Elev_{Max} - Elev_{Min})$	
Slope	Average slope of catchment (%)	
Hypsometric Integral	Area under the curve of cumulative height vs cumulative area	
Compactness	Ratio of catchment area to perimeter squared	Watershed polygon (derived)

Attribute	Description	Data Source
Ruggedness	Product of drainage density and relief	Stream Orders – 1:10,000 (Nova Scotia Environment, 2017) Enhanced 20-m Digital Elevation Model (Geomatics Centre, 2006)
Elongation Ratio	Ratio of the diameter of a circle with the same area as that of the catchment to catchment length	Stream Orders – 1:10,000 (Nova Scotia Environment, 2017)
Hydraulic conductivity	Area-weighted median hydraulic conductivity by bedrock groundwater region (from pumping test database)	Groundwater Regions of Nova Scotia – 1:500,000 (Kennedy and Drage, 2008)
Prop. sedimentary	Proportion of catchment underlain by sedimentary bedrock	
Prop. metamorphic	Proportion of catchment underlain by metamorphic bedrock	
Prop. drumlin	Proportion of catchment covered by drumlins	Surficial Groundwater Regions of Nova Scotia – 1:500,000 (Kennedy, 2013)
Prop. glaciofluvial sediments	Proportion of catchment covered by glaciofluvial sediments	

Attribute	**Description**	**Data Source**
Soil drainage	Area-weighted drainage	Detailed Soil Survey Compilations – 1:75,000 (Canadian Soil Information Service, 2010)
Mean annual precipitation	Mean annual precipitation over catchment (mm)	North America climate grid 1971 – 2000 (McKenney et al., n.d.)
Mean temperature coldest quarter	Mean temperature during the coldest quarter (°C)	
Mean precipitation seasonality	Coefficient of variation	
% Wetland	Percentage of catchment that is wetland	Wetland shapefile DNR
% Forested	Percentage of catchment that is forested	Canada's land cover – 30 m (Latifovic, Pouliot, and Olthof, 2017)
% Water	Percentage of catchment that is water	

Attribute	Description	Data Source
% Urban	Percentage of catchment that is urban/developed	
% Cropland	Percentage of catchment that is cropland	
% Barren	Percentage of catchment that is barren	

3.3.3 Inductive Classification

To reduce collinearity, a principal components analysis was performed on scaled and centred streamflow signatures. Principal components cumulatively representing 85% total variance were retained for analysis. To distinguish the different hydrologic regimes represented by the study sites, a clustering algorithm was applied to the extracted principal components. Cluster analysis is used to partition datasets into clusters or subgroups, where dissimilarity within groups is minimized and dissimilarity between groups is maximized (Johnson and Wichern, 2007). Cluster analysis can be limited, as some algorithms will produce clusters even given random data (Lawson and Jurs, 1990). Before performing the cluster analysis, the clustering tendency of the data was assessed using the Hopkins statistics (H value). The test was implemented using the *hopkins* function in the *clustertend* package for R statistical software (YiLan and RuTong, 2015). If the H value was less than a threshold of 0.5, the data was considered clusterable.

A dissimilarity matrix was computed from the extracted principal components using the Euclidean distance measure. The PAM clustering algorithm was chosen as it is less sensitive to noise and outliers than k-means (Kaufman and Rosseauw, 1990). Outlier insensitivity is of particular value for hydrologic classification given that outliers are common. The optimal number of clusters was determined by referencing the body of clustering solutions and the physical interpretability of the clusters (Sanborn and Bledsoe, 2006).

3.3.4 Deductive Classification

A wide variety of physical catchment descriptors and climate metrics were computed in an attempt to fully explain the quantity, movement, and storage of water within each catchment (Table 3.2). The intent was to determine which physical and climatic controls can separate catchments by the previously determined hydrologic clusters. A decision tree approach was selected to classify catchments as it is a practical tool that could be easily implemented by practitioners.

Classification and regression tree (CART) analysis is a supervised machine learning technique used to predict continuous and categorical response variables using binary recursive partitioning to create splits based on explanatory variables (Bell, 1999). Decision trees are especially useful for dealing with nonlinear relationships and hierarchical structures, which are common in fields like ecology (De'ath and Fabricius, 2000) and hydrology (Jencso and McGlynn, 2011; Sivakumar and Singh, 2012). This method produces an interpretable graphical output that can be used effectively by non-experts. Olden et al. (2008) review some significant disadvantages of CART, the most glaring of which with respect to catchment classification is that predictions are not associated with probability level or confidence interval. This makes it difficult to estimate the uncertainty of the predicted class of new data. Additionally, correlated explanatory variables can obscure interactions between variables. The final limitation, especially relevant to catchment classification, is that the end model is not necessarily the ideal tree. At each split, the best partition at that point in the tree is identified, regardless of the underlying hierarchy of processes.

Keeping these limitations in mind, a CART analysis was initially performed on the catchment descriptors. The hydrologic classes were treated as the response variable to explore influential splits for deductive classification. The results of this exploration were used to manually construct a logical and holistic decision tree. The decision tree was constructed using a random sample (70%) of the original dataset. To validate the decision tree, the catchment descriptors of the retained catchments (30%) were used to assign class membership according to the rules of the decision tree. These predicted classes were compared against their actual hydrologic classification to test the validity of the decision tree.

3.4 RESULTS & DISCUSSION

3.4.1 Inductive Classification

The principal components analysis produced a subset of uncorrelated variables to describe the variation in flow regimes across the study catchments (Figure 3.3). The original 17 explanatory variables were reduced to three components cumulatively describing 86% of the variation of the original dataset. The first principal component (PC1) described 45.2% of the variance in the original dataset and corresponded highly with the FL20, Q10, and average seasonal/annual flows. The second principal component (PC2) described 31.2% of the variance in the original dataset and corresponded with Q50, BFI, and RBI. The third principal component (PC3) described 9.6% of the variance in the original dataset and corresponded with SQ10lp3. Complete variable loadings on the first three components are provided in the appendix (Table A.2).

Ultimately, nine hydrologic clusters were identified using a PAM clustering algorithm (C1-C9; Figure 3.3). Streamflow signatures representing flood flows, high flows, and average seasonal/annual flows were loaded heavily on to PC1, positive values of which appeared to coincide with increasing precipitation. C8 and C9 are located in and around the Cape Breton highlands, while C1 and C3 are located in western Nova Scotia. Variables representing median flow, baseflow, and flashiness metrics were loaded heavily on to PC2: C4 and C8 had high baseflow and more positive PC2 scores, whereas C2, C5, and C6 were relatively flashy and had more negative PC2 scores (Figure 3.3).

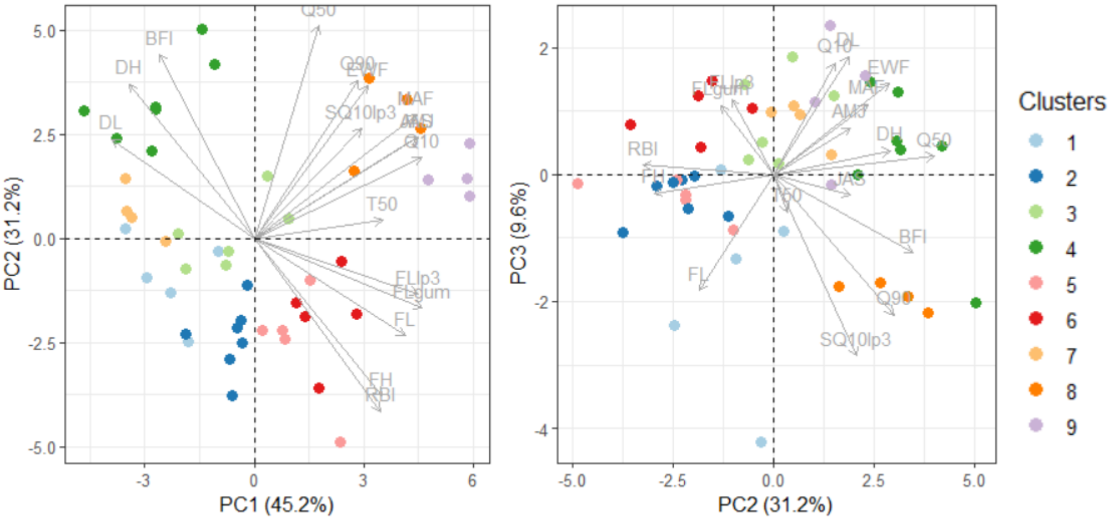

Figure 3.3. Biplots showing streamflow signature loadings and study catchment scores coloured by hydrologic cluster. A) Ordination plot of PC2 and PC2, B) Ordination plot of PC2 and PC3.

C9 catchments had highly positive PC1 scores, indicating high flood flows and annual/seasonal flows. The AMJ signature revealed that this is likely due to larger snowpack resulting in a high-magnitude spring freshet.

Variables representative of low flows tended to be loaded on PC3. Interestingly, C1 and C2 correspond with relatively high magnitude low flows (SQ10lp3), while C9, which receives similar precipitation inputs as C8 and is geographically proximate, experiences very low magnitude low flows.

Catchment clusters did not have the same ranges of streamflow signature values. For example, PC1 scores for C2 catchments varied little, suggesting that these catchments have similar flood flow and average flow characteristics, whereas there is more variability along PC2. The FL20 values for C2 catchments range from 0.72 to 1.02 $\text{m}^3\text{s}^{-1}\cdot\text{km}^{-2}$ compared to the larger spread of the whole dataset of 0.19–3.07 $\text{m}^3\text{s}^{-1}\cdot\text{km}^{-2}$. Spring flow (AMJ) varies from 0.032 – 0.045 $\text{m}^3\text{s}^{-1}\cdot\text{km}^{-2}$ (dataset spread: 0.025–0.10 $\text{m}^3\text{s}^{-1}\cdot\text{km}^{-2}$). C4 is much less compact on PC1, as flood flow values range from 0.19 – 0.40 $\text{m}^3\text{s}^{-1}\cdot\text{km}^{-2}$, but AMJ varies from 0.033 – 0.53 $\text{m}^3\text{s}^{-1}\cdot\text{km}^{-2}$. A table summarizing the intraclass streamflow signature variability is available in the appendix (Table A.2).

Figure 3.4 illustrates that although flow classes are mostly regionally clustered, they are not entirely contiguous (e.g., see cluster 4). This highlights that the common practice of using the nearest donor catchment for transferring hydrologic data may not always be appropriate.

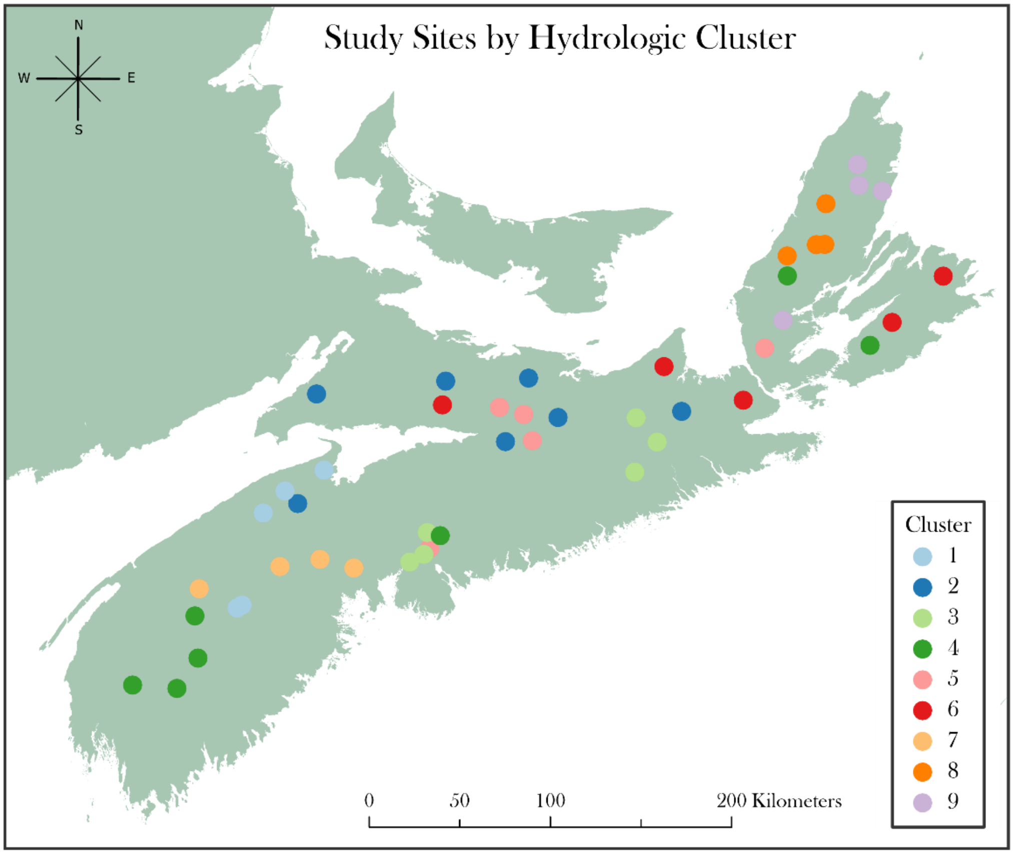

Figure 3.4. The 47 study catchments within the nine hydrologic classes extracted using a PAM clustering algorithm

Meta hydrologic classes

The nine hydrologic classes identified in the PAM cluster analysis were consolidated into five generalized meta classes: lake influenced baseflow, mainland transition, NS flashy, highland flashy, highland baseflow (Figure 3.5).

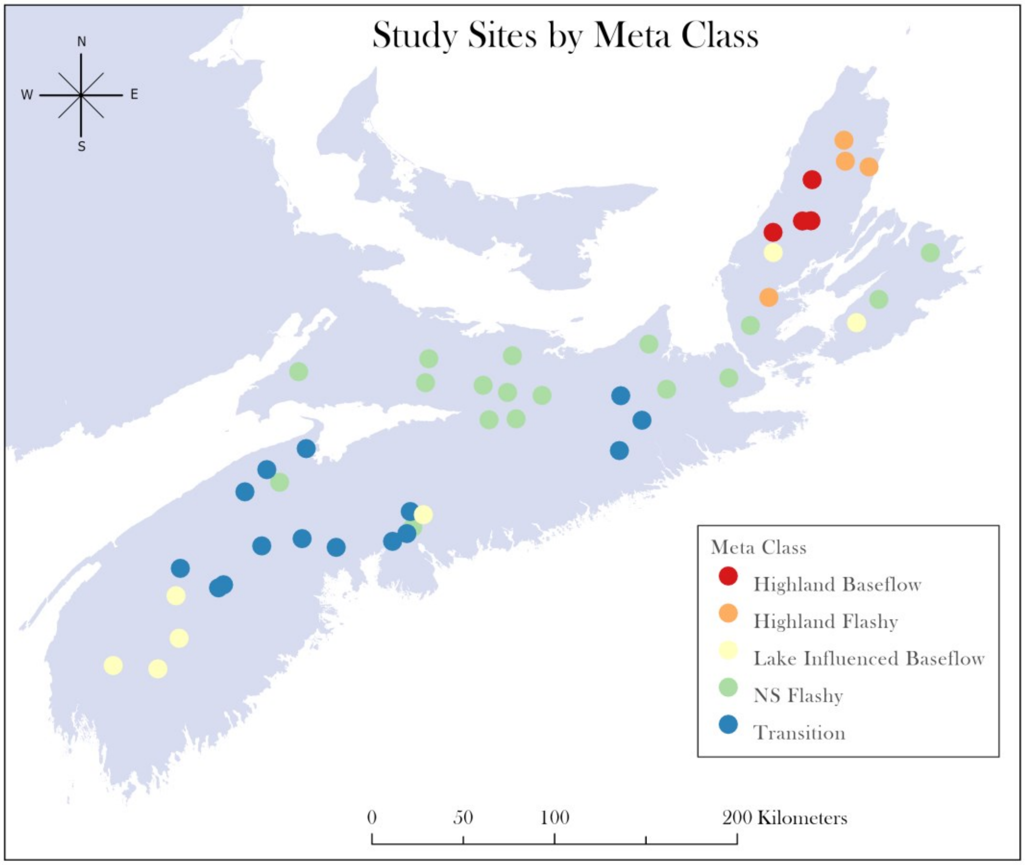

Figure 3.5. Study sites grouped into five generalized meta flow classes.

