Identifying streamflow changes in western North America from 1979 to 2021 using Deep Learning approaches

IDENTIFYING STREAMFLOW CHANGES IN WESTERN NORTH AMERICA FROM 1979 TO 2021 USING DEEP LEARNING APPROACHES

By Weigang TANG, Master of Science, Bachelor of Environmental Studies

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University September 30, 2022

McMaster University
Doctor of Philosophy (2022)
Hamilton, Ontario (School of Earth, Environment & Society)

TITLE: Identifying streamflow changes in western North America from 1979 to 2021 using Deep Learning approaches
AUTHOR: Weigang TANG (McMaster University)
SUPERVISOR: Dr. Sean K. CAREY
NUMBER OF PAGES: xii, 109

Abstract

Streamflow in Western North America (WNA) has been experiencing pronounced changes in terms of volume and timing over the past century, primarily driven by natural climate variability and human-induced climate changes. This thesis advances on previous work by revealing the most recent streamflow changes in WNA using a comprehensive suite of classical hydrometric methods along with novel Deep Learning (DL) based approaches for change detection and classification. More than 500 natural streams were included in the analysis across western Canada and the United States. Trend analyses based on the Mann-Kendall test were conducted on a wide selection of classic hydrometric indicators to represent varying aspects of streamflow over 43 years from 1979 to 2021. A general geographical divide at approximately $46°N$ degrees latitude indicates that total streamflow is increasing to the north while declining to the south. Declining late summer flows (July–September) were also widespread across the WNA domain, coinciding with an overall reduction in precipitation. Some changing patterns are regional specific, including: 1) increased winter low flows at high latitudes; 2) earlier spring freshet in Rocky Mountains; 3) increased autumns flows in coastal Pacific Northwest; and 4) dramatic drying in southwestern United States. In addition to classic hydrometrics, trend analysis was performed on Latent Features (LFs), which were extracted by Variation AutoEncoder (VAE) from raw streamflow data and are considered "machine-learned hydrometrics". Some LFs with direct hydrological implications were closely associated with the classical hydrometric indicators such as flow quantity, seasonal distribution, timing and magnitude of freshet, and snow-to-rain transition. The changing patterns of streamflows revealed by LFs show direct agreement with the hydrometric trends. By reconstructing hydrographs from select LFs, VAE also provides a mechanism to project changes in streamflow patterns in the future. Furthermore, a parametric t-SNE method based on DL technology was developed to visualize similarity among a large number of hydrographs on a 2-D map. This novel method allowed fast grouping of hydrologically similar rivers based on their flow regime type and provides new opportunities for streamflow classification and regionalization.

Acknowledgements

My PhD was a challenging but exciting journey, during which I received considerable help and support from my committee and colleagues. Firstly, I would like to thank my supervisor, Dr. Sean Carey, for his time, patience, kindness, open-mindedness, and generosity over the past few years. He respects my choice of research topic and provided me unreserved supports. I would also like to thank the members of my PhD committee, Dr. Paulin Coulibaly and Dr. Altaf Arain, for their helpful advice over the last few years. Dr. Mike Waddington and Dr. Ali Nazemi also provided me enlightening ideas and constructive suggestions for my PhD proposal and thesis. Additionally, many thanks to Global Water Future (GWF) and NSERC CREATE Program for their generous financial support and learning opportunities.

The McMaster Watershed Hydrology Group is the most hospitable and supportive team I could ever imagine. We had many great times and memories together at courses, conferences, meetings, and brewery. Special thanks to Nadine Shatilla, Kelly Biagi, Erin Nicholls, Graham Clark, and Arsh Grewal for your help and friendship. Also, special thanks to Keqi Wei from the ETFLab at McMaster ECE Department for those inspirational and enlightening ideas in Deep Learning technology. I am also thankful to Victor Cao, who is a friend of mine, for the comprehensive instruction and technical support in computer software and hardware.

I would also like to thank my family, especially my wife and parents, who gave me unconditional support throughout my entire PhD study. Their belief in me has kept my spirits and motivation high during this process.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Weigang TANG, declare that this thesis titled, “Identifying streamflow changes in western North America from 1979 to 2021 using Deep Learning approaches” and the work presented in it are my own. I confirm that:

- **Chapter 2**
  Tang, W., & Carey, S. K. (2022). Classifying annual daily hydrographs in Western North America using t-distributed stochastic neighbour embedding. *Hydrological Processes*, 36(1), e14473.
  Weigang Tang (dissertation author) is the first author, main researcher, and the corresponding author for this paper. Weigang Tang collected and processed daily streamflow data from USGS and WSC database and developed, trained, and tested the parametric t-SNE. The writing was executed by Weigang Tang with support and guidance from Dr. Carey.
- **Chapter 3**
  Tang, W., & Carey, S. K. (2022). Identifying the streamflow changes in western North America from 1979 to 2021 - Part 1: Insights from hydrometric trend analysis.
  Weigang Tang (dissertation author) is the first author, main researcher, and the corresponding author for this paper. Weigang Tang collected, processed, and analyzed streamflow data, ERA5 reanalysis climate variables, and shapefile of watersheds and glacier coverage and developed the companion dashboard application. The writing was executed by Weigang Tang with support and guidance from Dr. Carey.
- **Chapter 4**
  Tang, W., & Carey, S. K. (2022). Identifying the streamflow changes in western North America from 1979 to 2021 - Part 2: Insights from machine-learned features.
  Weigang Tang (dissertation author) is the first author, main researcher, and the corresponding author for this paper. Weigang Tang collected, processed, and analyzed streamflow data and developed, trained, and tested the VAE model. The writing was executed by Weigang Tang with support and guidance from Dr. Carey.