After generalization, the meta flow classes were more spatially contiguous than the original hydrologic clusters. The decision to generalize the original nine hydrologic clusters was due to difficulty in differentiating similarly acting clusters with varying water yields by their catchment descriptors. For example, C2, C5, and C6, which together form the NS Flashy class, are mainly underlain by sedimentary bedrock but have no real consistency in terms of size or precipitation. Most of the disparity in water yield could likely be attributed to precipitation differences across the catchments; however, the resolution and reliability of the available gridded precipitation data did not capture this. The precipitation grid used in this study does not incorporate maritime influence as part of its algorithm, so coastal precipitation effects could be dampened. The mainland transition class (C1, C3, and C7) was formed in the same manner. Variability exists between clusters that were assigned to the same meta flow class, but this variability either does not have an interpretable link to the available catchment descriptors, or not enough variability exists. Figure 3.6 illustrates the variation seen in a few key streamflow signatures between the new generalized meta classes, demonstrating adequate separation between meta classes for these six signatures.

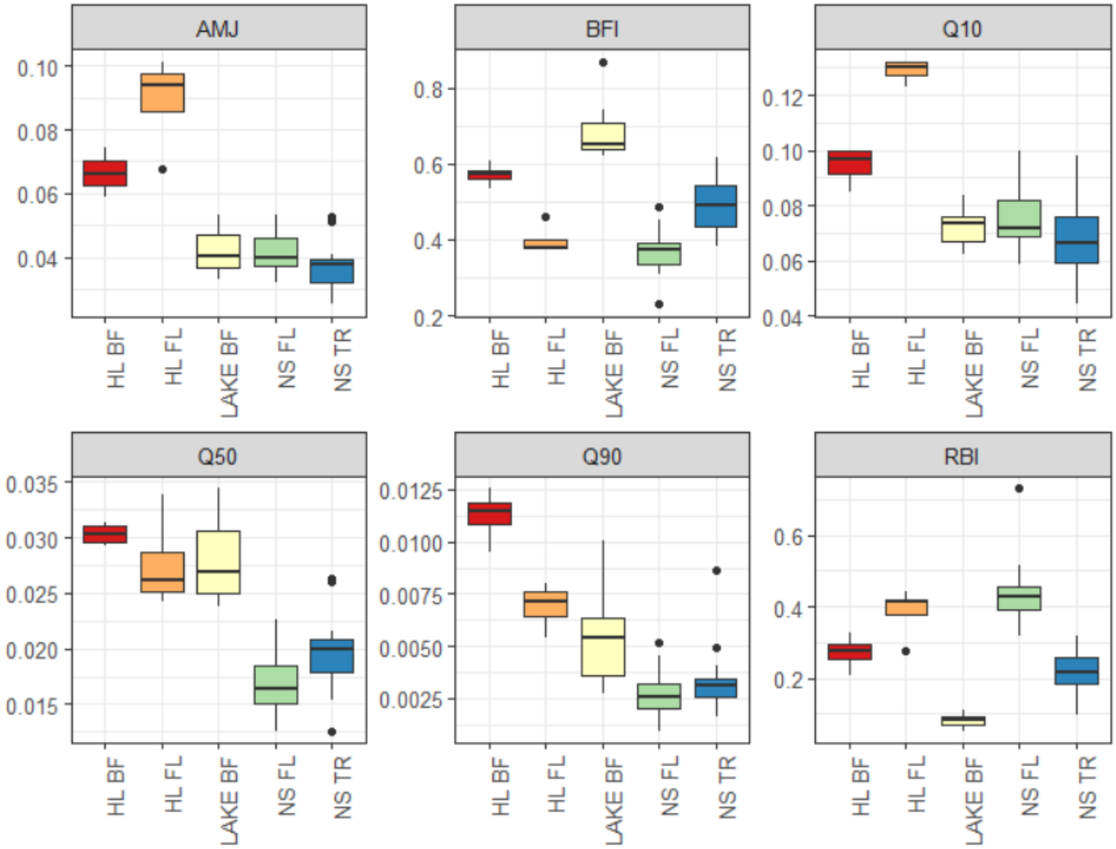

Figure 3.6. Boxplots showing range of streamflow signature values for each meta class, which reveals adequate separation between classes.

3.4.2 Deductive Classification

The catchment descriptors presented in Table 3.2 were used to partition catchments into their respective meta classes identified through the inductive classification. The final decision tree is provided in Figure 3.7. An initial CART analysis was used to identify some relevant thresholds in each catchment descriptor; however, theory-driven reasoning was then used to ensure the tree divisions were logical.

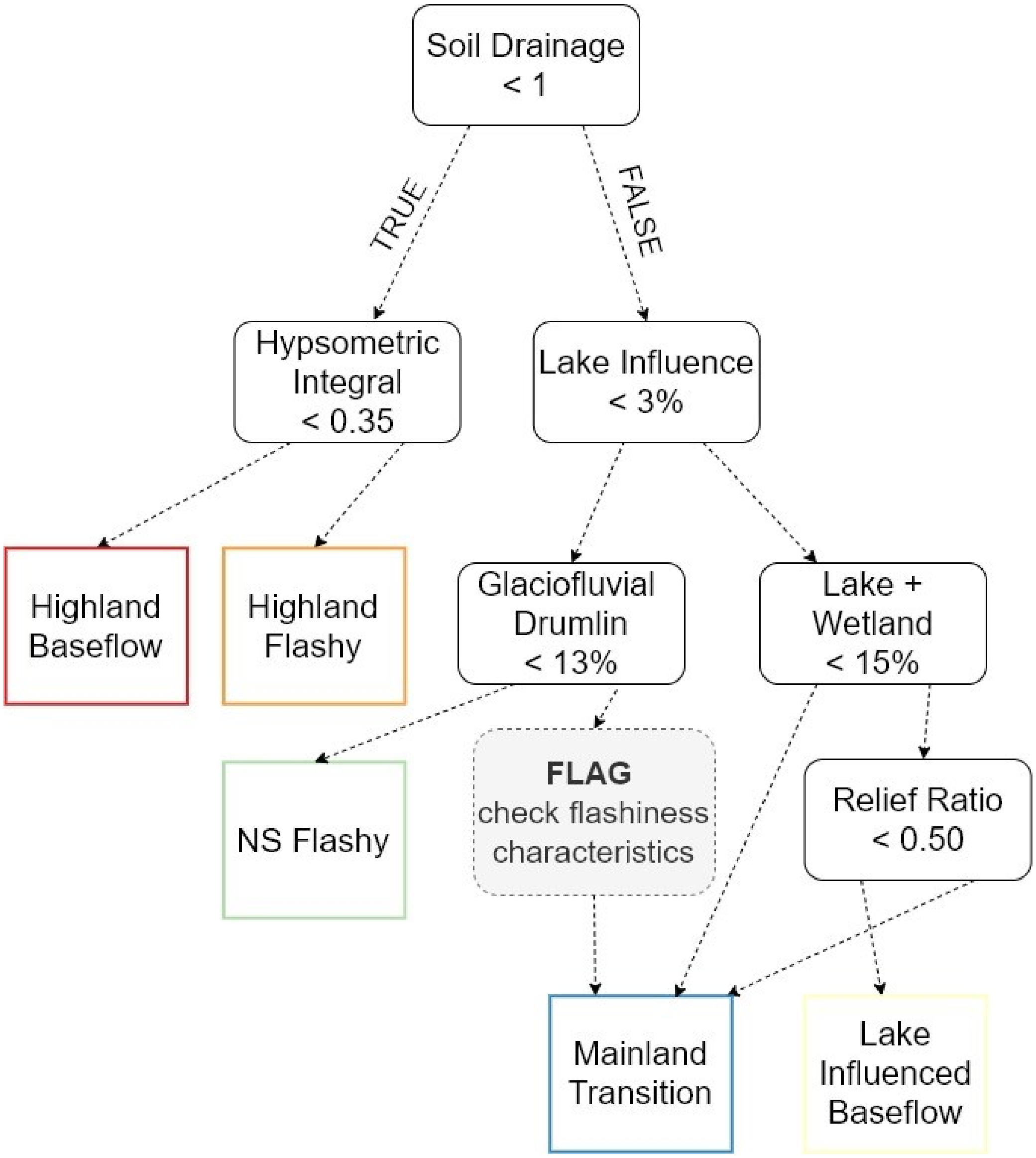

Figure 3.7. Decision tree for deductive classification into meta classes, where each split in the tree is a Boolean statement and moving to the left means the statement is true, while moving to the right means the statement is false.

When constructing the tree, an initial examination of the meta classes revealed a few important explanatory variables, such as lake-influenced baseflow. Large lakes act as storage, attenuating flows and providing a source of water during dry periods. Hannaford et al. (2013) incorporated a pre-existing lake influence index in their research, the Flood

Attenuation due to Reservoirs and Lakes (FARL) index (Bayliss, 1999). Todd et al. (2017) outlined large lake and glacier influence as a major factor impacting hydrology. Snelder and Biggs (2002) also classified lake influence as a "source of flow" in their top-down approach and placed it near the top of their hierarchy, only below climate forcing. The relationship between lake coverage as a fraction of the total catchment area and PC2, which corresponds with flashiness and baseflow, is illustrated in Figure 3.8. As lake coverage increased, catchments were less flashy and had higher baseflow indices. However, this relationship did not appear to hold for highland catchments, which had higher precipitation inputs.

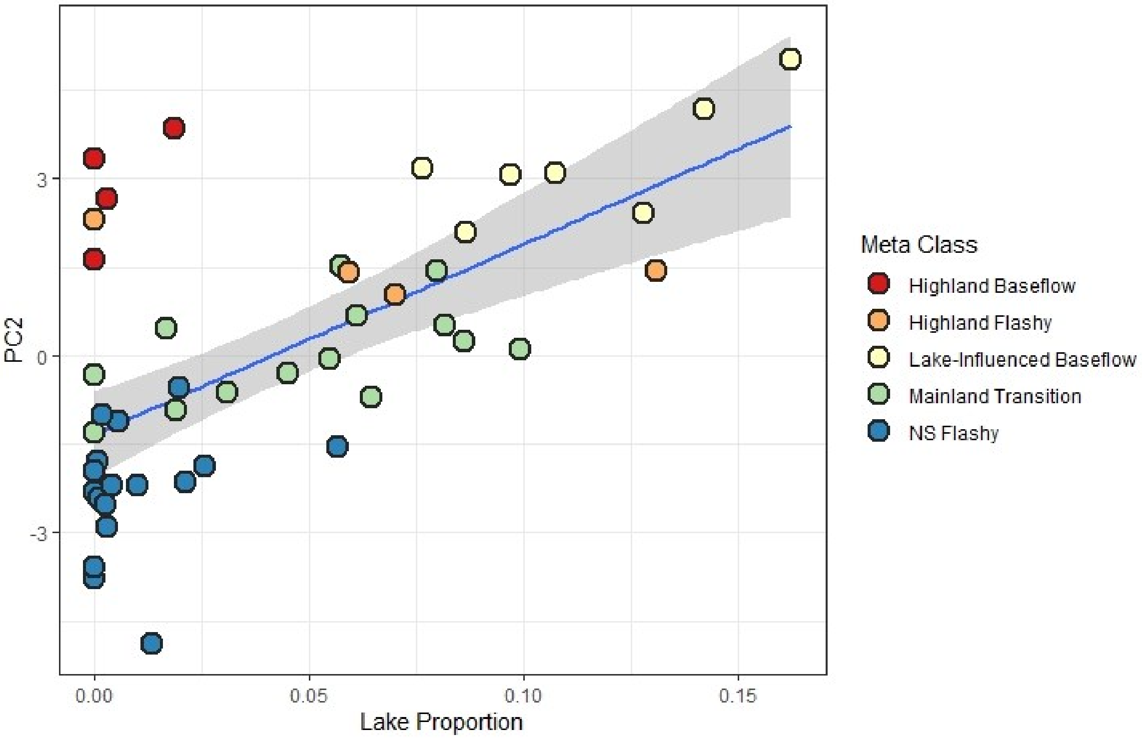

Figure 3.8. Plot showing PC2 vs. the lake influence as the fraction of the catchment covered by large lakes. BFI and RBI are primarily loaded on PC2. The linear fit (in blue) and 95% confidence interval (in grey) highlights the approximately linear relationship between lake coverage and flashiness.

Soil drainage was also noted as an influential variable and was used to distinguish Cape Breton highland catchments from the rest of the province. The distribution of the soil drainage index across the study catchments was bimodal (Figure 3.9), for which the majority of the highland catchments had values between 0 and 1. The soil drainage index is a numeric approximation of the qualitative soil drainage descriptions in the National Soil Database (Agriculture and Agri-Food Canada, 2010), from very rapidly drained (8) to not applicable (e.g. rock or ice; 8).

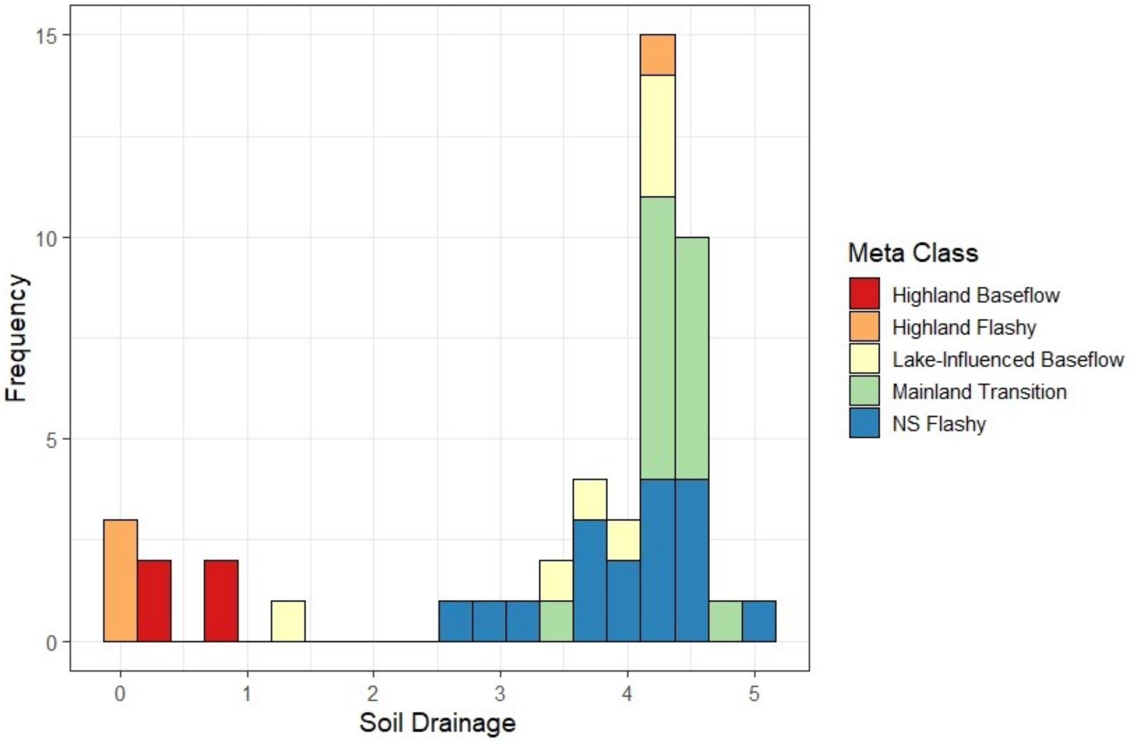

Figure 3.9. Plot showing the frequency of soil drainage values coloured by meta class. Most highland catchments had soil drainage index values <1, making it an effective break in the decision tree.

The physical characteristics of highland catchments were further examined to differentiate these meta classes. Qualitative examination revealed that highland baseflow catchments appeared to have deep canyons. The presence of these canyons was not well characterized by average slope or relief, but was characterized by the hypsometric integral. An example of this is the Northeast Margaree River (Water Survey of Canada station number 01FB001), which follows the Aspy Fault line (Figure 3.1A).

Outside of the Cape Breton highlands there are three generalized meta classes: flashy, transition, and lake influenced baseflow. As previously illustrated, lake influence is a relevant predictor of hydrologic function in Nova Scotia. Flashy catchments generally have very low lake storage. Similarly, some transition catchments also have little surface water, but have surficial geology such as glaciofluvial sediments and drumlins that promote infiltration and groundwater flow and storage. After partitioning by lake storage (<3%), a threshold of 13% glaciofluvial sediments or drumlin was used as a separator. Catchments with less than 13% glaciofluvial sediments or drumlin were assumed to be flashy, while if they are above this threshold, they should be checked for other flashiness characteristics (i.e., relief ratio, elongation ratio, size, and soil drainage) and classified as mainland transition or NS flashy.

If catchments have > 3% lake surface area, they are most likely transition or lake-influenced baseflow. They were further differentiated by their combined wetland and lake influence, and relief ratio. Catchments with more than 15% surface storage (combined lake and wetland area) and a low relief ratio (< 0.50) were assumed to be lake influenced baseflow. Catchments with less than 15% surface storage were assumed to be mainland transition.

Climate was considered the controlling factor in several other studies (Yadav et al., 2007; Liermann et al., 2012). Devito et al. (2005) presented a hierarchy of factors from which to form a conceptual model of dominant hydrologic controls in a particular catchment and the scale at which the interaction happens. The hierarchy they proposed was climate, bedrock geology, surficial geology, soil type and depth, and topography and drainage network. In this study, climate (annual precipitation and mean temperature) did not appear to be influential in separating meta classes and was therefore excluded from the decision tree. This may indicate that there is not enough climate variability in the region for it to be the controlling factor. However, we recommend that climate should be reserved as a check variable when selecting an appropriate gauged donor catchment. That is, ungauged catchments would first be classified using the decision tree, then a gauged donor catchment from the same meta class with similar precipitation inputs would be selected.

The decision tree was able to correctly classify 27 of 33 catchments from the training set, while correctly flagging five additional catchments to check their flashiness characteristics before classifying them as NS flashy or mainland transition. One catchment (01FG001 – River Denys at Big Marsh) was incorrectly flagged. This catchment was classified hydrologically as highland flashy but is not located in the highlands. The major differences between this catchment and others in its meta class are flow timing and early winter flows. River Denys has a flow centroid much earlier in the year than other rivers, and the magnitude of its early winter flow is roughly one third larger than the rest of the hydrologic cluster. River Denys has a unique shape (resembling a bowtie, with a constriction mid catchment), higher soil drainage than its counterparts, and receives less precipitation.

For the test set, the decision tree classified nine of 14 catchments correctly and flagged two of 14 to be checked for flashiness characteristics. Three of the 14 catchments were incorrectly classified, giving a misclassification rate of 21%. Clam Harbour River near Birchtown (01ER001) was misclassified as mainland transition rather than NS flashy, due to its relatively high lake storage component. However, it has low soil drainage, a high relief ratio, and a low hypsometric integral. Wreck Cove Brook near Wreck Cove (01FD001) was misclassified as highland baseflow rather than highland flashy due to a very low hypsometric integral. However, this catchment receives high precipitation inputs, which presumably masks the baseflow contributions. East River at St. Marys (01EO003) was misclassified as NS flashy rather than mainland transition due to its low lake storage and a low percentage of glaciofluvial sediments and drumlins.

The misclassified catchments highlight a key limitation with this type of hierarchical deductive classification: many different combinations of catchment characteristics can produce similar hydrologic function. Other authors have discussed this issue, referring to *process equifinality* in the runoff signal (Hellebrand et al., 2011; Ley et al., 2011).

Table 3.3 summarizes the original hydrologic clusters, their meta classes and pertinent landscape and hydrologic features, as well as the relative magnitude of water yield within each subtype.

Table 3.3. Summary table describing the key features of the hydrologic clusters

Meta Class	Cluster	Landscape Features	Hydrologic Features	Water Yield
Mainland transition	1	- Agricultural landuse - High soil drainage - Glaciofluvial sediments and drumlins	- Moderate baseflow	Low
	3	- Metamorphic bedrock - High precipitation - Elongated - Glaciofluvial sediments and drumlins		High
	7	- Moderate precipitation - Drumlins - Lake and wetland storage	- Prolonged periods of low flow	Medium
Lake influenced baseflow	4	- Lake and wetland storage	- High baseflow	Medium
NS flashy	2	- Sedimentary bedrock - Glaciofluvial sediments - Low precipitation	- Highly flashy	Low
	5	- Sedimentary bedrock - Moderate precipitation	- Highly flashy	Moderate

Meta Class	**Cluster**	**Landscape Features**	**Hydrologic Features**	**Water Yield**
	6	- Sedimentary bedrock - Moderate precipitation	- Highly flashy - High flood flows	High
Highland baseflow	8	- Canyonland (Baechler, 2015) - Limited soil coverage (exposed bedrock)	- High runoff from snowmelt	High
Highland flashy	9	- Limited soil coverage (exposed bedrock)	- High runoff from snowmelt - High flood flows - Late flow centroid	High

A more complete catchment descriptor dataset may be required for improved deductive classification, as discussed by Toth (2013). The content of the available datasets was not always the most relevant to hydrologic behavior. Incorporating factors like vegetation and more comprehensive climatic indices (e.g. evapotranspiration, snowpack) may improve discriminatory power. As previously discussed, climate did not emerge as a controlling factor, although it coincides with the separation of highland catchments from the rest of the province. The influence of increased precipitation was not integrated at the inductive classification stage. It is not clear if increased magnitude of flows should be attributed to increased precipitation, or catchment form. Further research would be needed to clarify this question.

The hydrologic classification scheme and resulting decision tree produced in this study should aid in understanding and managing Nova Scotia's surface water resources, and similar approaches could be used to develop a classification scheme elsewhere. However, this analysis could be expanded. In this study, we sought to categorize annual hydrologic regimes, but there is value in investigating how catchment classification changes when considering specific hydrologic conditions, such as low flow periods. Patil and Stieglitz (2011) examined the influence of flow condition on catchment similarity. They saw a seasonal influence on the factors that govern catchment behavior; at low flow conditions, local controls are dominant (e.g. soil moisture or micro-topography), while at higher flows, similarity is governed by climatic controls.

3.5 CONCLUSIONS & RECOMMENDATIONS

Catchment classification can provide a framework for identifying the variability in hydrologic behavior that exists at a regional or global scale. In this study, 47 small to medium sized (8 km^2 – 1300 km^2) watersheds in the province of Nova Scotia, Canada were characterized and classified using direct measures of hydrology (streamflow signatures), and environmental surrogates (form and forcing catchment descriptors). The hydrologic clusters and meta classes identified in this study are likely a subset of the potential classes that exist in the province, but this initial classification serves as a useful starting point for identifying the extent of variation in hydrologic regimes across the province. The decision tree that has been developed could be used as a tool for data transfer decisions or for a review of the monitoring network.

The central weakness of this classification is the small dataset. Since a large number of recent long-term records are not available in the study region, alternative methods could be investigated to expand the dataset and try to further connect specific catchment features to hydrologic function. Examples could include isotope tracers to determine transit time (Speed et al., 2010), distribution of water temperature to reveal hydrological and hydrogeological processes and controls (Briggs et al., 2018; Johnson et al., 2020), and DNA sequencing to characterize each stream's microbiome (Good et al., 2018). Additionally, a more complete catchment descriptor dataset may be required for better deductive classification. Including factors like vegetation and more comprehensive climatic indices (e.g., evapotranspiration) may improve discriminatory power. Despite the study limitations, the results demonstrate that this regional catchment classification system is useful for organizing the hydrological variability in regions with limited capacity to monitor their water resources. Other related studies have generally been performed for regions with more extensive spatial data and/or longer flow records. Our general approach is transferable for other regions with data challenges and is useful for uncovering regional controls on catchment function.

We are presently in an era of rapid environmental change, and there is increased awareness of the need to have representative hydrologic monitoring networks. The catchment classification approach developed in this study is useful for assessing gaps or redundancy in a hydrologic monitoring network and helping to inform the development of a cost-effective monitoring network that can be used to detect and quantify hydrologic regimes and their sensitivity to environmental change.

Chapter 4 Hydrologic assessment of a hydrometric monitoring network in a small maritime region

ABSTRACT

The National Hydrometric Program, operated by the Water Survey of Canada, is the primary source of surface water quantity data for the majority of the provinces and territories in Canada. The network is cost-shared between the federal and provincial/territorial governments, and decisions relating to station ownership and placement are made according to both federal and provincial/territorial interests. Nova Scotia is a small maritime province in Atlantic Canada that has a diverse climate and geology. The Nova Scotia hydrometric monitoring network currently consists of 31 stations in an area of 55,284 km^2, for a station density of one every 1,783 km^2. The province is comprised of 246 major watersheds, termed secondary watersheds, that are the principal spatial unit for surface water management in the province. The overall objective of this study was to determine how well the current hydrometric monitoring network captures the level of hydrologic variability expected in the province. This was accomplished through application of a regional catchment classification scheme, and a systematic analysis of the key hydrologic characteristics of both gauged and ungauged watersheds. According to guidelines for station density provided by the World Meteorological Organization (WMO) the current active hydrometric monitoring network would be considered adequate. However, the approach detailed in this study revealed that three of the five major watershed types in the region were not well represented by the current monitoring network. This study illustrated the importance of developing regional catchment classification systems

and provides a straightforward approach that could be applied by other jurisdictions to assess the representativeness of their hydrometric monitoring networks.

4.1 INTRODUCTION

The National Hydrometric Program (NHP), operated by the Water Survey of Canada (WSC), is the primary resource for surface water information in Canada. The WSC is responsible for the collection, interpretation, and dissemination of this data. The network is operated on a cost-sharing basis between the federal and provincial/territorial governments, and decisions on network arrangement are based on both parties' interests. The network also includes a variety of legacy purpose-built stations, e.g. for hydroelectric operation. These stations are often regulated, and thus less valuable for predicting streamflow in ungauged catchments.

Regular review of the monitoring network is essential to ensure it is robust and continuing to meet network objectives. Moving towards an optimized network, that meets targeted objectives, would increase the value arising from the data collected and better support practitioners and scientists. Walker (2002) explored the difficulty associated with establishing the true economic value of hydrometric data, as many who benefit from the data are not necessarily aware of its existence. This can make it particularly difficult for decision-makers to justify the costs associated with operating a network, despite the immense potential benefit a robust network can provide to a variety of stakeholders. Hydrometric data is used for a variety of important functions, including flood forecasting, reducing uncertainty associated with infrastructure design, and preparing for drought that could impact water supplies and agricultural producers. Long-term hydrometric data from undisturbed catchments is also desired for monitoring the effects of climate change on water resources. To this point, the Canadian Council of Ministers of the Environment (2011) produced a reference document to assist all levels of government in assessing their water monitoring network's fitness for preparing for and adapting to climate change. In this report, the authors identified Nova Scotia as a region at risk of increased water scarcity during summer months.

Since it last underwent a formal review in 1985, the province's hydrometric network has decreased significantly (Environment Canada and Nova Scotia Department of Environment, 1985). At that time, the network included 53 stations, and the previous review found little potential for data transfer between stations. Shortly after this review, due to federal budget cuts in the 1990s, nearly one third of the national network was decommissioned. The impact of this to Nova Scotia's network continues to be felt today, where the network has shrunk to 31 active stations (ECCC, 2019). For comparison, Vancouver Island, BC, a coastal region with a population similar to Nova Scotia's roughly 976,000 (Statistics Canada, 2019) and just over half the land area (but greater precipitation variability), has nearly double the number of active hydrometric stations (ECCC, 2019). This discrepancy generates questions with respect to the criteria that should be used to establish required hydrometric density. The World Meteorological Organization (WMO) has provided global guidelines for minimum hydrometric station density by physiographic region (WMO, 2008). However, these guidelines do not provide specifications, or methods, for deciding where these stations should be located within a specific region.

Approaches to hydrometric network review and design generally fit into one of six categories: 1) statistical-based methods, 2) information-theory, 3) user survey, 4) physiographic components, 5) sampling strategies, and 6) hybrid (Mishra and Coulibaly, 2009). In recent years, components of Canada’s National Hydrometric Program have been investigated using several of these approaches. Information theory methods are becoming increasingly popular and have been applied in several Canadian regions (e.g. British Columbia, Yukon, New Brunswick) as a way to rationalize or optimize hydrometric networks (e.g. Mishra and Coulibaly, 2010; Halverson et al., 2015; Boisvert et al., 2017). Using a physiographic component approach, Pellerin (2019) suggested revisions to the Canadian Reference Hydrometric Basin network. The breadth of hydrometric stations monitoring the Canadian pan-Arctic was assessed by Mlynowski et al. (2011) using statistical-based methods, who recommended gauging smaller rivers, as larger water courses had better coverage. Internationally, comprehensive studies have been carried out in jurisdictions with comparatively dense monitoring networks. For example, in the United Kingdom, networks have been evaluated for prediction in ungauged basins using statistical-based methods and sampling strategy approaches, and user surveys have been conducted to assess efficacy (Hannaford et al., 2013; Hynds et al., 2019). All the approaches outlined by Mishra and Coulibaly (2009) have merit, however, in a small jurisdiction like Nova Scotia, an applied approach that can be easily understood by decision-makers may be the most appropriate. For example, Laize (2004) presented a geographical information system-based approach to identify representative catchments and support catchment review, the Representative Catchment Index (RCI). The RCI is a deductive-based similarity index; catchments are compared to a reference area using spatial datasets and are scored according to coinciding spatial characteristics.

We can learn from jurisdictions that have undergone comprehensive network assessment, such as the United Kingdom, which focuses its resources on gauging benchmark catchments, artificial impacts, catchments with value for regionalization (representativeness), and integrated monitoring (Bradford and Marsh, 2003). In this paper, we focus on achieving a network of representative gauged catchments that provide long-term records which fully characterize the flow regimes seen in the province. Nova Scotia has a complex geology and moderate precipitation variability; its topography and coastal setting result in conditions that can be difficult to gauge efficiently. Instead of large drainage basins (allowing single stations to efficiently gauge sizable areas), there are nearly 250 meso-scale “secondary watersheds” draining directly to the ocean, in addition to hundreds of small coastal and island catchments.

The overall objective of this study was to determine how well the current hydrometric monitoring network in Nova Scotia captures the level of hydrologic variability expected in the province. To accomplish this, we applied a combined inductive-deductive catchment classification to the province’s active monitoring network and 246 ungauged major watersheds. Based on the results of this analysis we then identified gaps and potential redundancies in the active monitoring network.

4.2 METHODS

Study Region

The province of Nova Scotia (Figure 4.1) is located on the East Coast of Canada, connected to neighbouring New Brunswick and the rest of the Canadian mainland by a narrow land bridge (the Chignecto Ismuth). At 55,284 km^2, Nova Scotia is the country's second smallest province by land area, but home to a diversity of landscapes unique to a such a small region in Canada. The province is further divided into the mainland and Cape Breton Island.

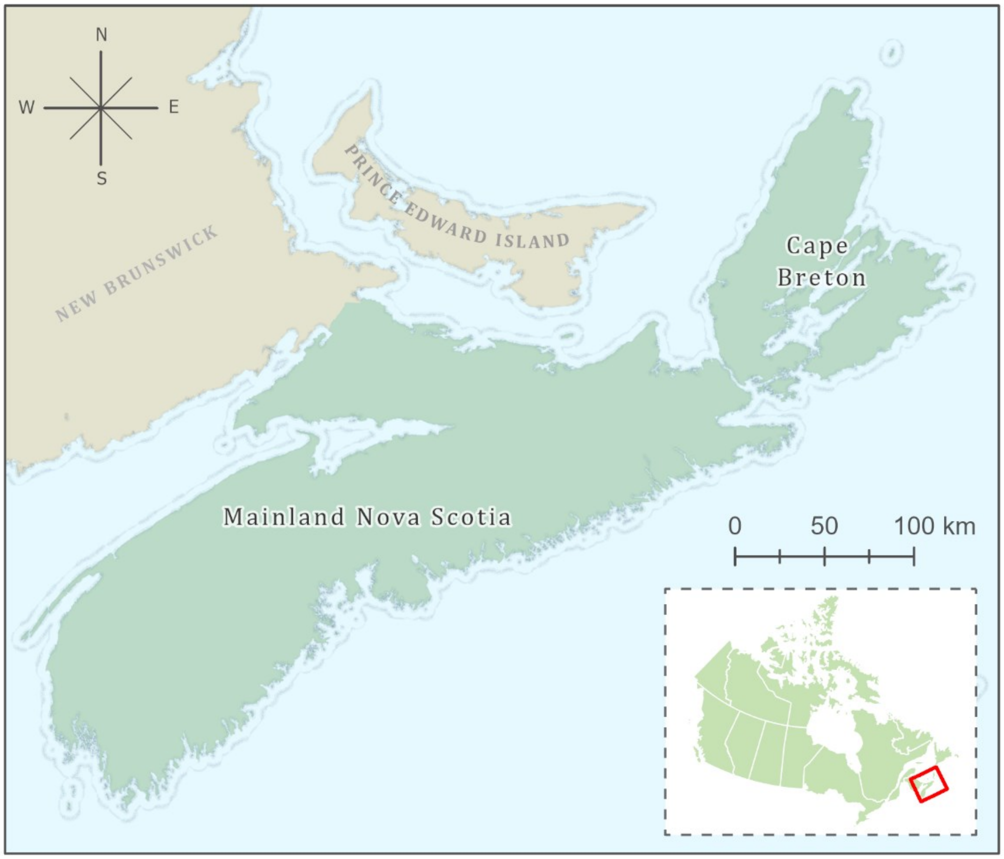

Figure 4.1. Locator map of Nova Scotia within Canada

The current hydrometric network was assessed against the WMO hydrometric density guidelines and for its coverage of known flow regimes. Initially, the minimum number of stations suggested by WMO (2008) was calculated for the province. The recommended minimum densities by physiographic unit are provided in Table 4.1.

Table 4.1. Recommended minimum densities of streamflow monitoring stations (area in km^2 per station) (WMO, 2008)

Physiographic Unit	Streamflow
Coastal	2750
Mountains	1000
Interior plains	1875
Hilly/undulating	1875
Small islands	300
Urban areas	–
Polar/arid	20000

The 31 active hydrometric stations in the province are listed in Table 4.2. Five of 31 stations are designated as regulated by ECCC (Table 4.2), and as such were considered to have poor value for regionalization and were removed from the analysis. In Nova Scotia, the land area is divided into primary, secondary, tertiary, and sub tertiary watersheds. These primary watersheds do not follow the definition of a watershed and as such were not used in this analysis. Secondary watersheds are generally used as the principal water management unit in Nova Scotia and refer to the major watersheds draining the hinterland of the province and terminating at a single point where the main stem meets the ocean. These watershed polygons are available for download from Nova Scotia Environment

(NSE) (NSE, 2010). These 246 secondary watersheds account for 79 percent (43,800 km^2) of the total land area.

A deductive-based decision tree developed in the Chapter 3 was applied to the current hydrometric network and the 246 secondary watersheds to allow for a straightforward comparison of the two. The decision tree classified ungauged catchments into five hydrologic meta-classes using only physical catchment characteristics. The five meta-classes were (i) highland baseflow, (ii) highland flashy, (iii) Nova Scotia flashy, (iv) lake-influenced baseflow, and (v) mainland transition. Descriptors characterizing the geology, landuse, storage, and meteorological forcing on the catchments were used to separate meta-classes. The deductive decision tree was manually constructed using soil drainage, large lake influence, hypsometric integral, the fraction of catchment with glaciofluvial sediments and drumlins, lake and wetland influence, and relief ratio as splits in the decision tree. The decision tree had been previously trained and tested using inductive-based classification of long-term gauged watersheds in the province (Johnston et al. 2020).

Table 4.2. List of active hydrometric monitoring stations with station code, contributing drainage area, Reference Hydrometric Basin Network status, and regulation status.

		Area	**Ref.**	**Reg.**
Annapolis River at Wilmot	01DC005	546		✓
Annapolis River at Lawrencetown	01DC007	1020		✓
Cornwallis River at Cambridge Station	01DD002	91		
North Brook at Sheffield Mills	01DD005	16		
Beaverbank River Near Kinsac	01DG003	97	✓	
St Andrews River at Stewiacke	01DG043	98		
McClures Brook at Cobequid Trail	01DH006	41		
Salmon River at Murray	01DH002	363		
North River at North River	01DH004	202		
Great Village River Near Scrabble Hill	01DJ005	89		
Kelley River (Mill Creek) At Eight Mile Ford	01DL001	63	✓	
Middle River of Pictou at Rocklin	01DP004	92	✓	
South River at St. Andrews	01DR001	177		

		Area	Ref.	Reg.
Tusket River at Wilson's Bridge	01EA003	1070		✓
Roseway River at Lower Ohio	01EC001	495	✓	
Mersey River Below George Lake	01ED005	723	✓	
Mersey River Below Mill Falls	01ED007	295	✓	
Shelburne River at Pollard's Falls Bridge	01ED013	268		
Moose Pit Brook at Tupper Lake	01EE005	18		
LaHave River at West Northfield	01EF001	1250	✓	
Sackville River at Bedford	01EJ001	146		
Little Sackville River at Middle Sackville	01EJ004	13		
St. Marys River at Stillwater	01EO001	1350	✓	
River Inhabitants at Glenora	01FA001	193	✓	
Northeast Margaree River at Margaree Valley	01FB001	368	✓	
Southwest Margaree River near Upper Margaree	01FB003	357	✓	
Cheticamp River Above Robert Brook	01FC002	190		✓
Indian Brook at Indian Brook	01FE002	125		✓
Middle River at MacLennans Cross	01FF001	123		
River Denys at Big Marsh	01FG001	14		
MacAskills Brook Near Birch Grove	01FJ002	17		

The decision tree framework was applied to the remaining 26 stations suitable for data transfer, ranging in size from 13 – 1350 km^2, as well as the 246 secondary watersheds. To visualize and compare the catchment attributes, a principal components analysis was performed in the R software environment (R Core Team, 2020) using the *FactoMineR* package (Le, Josse, and Husson, 2008) on individuals in each meta-class with precipitation, catchment area, and the catchments descriptors used in the deductive classification to define the respective meta-class. Additionally, histograms were constructed to visualize the distribution of annual precipitation amounts and watershed areas for gauged catchments and secondary watersheds.

4.3 RESULTS & DISCUSSION

4.3.1 WMO Station Density Guidelines

The World Meteorological Organization (WMO) (2008) has suggested minimum station densities to avoid “serious deficiencies” with the stipulation that adjustments should be made according to region specific socio-economic and physio-climatic factors. These recommendations differ by physiographic region, of which WMO defines six types: coastal, mountainous, interior plains, hilly/undulating, small islands, and polar/arid. Although it is not entirely clear if all of Nova Scotia should be defined as coastal rather than hilly/undulating, a range of 20 – 30 stations would be required according to these guidelines to avoid “serious deficiencies”. That is, a minimum of 20 stations if the entire province was defined as coastal, or a minimum of 30 station if the entire province was defined as hilly/undulating. The recommendation for coastal areas (one station per 2750 km^2) is considerably less dense than that for hilly/undulating or interior plains (one station per 1875 km^2). Therefore, Nova Scotia’s network of 31 hydrometric stations meets or exceeds WMO’s minimum densities before accounting for socio-economic and physio-climatic factors.

4.3.2 Representation of flow meta-classes in active hydrometric network

The watersheds of the active hydrometric network, classified deductively, are provided in Figure 4.2. Four of the five meta-classes appear in the active network, when classified deductively. This initial examination revealed that the majority of our previously identified generalized flow regimes are captured in the network, but the highland flashy meta-class is not represented.

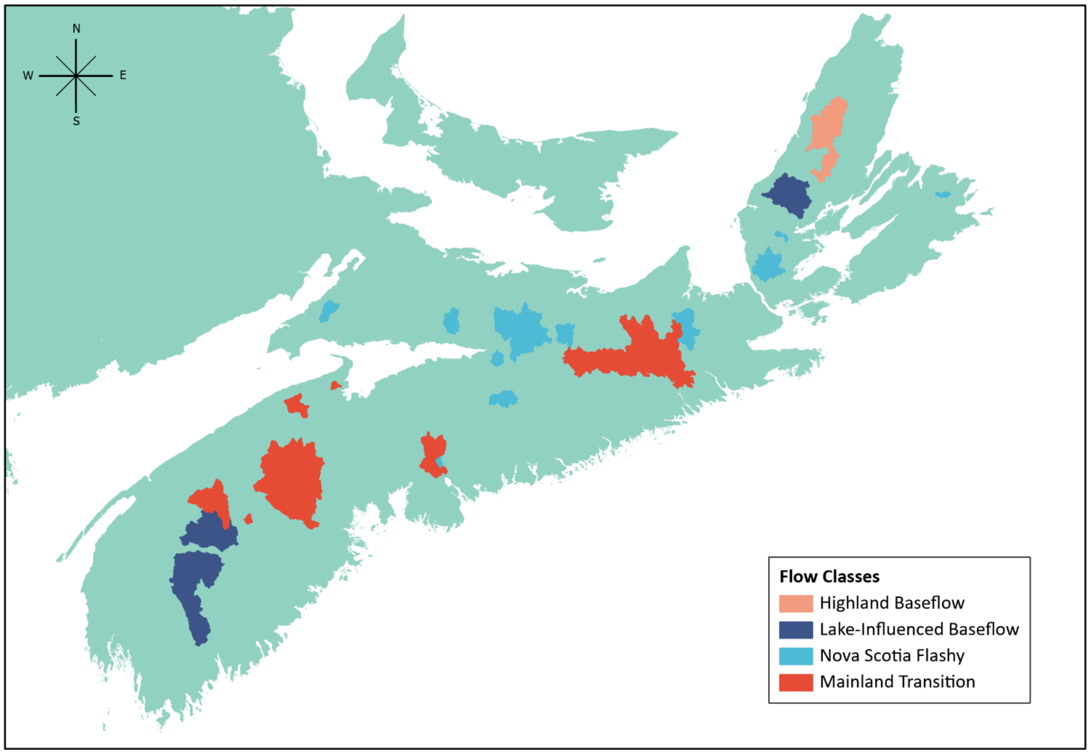

Figure 4.2. Map showing watersheds of active hydrometric stations without significant regulation, classified by flow class determined in the previous chapter.

4.3.3 Characterization of secondary (un-gauged) watersheds

The spatial distribution of meta-classes for all of the ungauged secondary watersheds are illustrated in Figure 4.3. **Nova Scotia flashy** catchments are dominant on the northern mainland and southwestern coast of Cape Breton, while **mainland transition** catchments drain a large portion of the mainland's interior. **Lake-influenced baseflow** catchments are prominent in southwestern Nova Scotia, as well as the southeastern lobe of Cape Breton. The watersheds that make up the Cape Breton highlands are mainly classified as **highland baseflow**, with no **highland flashy** catchments identified at this spatial scale. The absence of these catchments highlights one challenge with assessing how well the hydrometric monitoring network represents hydrologic variability at the spatial scale of assessment. In this study we generated comparisons at the secondary watershed scale; smaller subcatchments may display different types of hydrologic response, and these types of responses may not be apparent at the secondary watershed scale. The highland flashy meta-class was identified as a distinct flow regime which existed in the province using gauging records from WSC stations that are no longer active (Johnston et al. 2020). The lack of representation of this hydrologic regime within both the active hydrometric monitoring network, and the primary spatial unit for surface water management (secondary watersheds), illustrates a potential gap in both monitoring and management of water resources in the province. This could be addressed by gauging nested watersheds to better understand catchment controls, and the effect of spatial scale on hydrologic responses.

A preliminary visual comparison of the active hydrometric monitoring network (Figure 4.2) to the deductively classified secondary watersheds (Figure 4.3) demonstrates that the current monitoring network appears to capture the spatial variability in the other four meta-classes in the province. One notable exception is that the active hydrometric monitoring network does not include any small coastal catchments, which are numerous in the province.

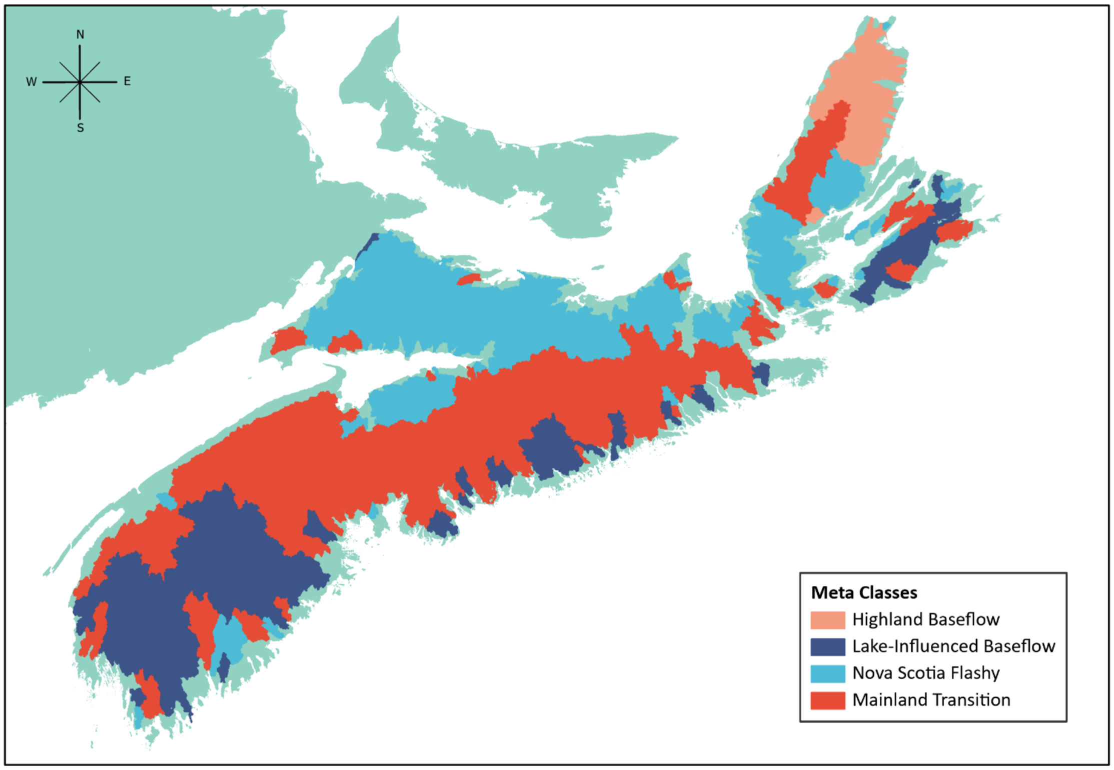

Figure 4.3. Map of deductively classified secondary watersheds.

4.3.4 Comparison of the characteristics of gauged (hydrometric network) vs ungauged (secondary) watersheds

Specific characteristics of the watersheds in the active hydrometric network were compared to those found within the secondary watersheds. The variation in key watershed characteristics were first visualized using PCA biplots (Figure 4.4). A visual inspection of the biplots for each of the four meta-classes reveals some key gaps and potential redundancies in the active hydrometric monitoring network. To start, the two gauged **highland baseflow** catchments have very similar physio-climatic characteristics (soil drainage, hypsometric integral, and annual precipitation) compared to the secondary watersheds classified as highland baseflow (Figure 4.4a). Likewise, for the **mainland transition** meta-class, the six gauged catchments are positioned in three clusters with comparable physio-climatic characteristics (Figure 4.4b). The biplot illustrates that watersheds with higher relief ratio (relratio) and higher surface storage (stor) are not currently represented in the active hydrometric monitoring network, and that a large proportion of secondary watersheds in this meta-class have these characteristics. Within the **lake-influenced baseflow** meta-class (Figure 4.4c), the four gauged catchments have similar water storage components (water), and do not represent secondary watersheds that possess both large lake influences and high wetland storage. The **Nova Scotia flashy** meta-class (Figure 4.4d) is the most comprehensively gauged meta-class, with the active hydrometric monitoring appearing to capture most variability in secondary watershed characteristics. However, several potential redundancies are apparent, as there are several pairs of gauged stations which cluster together and therefore are capturing watersheds with similar physio-climatic characteristics.

Table 4.3. Summary table describing the frequency and areal coverage of each meta-class for gauged and secondary watersheds, where total area refers to the sum of areas for each type (i.e. percent of all gauged area and percent of all secondary watersheds).

Meta-class	Type	Count	Area Covered	Percent of Total Area
Highland Baseflow	Gauged	2	487	7
	Secondary	31	2075	4.7
Lake-influenced Baseflow	Gauged	4	1854	26.7
	Secondary	38	9865	22.5
Nova Scotia Flashy	Gauged	12	1358	19.6
	Secondary	95	12194	27.9
Mainland Transition	Gauged	8	3245	46.7
	Secondary	80	19644	44.9

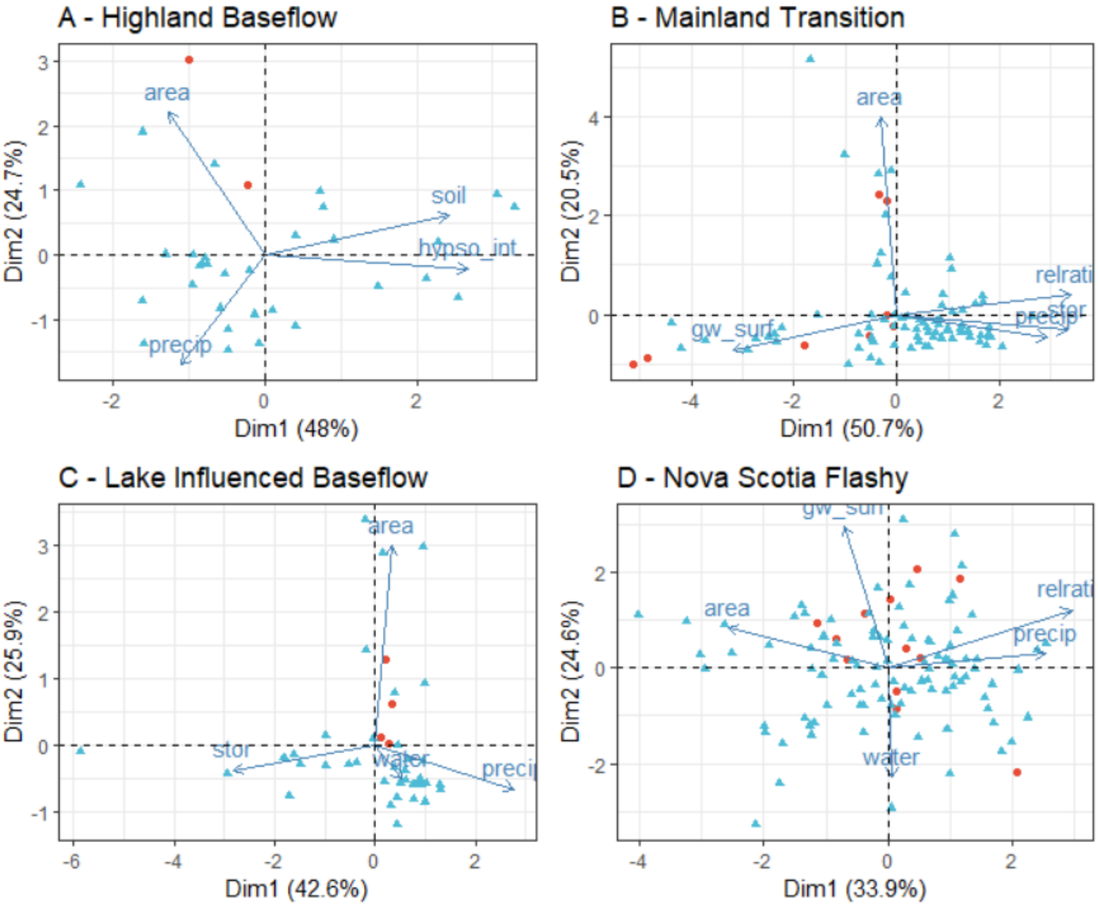

Figure 4.4. Explanatory variables for each meta-class visualized in two dimensions using PCA. Watersheds in the active hydrometric network are represented by red dots while secondary watersheds are represented by blue triangles.

Watershed area and annual precipitation were examined independently given their dominant influence on water yield and temporal hydrologic responses. The distribution of watershed areas within the active hydrometric monitoring network, as compared to the secondary watersheds, are illustrated for each meta-class in Figure 4.5. The two gauged watersheds in the **highland baseflow** meta-class are both mid-sized (120 and 365 km^2), while in contrast, secondary watersheds in this same class are as small as 11.5 km^2. This further illustrates that this meta-class is currently under-represented within the active hydrometric monitoring network. In the **lake-influenced baseflow** meta-class, there appears to be a gap in monitoring very small to small (< 100 km^2) and very large (> 1000 km^2) catchments of this class. Establishing a lake-influenced baseflow station on a very large watershed (> 1000 km^2) would not be justified, as there are very few lake-influenced baseflow watersheds of this size in the province. However, reallocating resources to a smaller, potentially nested catchment (10 – 100 km^2) could be valuable. Figure 4.5 further illustrates that the **Nova Scotia flashy** meta-class is the most comprehensively gauged meta-class. The range of watershed areas in the **mainland transition** meta-class also appears to be adequately represented in the active monitoring network.

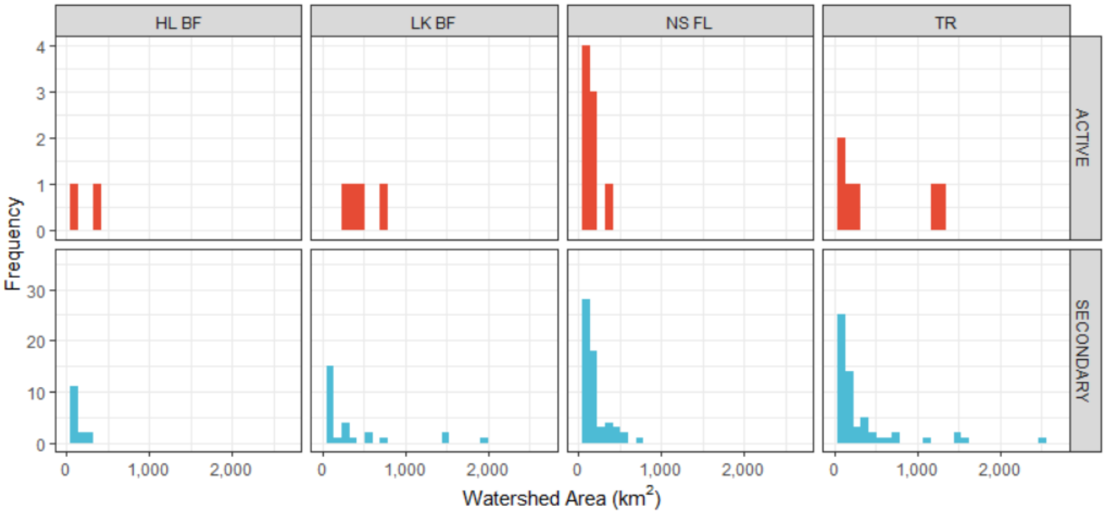

Figure 4.5. Histograms showing distribution of drainage area by meta-class for the active hydrometric network (red) and secondary watersheds (blue).

Figure 4.6 illustrates the distribution of annual precipitation for each watershed in the active hydrometric monitoring network as compared to the secondary watersheds. Catchments in the active hydrometric monitoring network are not exposed to the higher precipitation seen in **highland baseflow** (and presumably **highland flashy**) secondary watersheds. This gap is notable, as it is important to gauge the influence of these higher precipitation levels (> 1700 mm) that do not occur elsewhere in the province. In the **lake-influenced baseflow** meta-class, watersheds that experience moderate (< 1350 mm) and higher precipitation (> 1530 mm) are also not captured in the active hydrometric monitoring network. The active monitoring network better represents the range of precipitation in the **Nova Scotia flashy** and **mainland transition** meta-classes**.**

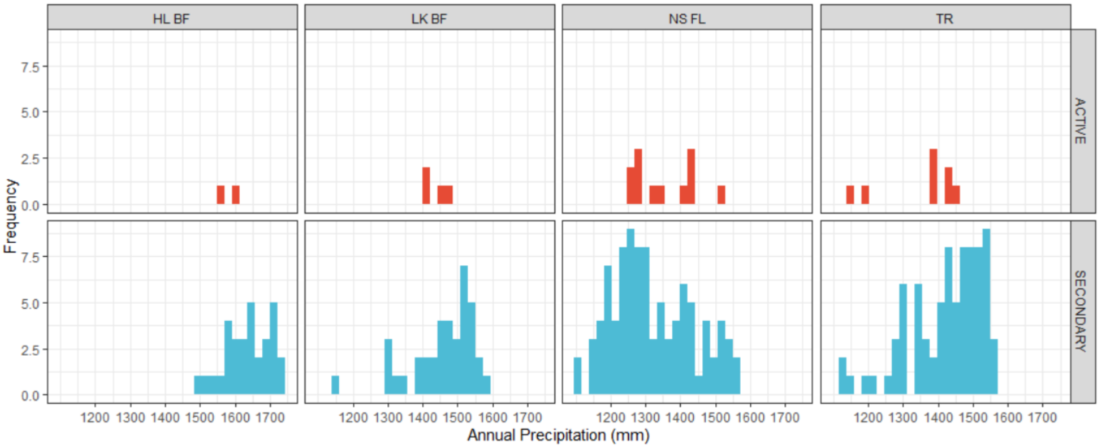

Figure 4.6 Histograms showing distribution of annual precipitation in mm by meta-class for the active hydrometric network (red) and secondary watersheds (blue).

4.4 CONCLUSIONS

In this study, we used two methods to evaluate a maritime hydrometric monitoring network's coverage and representativeness. The monitoring network was first assessed using the WMO (2008) guidelines for hydrometric station density. Then, a regional catchment classification scheme was used to categorize watersheds into generalized flow classes based on their physiographic characteristics. Additional data analysis and visualization approaches were used to determine how well these flow classes, and the variability that exists within them, are monitored. Before accounting for physio-climatic and socio-economic factors, the network meets WMO (2008) guidelines for minimum station density. However, the results of the assessment presented in this study illustrate that the current active hydrometric monitoring network does not adequately capture the range of hydrologic variability within this region: three of the five primary flow regimes observed in the province are not well represented by the active network. Although the province of Nova Scotia is relatively small, it possesses substantial physiographic variability, and moderate precipitation variability. This variability is difficult to monitor with a small network, as the province is made up of hundreds of secondary watersheds draining directly into the ocean. This differs from having a few watersheds draining large swaths of land, which might be typical of continental regions. This analysis also illustrated that the stations in the active monitoring network are not evenly distributed among the five generalized flow classes seen in Nova Scotia. Although the WMO (2008) provides a useful initial estimate of minimum station density, this study demonstrates the value in conducting additional analysis of the variability in key factors which influence hydrologic response in maritime regions. The methods applied in this study could serve as a useful framework for assessing hydrometric networks in other jurisdictions.

Chapter 5 Conclusions & Recommendations

5.1 SUMMARY OF RESEARCH

In this thesis, we explored the hydrologic behaviour of catchments in Nova Scotia, Canada. A group of 47 gauged catchments were characterized and classified using direct measures of hydrology (streamflow signatures). This revealed nine hydrologic clusters which were then merged into five generalized metaclasses: NS flashy, lake-influenced baseflow, mainland transition, highland flashy, and highland baseflow. Consolidating the nine hydrologic clusters into five metaclasses increased the spatial contiguousness of the catchments. Local hydrologic controls that differentiated these metaclasses were investigated, and the classification was reproduced using the identified form and forcing catchment descriptors to partition the classes, structured as a decision tree. Surface water storage emerged as a strong control on hydrology, initiating its own class (lake-influenced baseflow). We then assessed the coverage and representativeness of the province's hydrometric monitoring network using the decision tree to classify catchments into their respective meta-classes.

5.2 NETWORK RECONFIGURATION

The following modifications are suggested for network reconfiguration based on the results of the work completed:

1. The meta class **highland flashy** is not represented in the current monitoring network. We recommend establishing a minimum of two stations to monitor catchments in this meta-class.
2. We examined the network coverage of the remaining four metaclasses in terms of physiographic characteristics, drainage area, and precipitation.
    a. For **highland baseflow**, we recommend gauging a smaller catchment, nested within the already monitored Northeast Margaree River at Margaree Valley watershed, as well as another small catchment further north in the Cape Breton highlands (e.g. Effie's Brook in Victoria County).
    b. For the **Nova Scotia flashy meta-class**, we recommend discontinuing 2 – 3 of the twelve stations in this meta-class and reallocating these resources to other meta-classes.
    c. For the **lake-influenced baseflow meta-class**, we recommend gauging a smaller catchment. Again, this is an opportunity to introduce a nested station into an already monitored watershed (e.g. Shelburne Rive at Pollard's Fall Bridge or Roseway River at Lower Ohio). Secondarily, monitoring watersheds with different levels of precipitation (< 1350 and > 1500 mm) or a larger proportion of dominant surface storage (lake vs. wetland) would be valuable.
3. As suggested in previous network reviews, we recommend that new stations be coastal or nested if possible, as that is a gap in the current network.

5.3 INTEGRATED MONITORING

In the previous chapter, we established that although the network is currently meeting WMO (2008) minimum station density guidelines, there are opportunities to reconfigure the current monitoring network into a more representative system. Moving forward, when establishing or reactivating a station, network managers should bear in mind the location of permanent meteorological stations collecting precipitation data (Figure 5.1). Reference meteorological stations receive priority in operational funding and are less likely to be deactivated in the future. The information collected from hydrometric stations has much more value when the station is collocated with a suitable meteorological station.

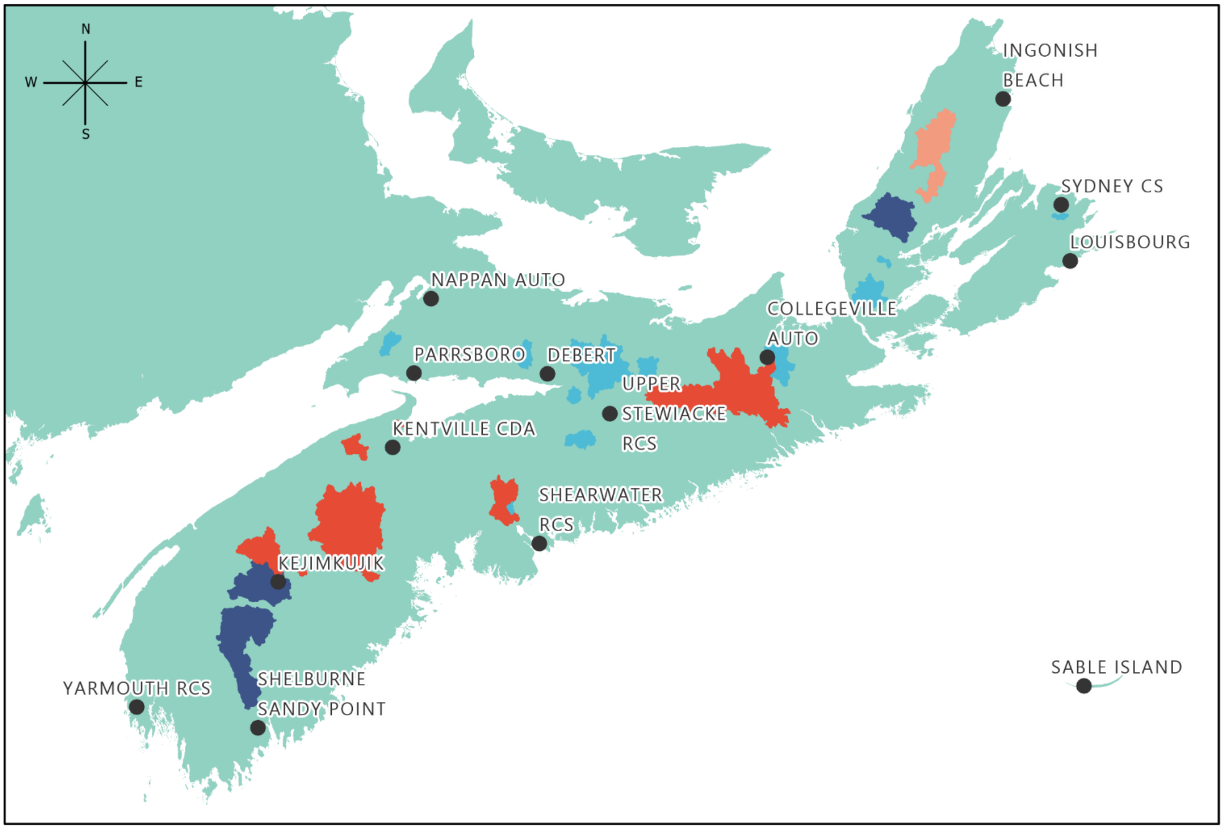

Figure 5.1. Map showing distribution of Water Survey of Canada monitored watersheds and climate reference stations across Nova Scotia; these stations are prioritized for continued operation

5.4 RECOMMENDATIONS FOR FUTURE WORK

The classification system presented in this thesis could serve as an organizing framework to conduct further research into the province's hydrology and related areas. This system serves a useful starting point for more detailed and specialized investigations and could be extended to other small jurisdictions.

1. The classification system contained in this thesis describes the overall similarity of the study catchments. This research can be extended in several ways:

    a. Individual classifications should be developed that are representative of seasonal catchment function (e.g. low flow classification) and used to assess the hydrometric network. Mishra et al. (2014) indicated that seasonality should be incorporated into network design (see also: Gnann et al., 2020).
    b. Development of a temperature-based catchment classification system and comparison to the hydrologic classification presented in this thesis (e.g. McManamay and DeRolph, 2019; Briggs et al., 2018; Johnson et al., 2020).
    c. Application of isotopic tracer methods and genohydrology (DNA-based methods; Good et al., 2018) to further validate the catchment classification system when extended to ungauged catchments.
2. Assess data quality and record length associated with each monitored meta-class (Hynds et al., 2019; Hannaford et al., 2013). For example, this would be useful in identifying maximum return period estimates and other data descriptors that could be reliably generated using these records.
3. Investigating beyond the current classification system:
    a. Establish a series of short-term stations in previously ungauged regions to determine if any flow regimes exist outside of the already identified meta-classes.
    b. Investigate hydrologic controls that are unique to the region's coastal catchments. How do catchment form factors and climate forcings influence these catchments?

References

Addor, N., Clark, M. P., Prieto, C., Newman, A. J., Mizukami, N., Nearing, G., & Le Vine, N. (2017). On the information content of hydrological signatures and their relationship to catchment attributes. *EGU General Assembly Conference Abstracts*, *19*, 9718. http://adsabs.harvard.edu/abs/2017EGUGA..19.9718A

Albers, S., Hutchinson, D., Dunnington, D., Whaley, R., Columbia, P. of B., rOpenSci), L. W.(Reviewed for, & rOpenSci), L. D. (Reviewed for. (2017). *tidyhydat: Extract and Tidy Canadian "Hydrometric" Data* (0.3.5) [Computer software]. https://CRAN.R-project.org/package=tidyhydat

Alfonso, L., He, L., Lobbrecht, A., & Price, R. (2013). Information theory applied to evaluate the discharge monitoring network of the Magdalena River. Journal of Hydroinformatics, 15(1), 211–228. https://doi.org/10.2166/hydro.2012.066
Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J., & Capell, R. (2012). A comparison of similarity indices for catchment classification using a cross-regional dataset. *Advances in Water Resources*, *40*, 11–22. https://doi.org/10.1016/j.advwatres.2012.01.008

Allchin, M. (n.d.). *Characterisation and Classification of Hydrological Catchments in Alberta, Canada Using Growing Self-Organising Maps*. Retrieved August 22, 2017, from http://leeds.gisruk.org/abstracts/GISRUK2015_submission_86.pdf

Auerbach, D. A., Buchanan, B. P., Alexiades, A. V., Anderson, E. P., Encalada, A. C., Larson, E. I., McManamay, R. A., Poe, G. L., Walter, M. T., & Flecker, A. S. (2016). Towards catchment classification in data-scarce regions. *Ecohydrology*, *9*(7), 1235–1247. https://doi.org/10.1002/eco.1721

Baechler, F. (2015). The geology and hydrogeology of faults on Cape Breton Island, Nova Scotia, Canada: An overview. *Atlantic Geology*, *51*, 242–268.

Baker, D. B., Richards, R. P., Loftus, T. T., & Kramer, J. W. (2004). A New Flashiness Index: Characteristics and Applications to Midwestern Rivers and Streams1. *JAWRA Journal of the American Water Resources Association*, *40*(2), 503–522. https://doi.org/10.1111/j.1752-1688.2004.tb01046.x

Beck, H. E., de Roo, A., & van Dijk, A. I. J. M. (2015). Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments*. *Journal of Hydrometeorology*, *16*(4), 1478–1501. https://doi.org/10.1175/JHM-D-14-0155.1

Bell, J. F., (1999). Tree-based methods. Pages 89 – 105 in Machine Learning Methods for Ecological Applications, edited by A. H. Fielding. Boston (MA): Kluwer Academic.

Black, A. R., Lees, M. L., Marsh, T. J., Law, F., & Dixon, J. M. (1995). *A review of the*

Northern Ireland hydrometric network. Report to the Department of the Environment for Northern Ireland Environment Service [Publication - Report]. http://nora.nerc.ac.uk/506977/

Boisvert, J., El-Jabi, N., El Adlouni, S.-E., Caissie, D., & Thiombiano, A. N. (2017). New Brunswick hydrometric network analysis and rationalization. *Canadian Journal of Civil Engineering*, *44*(10), 829–837. https://doi.org/10.1139/cjce-2016-0487

Boscarello, L., Ravazzani Giovanni, Cislaghi Alessio, & Mancini Marco. (2016). Regionalization of Flow-Duration Curves through Catchment Classification with Streamflow Signatures and Physiographic–Climate Indices. *Journal of Hydrologic Engineering*, *21*(3), 05015027. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001307

Bradford, R. B., & Marsh, T. J. (2003). Defining a network of benchmark catchments for the UK. *Proceedings of the Institution of Civil Engineers - Water and Maritime Engineering*, *156*(2), 109–116. https://doi.org/10.1680/wame.2003.156.2.109

Briggs, M. A., Johnson, Z. C., Snyder, C. D., Hitt, N. P., Kurylyk, B. L., Lautz, L., Irvine, D. J., Hurley, S. T., & Lane, J. W. (2018). Inferring watershed hydraulics and cold-water habitat persistence using multi-year air and stream temperature signals. Science of The Total Environment, 636, 1117–1127. https://doi.org/10.1016/j.scitotenv.2018.04.344

Brimley, B., Cantin, J. F., Harvey, D., Kowalchuk, M., Marsh, P., Ouarda, T. M. B. J., ... & Wedel, R. (1999). Establishment of the reference hydrometric basin network (RHBN) for Canada. Environment Canada, 41.

Brown, S. C., Lester, R. E., Versace, V. L., Fawcett, J., & Laurenson, L. (2014). Hydrologic Landscape Regionalisation Using Deductive Classification and Random Forests. *PLOS ONE*, *9*(11), e112856. https://doi.org/10.1371/journal.pone.0112856

Burn, D. H., & Goulter, I. C. (1991). An approach to the rationalization of streamflow data collection networks. *Journal of Hydrology*, *122*(1–4), 71–91.

Canada, E. and C. C., & Canada, E. and C. C. (2007, January 31). *About the Water Survey of Canada* [Program descriptions]. Aem. https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/about.html

Canada, E. and C. C., & Canada, E. and C. C. (2009, December 23). *Hydrometric program agreements and how they work* [Program descriptions]. Aem. https://www.canada.ca/en/environment-climate-change/services/water-overview/quantity/monitoring/survey/hydrometric-program-national-partnership/how-they-work.html

Canada, G. of C. and A.-F. (2013, February 7). *Canadian Soil Information Service*. http://sis.agr.gc.ca/cansis/publications/manuals/1998-cssc-ed3/index.html

Canadian Council of Ministers of the Environment. (2011). Selected Tools to Evaluate Water Monitoring Networks for Climate Change Adaptation. Canadian Council of Ministers of the Environment.

Caratti, J. F., Nesser, J. A., & Maynard, C. L. (2004). Watershed Classification Using Canonical Correspondence Analysis and Clustering Techniques: A Cautionary Note. *JAWRA Journal of the American Water Resources Association*, *40*(5), 1257–1268. https://doi.org/10.1111/j.1752-1688.2004.tb01584.x

Carlisle, D. M., Falcone, J., Wolock, D. M., Meador, M. R., & Norris, R. H. (2010). Predicting the natural flow regime: Models for assessing hydrological alteration in streams. *River Research and Applications*, *26*(2), 118–136. https://doi.org/10.1002/rra.1247

Carrillo, G., Troch, P. A., Sivapalan, M., Wagener, T., Harman, C., & Sawicz, K. (2011). Catchment classification: Hydrological analysis of catchment behavior through process-based modeling along a climate gradient. *Hydrology and Earth System Sciences*, *15*(11), 3411–3430. https://doi.org/10.5194/hess-15-3411-2011

Chiverton, A., Hannaford, J., Holman, I., Corstanje, R., Prudhomme, C., Bloomfield, J., & Hess, T. M. (2015). Which catchment characteristics control the temporal dependence structure of daily river flows? *Hydrological Processes*, *29*(6), 1353–1369. https://doi.org/10.1002/hyp.10252

Cobb, E. D., & Biesecker, J. E. (1971). The National Hydrologic Bench-mark Network. U.S. Department of the Interior, Geological Survey.

Coulibaly, P., Samuel, J., Pietroniro, A., & Harvey, D. (2013). Evaluation of Canadian National Hydrometric Network density based on WMO 2008 standards. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *38*(2), 159–167. https://doi.org/10.1080/07011784.2013.787181

Daly, C. (2006). Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, *26*(6), 707–721. https://doi.org/10.1002/joc.1322

De'ath, G., & Fabricius, K. E. (2000). Classification and Regression Trees: A Powerful yet Simple Technique for Ecological Data Analysis. *Ecology*, *81*(11), 3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Detenbeck, N. E., Brady, V. J., Taylor, D. L., Snarski, V. M., & Batterman, S. L. (2005). Relationship of stream flow regime in the western Lake Superior basin to watershed type characteristics. *Journal of Hydrology*, *309*(1), 258–276. https://doi.org/10.1016/j.jhydrol.2004.11.024

Devito, K., Creed, I., Gan, T., Mendoza, C., Petrone, R., Silins, U., & Smerdon, B. (2005). A framework for broad-scale classification of hydrologic response units on the Boreal Plain: Is topography the last thing to consider? *Hydrological Processes*, *19*(8), 1705–1714. https://doi.org/10.1002/hyp.5881

Environment and Climate Change Canada, (2019). Extracted from ECCC's HYDAT.mdb, released on January 17, 2019.

Environment Canada, & Nova Scotia Department of the Environment. (1985). *Hydrologic Network Review*.

ArcGIS Desktop - version 10.5.1. (2016). Software. Redlands, CA: ESRI.

Flores, A. N., Bledsoe, B. P., Cuhaciyan, C. O., & Wohl, E. E. (2006). Channel-reach morphology dependence on energy, scale, and hydroclimatic processes with implications for prediction using geospatial data. *Water Resources Research*, *42*(6), W06412. https://doi.org/10.1029/2005WR004226

Gnann, S. J., Howden, N. J. K., & Woods, R. A. (2020). Hydrological signatures describing the translation of climate seasonality into streamflow seasonality. Hydrology and Earth System Sciences, 24(2), 561–580. https://doi.org/10.5194/hess-24-561-2020

Good, S. P., URycki, D. R., & Crump, B. C. (2018). Predicting Hydrologic Function With Aquatic Gene Fragments. *Water Resources Research*, *54*(3), 2424–2435. https://doi.org/10.1002/2017WR021974

Grigg, D. (1965). The Logic of Regional Systems. *Annals of the Association of American Geographers*, *55*(3), 465–491. https://doi.org/10.1111/j.1467-8306.1965.tb00529.x

Grigg, D. B. (1967). Regions, Models and Classes. In *Models in Geography*. Methuen.

Halverson, M. J., & Fleming, S. W. (2015). *Complex network theory, streamflow, and hydrometric monitoring system design*. https://doi.org/10.5194/hess-19-3301-2015

Hamed, K. H., & Rao, A. R. (1998). *A modified Mann-Kendall trend test for autocorrelated data*. https://www.sciencedirect.com/science/article/pii/S002216949700125X

Hannaford, J., Holmes, M. G. R., Laizé, C. L. R., Marsh, T. J., & Young, A. R. (2013a). Evaluating hydrometric networks for prediction in ungauged basins: A new methodology and its application to England and Wales. *Hydrology Research*, *44*(3), 401. https://doi.org/10.2166/nh.2012.115

Hannaford, J., Holmes, M. G. R., Laizé, C. L. R., Marsh, T. J., & Young, A. R. (2013b). Evaluating hydrometric networks for prediction in ungauged basins: A new methodology and its application to England and Wales. *Hydrology Research*, *44*(3), 401–418.

https://doi.org/10.2166/nh.2012.115

Harrigan, S., Hannaford, J., Muchan, K., & Marsh, T. (2017). Designation and trend analysis of the updated UK Benchmark Network of river flow stations: The UKBN2 dataset. *Hydrology Research*, nh2017058. https://doi.org/10.2166/nh.2017.058

Hellebrand, H., Müller, C., Matgen, P., Fenicia, F., & Savenije, H. (2011). A process proof test for model concepts: Modelling the meso-scale. *Physics and Chemistry of the Earth, Parts A/B/C*, *36*(1), 42–53. https://doi.org/10.1016/j.pce.2010.07.019

Hijmans, R. J., Etten, J. van, Sumner, M., Cheng, J., Bevan, A., Bivand, R., Busetto, L., Canty, M., Forrest, D., Ghosh, A., Golicher, D., Gray, J., Greenberg, J. A., Hiemstra, P., Geosciences, I. for M. A., Karney, C., Mattiuzzi, M., Mosher, S., Nowosad, J., … Wueest, R. (2019). *raster: Geographic Data Analysis and Modeling* (3.0-7) [Computer software]. https://CRAN.R-project.org/package=raster

Huntington, T. G. (2006). Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology*, *319*(1), 83–95. https://doi.org/10.1016/j.jhydrol.2005.07.003

Hynds, P. D., Nasr, A. E., & O’Dwyer, J. (2019). Evaluation of hydrometric network efficacy and user requirements in the Republic of Ireland via expert opinion and statistical analysis. *Journal of Hydrology*, *574*, 851–861. https://doi.org/10.1016/j.jhydrol.2019.04.086

Ingledow, T. and Associates Limited, (1970). Hydrometric Network Plan for the Provinces of Newfoundland, New Brunswick, Nova Scotia, and Prince Edward Island.

Jin, Y., Liu, J., Lin, L., Wang, A., & Chen, X. (2017). Exploring hydrologically similar catchments in terms of the physical characteristics of upstream regions. *Hydrology Research*, nh2017191. https://doi.org/10.2166/nh.2017.191

Johnson, Z. C., Johnson, B. G., Briggs, M. A., Devine, W. D., Snyder, C. D., Hitt, N. P., Hare, D. K., & Minkova, T. V. (2020). Paired air-water annual temperature patterns reveal hydrogeological controls on stream thermal regimes at watershed to continental scales. Journal of Hydrology, 587, 124929. https://doi.org/10.1016/j.jhydrol.2020.124929

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.

Kendall, M. G. (1975). *Rank Correlation Methods*. Charles Griffin. https://www.abebooks.com/RANK-CORRELATION-METHODS-Kendall-M.G-Charles/10027955631/bd

Kennard, M. J., Pusey, B. J., Olden, J. D., Mackay, S. J., Stein, J. L., & Marsh, N. (2010a). Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology*, *55*(1), 171–193. https://doi.org/10.1111/j.1365-2427.2009.02307.x

Kennard, M. J., Mackay, S. J., Pusey, B. J., Olden, J. D., & Marsh, N. (2010b). Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. River Research and Applications, 26(2), 137–156. https://doi.org/10.1002/rra.1249

Kennedy, G. W. (2014). *Surficial Groundwater Regions of Nova Scotia* [Data set]. https://novascotia.ca/natr/meb/download/dp428md.asp

Kennedy, G. W., Drage, J., & Fisher, B. E. (2008). *Groundwater Regions of Nova Scotia* [Data set]. https://novascotia.ca/natr/meb/download/dp428md.asp

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (n.d.). A Quantitative Hydrological Climate Classification Evaluated with Independent Streamflow Data. *Water Resources Research*, *0*(ja). https://doi.org/10.1029/2018WR022913

Kuentz, A., Arheimer, B., Hundecha, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences; Katlenburg-Lindau*, *21*(6), 2863–2879. http://dx.doi.org/10.5194/hess-21-2863-2017

Laizé, C. L. R. (2004). Integration of spatial datasets to support the review of hydrometric networks and the identification of representative catchments. *Hydrology and Earth System Sciences*, *8*(6), 1103–1117.

Laizé, Cédric L. R., & Hannah, D. M. (2010). Modification of climate–river flow associations by basin properties. *Journal of Hydrology*, *389*(1–2), 186–204. https://doi.org/10.1016/j.jhydrol.2010.05.048

Lanfear, K. J., & Hirsch, R. M. (1999). USGS Study reveals a decline in long-record streamgages. *Eos, Transactions American Geophysical Union*, *80*(50), 605–607. https://doi.org/10.1029/99EO00406

Langbein, W.B. and Hoyt, W.G., (1959). Water facts for the nation’s future: Uses and benefit of hydrologic data programs. New York: The Ronald Press Company.

Latifovic, R., Pouliot, D., & Olthof, I. (2017). Circa 2010 Land Cover of Canada: Local Optimization Methodology and Product Development. Remote Sensing, 9(11), 1098. https://doi.org/10.3390/rs9111098

Leach, J. M., Kornelsen, K. C., Samuel, J., & Coulibaly, P. (2015). Hydrometric network design using streamflow signatures and indicators of hydrologic alteration. *Journal of Hydrology*, *529, Part 3*, 1350–1359. https://doi.org/10.1016/j.jhydrol.2015.08.048

Leibowitz, S. G., Comeleo, R. L., Wigington, P. J., Weber, M. H., Sproles, E. A., &

Sawicz, K. A. (2016). Hydrologic Landscape Characterization for the Pacific Northwest, USA. *JAWRA Journal of the American Water Resources Association*, *52*(2), 473–493. https://doi.org/10.1111/1752-1688.12402

Ley, R., Casper, M. C., Hellebrand, H., & Merz, R. (2011). Catchment classification by runoff behaviour with self-organizing maps (SOM). *Hydrol. Earth Syst. Sci.*, *15*(9), 2947–2962. https://doi.org/10.5194/hess-15-2947-2011

Li, C., Singh, V. P., & Mishra, A. K. (2012). Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information minimum redundancy. *Water Resources Research*, *48*(5). https://doi.org/10.1029/2011WR011251

Liermann, C. A. R., Olden, J. D., Beechie, T. J., Kennard, M. J., Skidmore, P. B., Konrad, C. P., & Imaki, H. (n.d.). Hydrogeomorphic Classification of Washington State Rivers to Support Emerging Environmental Flow Management Strategies. *River Research and Applications*, *28*(9), 1340–1358. https://doi.org/10.1002/rra.1541

Mackey, B. G., Berry, S. L., & Brown, T. (2008). Reconciling approaches to biogeographical regionalization: A systematic and generic framework examined with a case study of the Australian continent. *Journal of Biogeography*, *35*(2), 213–229. https://doi.org/10.1111/j.1365-2699.2007.01822.x

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, *13*(3), 245–259. JSTOR. https://doi.org/10.2307/1907187

McDonnell, J. J., & Woods, R. (2004). On the need for catchment classification. *Journal of Hydrology*, *299*(1–2), 2–3. https://doi.org/10.1016/j.jhydrol.2004.09.003

McKenney, D., Papadopol, P., Campbell, K., Lawrence, K., & Hutchinson, M. (n.d.). *Spatial models of Canada- and North America-wide 1971/2000 minimum and maximum temperature, total precipitation and derived bioclimatic variables.* 9.

McLeod, A. I. (2011). *Kendall: Kendall rank correlation and Mann-Kendall trend test* (2.2) [Computer software]. https://CRAN.R-project.org/package=Kendall

McManamay, R. A., & DeRolph, C. R. (2019). A stream classification system for the conterminous United States. *Scientific Data*, *6*, 190017. https://doi.org/10.1038/sdata.2019.17

McMillan, H., Westerberg, I., & Branger, F. (2017). Five guidelines for selecting hydrological signatures. *Hydrological Processes*, *31*(26), 4757–4761. https://doi.org/10.1002/hyp.11300

Mlynowski, T. J., Hernández-Henríquez, M. A., & Déry, S. J. (2011). An evaluation of hydrometric monitoring across the Canadian pan-Arctic region, 1950–2008. *Hydrology Research*, *42*(6), 479. https://doi.org/10.2166/nh.2011.105

Monk, W. A., Peters, D. L., Allen Curry, R., & Baird, D. J. (2011). Quantifying trends in indicator hydroecological variables for regime-based groups of Canadian rivers. *Hydrological Processes*, *25*(19), 3086–3100. https://doi.org/10.1002/hyp.8137

Nemec, J., & Askew, A. (1986). Mean and variance in network-design philosophies. *IAHS*, 123–131.

Nova Scotia Department of Natural Resources. (2006). *Enhanced Digital Elevation Model*.

Nova Scotia Environment. (2010). *Water for Life: Nova Scotia's Water Resource Management Strategy*. https://novascotia.ca/nse/water.strategy/docs/WaterStrategy_Water.Resources.Management.Strategy.pdf

Nova Scotia Geomatics Centre. (2006). *Enhanced Digital Elevation Model* [Map].

Nova Scotia Museum of Natural History Nova Scotia. (1996). Freshwater. In *Natural History of Nova Scotia* (Vol. 3). https://ojs.library.dal.ca/NSM/article/view/3755

Olden, J. D., Kennard, M. J., & Pusey, B. J. (2012). A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology*, *5*(4), 503–518. https://doi.org/10.1002/eco.251

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine Learning Methods Without Tears: A Primer for Ecologists. *The Quarterly Review of Biology*, *83*(2), 171–193. https://doi.org/10.1086/587826

Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, *19*(2), 101–121. https://doi.org/10.1002/rra.700

Oueslati, O., Girolamo, A. M. D., Abouabdillah, A., Kjeldsen, T. R., & Porto, A. L. (2015). Classifying the flow regimes of Mediterranean streams using multivariate analysis. *Hydrological Processes*, *29*(22), 4666–4682. https://doi.org/10.1002/hyp.10530

Patil, S. D., Wigington, P. J., Leibowitz, S. G., & Comeleo, R. L. (2014). Use of Hydrologic Landscape Classification to Diagnose Streamflow Predictability in Oregon. *JAWRA Journal of the American Water Resources Association*, *50*(3), 762–776. https://doi.org/10.1111/jawr.12143

Patil, S., & Stieglitz, M. (2011). Hydrologic similarity among catchments under variable flow conditions. *Hydrology and Earth System Sciences*, *15*(3), 989–997. https://doi.org/10.5194/hess-15-989-2011

Pearson, C. P. (1998). *Changes to New Zealand's national hydrometric network in the 1990s*. http://www.hydrologynz.co.nz/downloads/JoHNZ_1998_v37_1_Pearson(1).pdf

Pebesma, E. (2018). *Simple Features for R: Standardized Support for Spatial Vector Data*. *10*, 8.

Pellerin, J. (n.d.). *Updating the Canadian Reference Hydrometric Basin Network to detect climate-related trends in streamflow*. 142.

Peñas, F., Barquín, P., & Álvarez, C. (2016). *Sources of variation in hydrological classifications: Time scale, flow series origin and classification procedure*. *538*. https://doi.org/10.1016/j.jhydrol.2016.04.049

Philip, E., & McLaughlin, J. (2018). Evaluation of Stream Gauge Density and Implementing the Concept of Virtual Gauges in Northern Ontario for Watershed Modeling. *Journal of Water Management Modeling*. https://doi.org/10.14796/JWMM.C438

Pilon, P. J., Yuzyk, T. R., Hale, R. A., & Day, T. J. (1996). Challenges Facing Surface Water Monitoring in Canada. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *21*(2), 157–164. https://doi.org/10.4296/cwrj2102157

Post, D. A., & Jakeman, A. J. (1996). Relationships Between Catchment Attributes and Hydrological Response Characteristics in Small Australian Mountain Ash Catchments. *Hydrological Processes*, *10*(6), 877–892. https://doi.org/10.1002/(SICI)1099-1085(199606)10:6<877::AID-HYP377>3.0.CO;2-T

Pyne, M. I., Carlisle, D. M., Konrad, C. P., & Stein, E. D. (2017). Classification of California streams using combined deductive and inductive approaches: Setting the foundation for analysis of hydrologic alteration. *Ecohydrology*, *10*(3), n/a-n/a. https://doi.org/10.1002/eco.1802

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Razavi, T., & Coulibaly, P. (2013). Classification of Ontario watersheds based on physical attributes and streamflow series. *Journal of Hydrology*, *493*, 81–94. https://doi.org/10.1016/j.jhydrol.2013.04.013

Richter, B. D., Baumgartner, J. V., Powell, J., & Braun, D. P. (1996). A Method for Assessing Hydrologic Alteration within Ecosystems. Conservation Biology, 10(4), 1163–1174.

Sanborn, S. C., & Bledsoe, B. P. (2006). Predicting streamflow regime metrics for ungauged streamsin Colorado, Washington, and Oregon. *Journal of Hydrology*, *325*(1–4), 241–261. https://doi.org/10.1016/j.jhydrol.2005.10.018

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011). Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology & Earth System Sciences*, *15*(9), 2895–2911. https://doi.org/10.5194/hess-15-2895-2011

Sawicz, Keith. (2013). *Catchment classification- understanding hydrologic similarity through catchment function* [ProQuest Dissertations Publishing]. http://search.proquest.com/docview/1467508727/

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shiklomanov, A. I., Lammers, R. B., & Vörösmarty, C. J. (2002). Widespread decline in hydrological monitoring threatens Pan-Arctic Research. *Eos, Transactions American Geophysical Union*, *83*(2), 13–17. https://doi.org/10.1029/2002EO000007

Sivapalan, M. (2003). Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrological Processes*, *17*(15), 3163–3170. https://doi.org/10.1002/hyp.5155

Sivapalan, M. (2006). Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale. In *Encyclopedia of Hydrological Sciences*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470848944.hsa012

Snelder, T. H., & Biggs, B. J. F. (2002). Multiscale River Environment Classification for Water Resources Management1. *JAWRA Journal of the American Water Resources Association*, *38*(5), 1225–1239. https://doi.org/10.1111/j.1752-1688.2002.tb04344.x

Speed, M., Tetzlaff, D., Soulsby, C., Hrachowitz, M., & Waldron, S. (2010). Isotopic and geochemical tracers reveal similarities in transit times in contrasting mesoscale catchments. *Hydrological Processes*, *24*(9), 1211–1224. https://doi.org/10.1002/hyp.7593

Spence, C., Saso, P., & Rausch, J. (2007). Quantifying the Impact of Hydrometric Network Reductions on Regional Streamflow Prediction in Northern Canada. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *32*(1), 1–20. https://doi.org/10.4296/cwrj3201001

Ssegane, H., Tollner, E. W., Mohamoud, Y. M., Rasmussen, T. C., & Dowd, J. F. (2012a). Advances in variable selection methods I: Causal selection methods versus stepwise regression and principal component analysis on data of known and unknown functional relationships. *Journal of Hydrology*, *438*, 16–25. https://doi.org/10.1016/j.jhydrol.2012.01.008

Ssegane, H., Tollner, E. W., Mohamoud, Y. M., Rasmussen, T. C., & Dowd, J. F. (2012b). Advances in variable selection methods II: Effect of variable selection method on classification of hydrologically similar watersheds in three Mid-Atlantic ecoregions. *Journal of Hydrology*, *438*, 26–38. https://doi.org/10.1016/j.jhydrol.2012.01.035

Statistics Canada. (2019). Population estimates on July 1st, 2019. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501

Todd, M. J., Wigington, P. J., & Sproles, E. A. (2017). Hydrologic Landscape Classification to Estimate Bristol Bay, Alaska Watershed Hydrology. *JAWRA Journal of the American Water Resources Association*, *53*(5), 1008–1031. https://doi.org/10.1111/1752-1688.12544

Toth, E. (2013). Catchment classification based on characterisation of streamflow and precipitation time series. *Hydrology and Earth System Sciences; Katlenburg-Lindau*, *17*(3), 1149. http://dx.doi.org/10.5194/hess-17-1149-2013

Vörösmarty, C. J. (2002). Global water assessment and potential contributions from Earth Systems Science. *Aquatic Sciences*, *64*(4), 328–351. https://doi.org/10.1007/PL00012590

Wagener, T., Sivapalan, M., Troch, P., & Woods, R. (2007). Catchment Classification and Hydrologic Similarity. *Geography Compass*, *1*(4), 901–931. https://doi.org/10.1111/j.1749-8198.2007.00039.x

Walker, S. (2000). The value of hydrometric information in water resources management and flood control. *Meteorological Applications*, *7*(4), 387–397. https://doi.org/10.1017/S1350482700001626

Wallner, M., Haberlandt, U., & Dietrich, J. (2013). A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps. *Journal of Hydrology*, *494*, 59–71. https://doi.org/10.1016/j.jhydrol.2013.04.022

Whitfield, P. H., Burn, D. H., Hannaford, J., Higgins, H., Hodgkins, G. A., Marsh, T., & Looser, U. (2012). Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. *Hydrological Sciences Journal*, *57*(8), 1562–1579. https://doi.org/10.1080/02626667.2012.728706

Wigington, P. J., Leibowitz, S. G., Comeleo, R. L., & Ebersole, J. L. (2013). Oregon Hydrologic Landscapes: A Classification Framework1. *JAWRA Journal of the American Water Resources Association*, *49*(1), 163–182. https://doi.org/10.1111/jawr.12009

Williams, C. (n.d.). Nova Scotia Stream Orders [Map].

Winter, T. C. (2001). The Concept of Hydrologic Landscapes. *JAWRA Journal of the American Water Resources Association*, *37*(2), 335–349. https://doi.org/10.1111/j.1752-1688.2001.tb00973.x

Wolfe, J. D., Shook, K. R., Spence, C., & Whitfield, C. J. (2019). A watershed classification approach that looks beyond hydrology: Application to a semi-arid, agricultural region in Canada. *Hydrology and Earth System Sciences*, *23*(9), 3945–3967. https://doi.org/10.5194/hess-23-3945-2019

Wolock, D. M., Winter, T. C., & McMahon, G. (2004). Delineation and Evaluation of Hydrologic-Landscape Regions in the United States Using Geographic Information System Tools and Multivariate Statistical Analyses. *Environmental Management*, *34*(1), S71–S88. https://doi.org/10.1007/s00267-003-5077-9

World Meteorological Organization (Ed.). (2008). *Guide to hydrological practices* (6th ed). WMO.

Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, *30*(8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005

YiLan, L., & RuTong, Z. (2015). *clustertend: Check the Clustering Tendency* (1.4) [Computer software]. https://CRAN.R-project.org/package=clustertend

Appendix A

Table A.1 Variable loadings on first three principal components

	PC1	**PC2**	**PC3**
Q10	**0.82**	0.35	0.40
Q50	0.31	**0.91**	0.07
Q90	0.51	**0.68**	-0.50
FH	**0.63**	**-0.68**	-0.07
DH	**-0.61**	**0.66**	0.09
MAF	**0.80**	0.53	0.25
JAS	**0.80**	0.43	-0.07
AMJ	**0.81**	0.43	0.17
BFI	-0.47	**0.79**	-0.28
RBI	**0.62**	**-0.74**	0.03
T50	**0.63**	0.08	-0.14
EWF	0.56	**0.66**	0.33
FL	**0.74**	-0.42	-0.41
DL	-0.70	0.43	0.42
SQ10lp3	0.53	0.47	-0.65
FLgum	**0.82**	-0.30	0.25
FLlp3	**0.81**	-0.24	0.27

Table S2. Principal component score summary table by cluster

PC	Cluster	Min	Max	Median	SD
1	1	-3.53	-0.99	-2.29	0.99
	2	-1.89	-0.20	-0.47	0.57
	3	-2.07	0.94	-0.74	1.19
	4	-4.65	-1.08	-2.72	1.24
	5	0.21	2.36	0.83	0.83
	6	1.17	2.79	1.77	0.68
	7	-3.53	-2.43	-3.43	0.52
	8	2.73	4.57	3.66	0.86
	9	4.78	5.92	5.87	0.56
2	1	-2.47	0.24	-0.93	1.03
	2	-3.77	-1.12	-2.30	0.82
	3	-0.72	1.51	-0.09	0.84
	4	2.09	5.03	3.09	1.01
	5	-4.89	-1.01	-2.21	1.42
	6	-3.58	-0.54	-1.81	1.10
	7	-0.07	1.42	0.59	0.62
	8	1.62	3.84	3.00	0.96
	9	1.02	2.29	1.43	0.53
3	1	-4.22	0.08	-1.33	1.64

	2	-0.92	-0.03	-0.19	0.34
	3	0.18	1.85	0.88	0.70
	4	-2.02	1.45	0.44	1.15
	5	-0.88	-0.08	-0.33	0.32
	6	0.43	1.47	1.05	0.41
	7	0.31	1.09	0.96	0.35
	8	-2.17	-1.71	-1.84	0.21
	9	-0.16	2.34	1.35	1.05

Table A.2 Signature Variability by Cluster

Signature	**Cluster**	**Median**	**Minimum**	**Maximum**	**SD**
AMJ	1	0.030	0.025	0.033	0.003
AMJ	2	0.038	0.032	0.045	0.004
AMJ	3	0.041	0.038	0.053	0.007
AMJ	4	0.041	0.033	0.053	0.008
AMJ	5	0.041	0.033	0.046	0.005
AMJ	6	0.047	0.040	0.053	0.005
AMJ	7	0.037	0.035	0.039	0.002
AMJ	8	0.066	0.059	0.075	0.007
AMJ	9	0.094	0.068	0.101	0.015
BFI	1	0.53	0.43	0.60	0.07

BFI	2	0.38	0.34	0.49	0.05
BFI	3	0.43	0.38	0.52	0.05
BFI	4	0.65	0.62	0.87	0.09
BFI	5	0.37	0.23	0.42	0.07
BFI	6	0.35	0.31	0.38	0.03
BFI	7	0.54	0.49	0.62	0.05
BFI	8	0.57	0.54	0.61	0.03
BFI	9	0.38	0.38	0.46	0.04
DH	1	4.3	3.0	7.7	1.8
DH	2	2.8	2.3	3.3	0.4
DH	3	3.9	3.3	4.6	0.5
DH	4	8.6	6.4	12.1	1.9
DH	5	2.3	1.8	2.7	0.4
DH	6	2.6	2.3	2.8	0.2
DH	7	5.4	4.0	8.3	1.8
DH	8	3.1	2.9	4.3	0.6
DH	9	3.0	2.4	3.7	0.6
DL	1	11.8	6.3	13.3	3.1
DL	2	9.6	7.5	11.2	1.2
DL	3	16.5	9.8	19.7	3.5
DL	4	21.2	14.2	34.3	6.7
DL	5	7.1	4.5	8.3	1.4
DL	6	8.7	7.1	11.1	1.5

DL	7	25.3	19.9	32.7	5.5
DL	8	7.6	6.3	7.8	0.7
DL	9	7.4	5.4	9.7	1.8
EWF	1	0.031	0.030	0.038	0.004
EWF	2	0.038	0.030	0.041	0.004
EWF	3	0.048	0.041	0.055	0.006
EWF	4	0.052	0.044	0.060	0.005
EWF	5	0.037	0.036	0.048	0.005
EWF	6	0.049	0.037	0.057	0.007
EWF	7	0.043	0.037	0.043	0.003
EWF	8	0.059	0.053	0.062	0.004
EWF	9	0.056	0.053	0.077	0.011
FH	1	8.5	4.5	12.1	2.8
FH	2	12.8	10.8	15.5	1.6
FH	3	9.3	8.0	10.8	1.0
FH	4	4.2	2.7	5.0	0.8
FH	5	13.2	7.5	19.8	4.5
FH	6	14.3	12.8	15.5	1.1
FH	7	6.7	4.4	9.1	1.9
FH	8	11.5	8.2	12.1	1.8
FH	9	11.7	9.3	15.4	2.5
FL	1	3.1	2.8	5.4	1.2
FL	2	3.8	3.1	4.9	0.6

FL	3	2.2	1.9	3.7	0.7
FL	4	1.7	1.0	2.6	0.5
FL	5	5.1	4.4	8.0	1.4
FL	6	4.2	3.3	5.1	0.7
FL	7	1.5	1.1	1.8	0.3
FL	8	4.8	4.6	5.8	0.5
FL	9	5.0	3.7	6.7	1.2
FLgum	1	0.40	0.19	0.52	0.12
FLgum	2	0.94	0.72	1.02	0.12
FLgum	3	0.68	0.55	1.39	0.31
FLgum	4	0.23	0.19	0.40	0.07
FLgum	5	1.14	0.69	1.77	0.43
FLgum	6	1.55	1.08	2.38	0.53
FLgum	7	0.38	0.31	0.42	0.05
FLgum	8	1.27	1.01	1.53	0.22
FLgum	9	1.96	1.30	2.60	0.62
FLlp3	1	0.40	0.18	0.59	0.15
FLlp3	2	0.80	0.75	0.96	0.09
FLlp3	3	0.58	0.51	1.49	0.37
FLlp3	4	0.21	0.15	0.42	0.09
FLlp3	5	1.24	0.55	1.48	0.36
FLlp3	6	1.61	1.10	2.91	0.70
FLlp3	7	0.34	0.26	0.42	0.06

FLlp3	8	1.34	1.01	1.53	0.23
FLlp3	9	2.11	1.36	3.07	0.72
JAS	1	0.009	0.008	0.013	0.002
JAS	2	0.010	0.006	0.011	0.002
JAS	3	0.012	0.010	0.016	0.002
JAS	4	0.012	0.009	0.017	0.003
JAS	5	0.013	0.012	0.016	0.002
JAS	6	0.014	0.011	0.018	0.003
JAS	7	0.009	0.008	0.009	0.001
JAS	8	0.021	0.017	0.022	0.002
JAS	9	0.018	0.014	0.027	0.006
MAF	1	0.024	0.022	0.027	0.002
MAF	2	0.029	0.024	0.031	0.002
MAF	3	0.034	0.030	0.042	0.005
MAF	4	0.035	0.029	0.041	0.004
MAF	5	0.031	0.030	0.037	0.003
MAF	6	0.036	0.030	0.042	0.004
MAF	7	0.029	0.029	0.030	0.001
MAF	8	0.046	0.041	0.048	0.003
MAF	9	0.053	0.050	0.058	0.003
Q10	1	0.051	0.044	0.065	0.008
Q10	2	0.067	0.059	0.071	0.005
Q10	3	0.079	0.063	0.098	0.012

Q10	4	0.074	0.062	0.084	0.007
Q10	5	0.075	0.069	0.085	0.006
Q10	6	0.088	0.071	0.100	0.011
Q10	7	0.069	0.065	0.071	0.003
Q10	8	0.097	0.085	0.100	0.007
Q10	9	0.130	0.123	0.132	0.004
Q50	1	0.017	0.013	0.018	0.002
Q50	2	0.016	0.013	0.018	0.002
Q50	3	0.021	0.018	0.026	0.003
Q50	4	0.027	0.024	0.034	0.004
Q50	5	0.017	0.013	0.021	0.003
Q50	6	0.019	0.015	0.023	0.003
Q50	7	0.020	0.018	0.021	0.001
Q50	8	0.030	0.029	0.031	0.001
Q50	9	0.026	0.024	0.034	0.004
Q90	1	0.0036	0.0017	0.0086	0.0026
Q90	2	0.0023	0.0009	0.0038	0.0010
Q90	3	0.0032	0.0016	0.0050	0.0011
Q90	4	0.0054	0.0027	0.0100	0.0025
Q90	5	0.0030	0.0021	0.0052	0.0012
Q90	6	0.0026	0.0019	0.0045	0.0011
Q90	7	0.0026	0.0017	0.0032	0.0006
Q90	8	0.0115	0.0095	0.0126	0.0013

Q90	9	0.0071	0.0054	0.0080	0.0011
RBI	1	0.22	0.11	0.32	0.08
RBI	2	0.38	0.32	0.47	0.06
RBI	3	0.26	0.20	0.31	0.04
RBI	4	0.08	0.05	0.11	0.02
RBI	5	0.43	0.40	0.73	0.14
RBI	6	0.44	0.39	0.52	0.05
RBI	7	0.16	0.10	0.21	0.05
RBI	8	0.27	0.21	0.32	0.05
RBI	9	0.41	0.28	0.44	0.07
SQ10lp3	1	1.68	0.18	5.10	2.02
SQ10lp3	2	0.59	0.11	1.61	0.54
SQ10lp3	3	0.25	0.03	0.61	0.22
SQ10lp3	4	0.75	0.30	3.96	1.26
SQ10lp3	5	0.65	0.22	2.13	0.77
SQ10lp3	6	0.31	0.07	1.49	0.57
SQ10lp3	7	0.32	0.14	0.38	0.10
SQ10lp3	8	5.97	4.97	7.04	0.86
SQ10lp3	9	2.84	1.50	3.60	0.90
T50	1	163	146	171	9.6
T50	2	164	154	171	6.0
T50	3	160	144	166	8.7
T50	4	148	144	168	9.4

T50	5	164	147	175	10.6
T50	6	160	157	170	5.8
T50	7	155	151	160	3.5
T50	8	175	153	186	13.6
T50	9	197	152	206	24.7

Appendix B

Table B.1 Data sources required to calculate explanatory variables in decision tree.

Variable	Data Source
Soil Drainage	Detailed Soil Survey Compilations (Canada Soil Information Service, 2013)
Hypsometric Integral	Enhanced 20-m Digital Elevation Model (Geomatics Centre, 2006)
Lake Influence	Canada Land Cover MODIS
Glaciofluvial + Drumlin	Surficial Groundwater Regions of Nova Scotia (Kennedy, 2013)
Lake + Wetland	Canada Land Cover MODIS (Latifovic et al., 2017)
Relief Ratio	Enhanced 20-m Digital Elevation Model (Geomatics Centre, 2006)