Chapter 1

Background

1.1 Climate Change and Streamflow in Western North America

Changes to streamflow volume and timing have been widely reported for Western North America (WNA) over the past century and are largely driven by human-induced climate change and natural climate variability. Increasing trends in air temperature in WNA have been observed since the 1960's, with warming patterns particularly strong in high latitudes and during winter (Hicke et al., 2022). WNA, defined here as eleven American states (Washington, Oregon, California, Idaho, Nevada, Montana, Utah, Wyoming, Colorado, Arizona, and New Mexico) and four Canadian provinces (British Columbia and Alberta) and territories (Yukon and Northwest Territory) that have hydrological regimes that are strongly affected by alpine and/or cryosphere processes (Fig. 2.1). In WNA, warming climate has altered streamflow regimes through a number of processes including: glacier wastage, permafrost thaw, snow-to-rain transition and enhanced evapotranspiration. In addition, there are associated change to general atmospheric circulation the absolute amount and timing of precipitation.

There is a diversity of flow regimes in WNA rivers and streams. Strong latitude and altitude gradients influence temperature the disposition of snow and rain throughout the year, whereas physiographic features such as the Cordillera have fundamental controls on moisture transport and the presence of glaciers. Furthermore, in the northern reaches of WNA permafrost is widespread and its presence strongly influences streamflow response.

Glacier mass loss in WNA has been the greatest globally since the 1960s (Zemp et al., 2019), and the losses have strongly accelerated in the recent decades (Hugonnet et al., 2021). Water derived from glacial wastage can provide important compensation for reduced summer flows, as glaciated watersheds have subdued losses during the late summer compared with unglaciated watersheds (Hodgkins, 2009; La Frenierre and Mark, 2014). However, this addition of water will decline as wastage continues, and (Moore et al., 2020) indicate that flows have already passed peak water at watersheds with relatively smaller glaciated areas; indicating the buffering effect of glaciers will recede in the coming decades. Permafrost thaw is accelerating and widespread in high latitudes of WNA (Smith et al., 2010), and there is considerable evidence that this is changing streamflow regimes. Most notably, significant upward trends in baseflow and winter flows have been observed for major north-draining basins (e.g. Mackenzie and Yukon Rivers) this is attributed to increases in groundwater associated with thawing permafrost and enhanced recharge and subsurface connectivity through deeper flow paths (Aziz and Burn, 2006; Walvoord and Striegl, 2007).

A more general phenomenon associated with warming is the snow-to-rain transition and earlier melt of snowpacks, which have been observed and is projected in climate models across the entire WNA domain (Knowles et al., 2006; Bintanja and Andry, 2017). Lower winter snow accumulation often results in a reduced spring freshet (Adam et al., 2009) whereas increased winter rainfall and intermittent melt led to more active winter flows (Musselman et al., 2018; Musselman et al., 2021). Widespread earlier snowmelt in WNA is a primary cause for declines in summer flows (Rood et al., 2008). In addition, while there is considerable spatial variability, annual precipitation decreased across WNA at the end of the last century (Hicke et al., 2022). Reduced precipitation coupled with enhanced evapotranspiration driven by warming collectively have caused the recent mega-droughts in south-western North America (Williams et al., 2020) and droughts are expected to occur with increasing frequency and severity in the coming century (Cook et al., 2015).

Streamflow variability in WNA is strongly influenced by large-scale oscillations in climate, including the Pacific Decadal Oscillation (PDO), Arctic Oscillation (AO), Pacific North America (PNA), and El Nino-Southern Oscillation (ENSO). Of these, PDO is considered the most influential as it is well correlated with regional air temperatures and precipitation anomalies across WNA, particularly during the cold season (Nov- Apr) and in high latitudes (Mantua and Hare, 2002). The warm phase of the PDO typically coincides with anomalously warmer temperatures in northwestern North America (i.e. western Canada and Alaska) and anomalously dry periods in interior Alaska and the Pacific Northwest and wet periods in coals Gulf Alaska and southwestern United States (Mantua and Hare, 2002). Typically, the cool and warm PDO phases shift every 20 to 35 years, with the last complete phase of PDO was a 30-year warm phase from 1977 to 2006 (Hodgkins, 2009). This warm phase of PDO is considered a main driver of accelerated mass loss and permafrost thaw across WNA over the last several decades (Brabets and Walvoord, 2009; Moore et al., 2009). In recent years, the regularity of the PDO has substantially declined, with shorter periodicity and weaker amplitude (Li et al., 2020). This reduced predictability and higher frequency shifts in the cool/arm phase act to complicate resolving patterns of change in WNA.

Changes to streamflow have significant ecological and societal impacts. For example, in headwater streams of the Rocky Mountains, decline summer streamflow when evaporative demand was greatest resulted in drought stress on floodplain forests (Rood et al., 2008). Many species of salmonoid fish are highly sensitive to stream thermal regimes, and declining flows and associated warmer water temperatures in the Canadian Rocky Mountains and the Pacific Northwest have negatively affected spawning, growth, and populations (Beechie et al., 2006; Battin et al., 2007; Williams et al., 2015). From a societal perspective, increases in water availability in the north have expanded interest in hydropower generation as there are numerous projects in the planning and development phase (Cherry et al., 2017). In contrast, severe and persistent drought in the southwestern United States brings massive challenges for existing water supply and governance systems. Water shortages for domestic, industrial and agricultural use have resulted in great loss of the economy of US Southwest and threaten food security (MacDonald, 2010). In 2022, Lake Meade was at its lowest level since it began filling behind the Hoover Dam in 1936. These are a small selection of the hundreds of challenges and opportunities facing WNA water resources. Considering the complexity of water supply, management and governance systems in North America, and the litany of challenges, understanding how climate change affects streamflow, and how we can detect it, provides key information for adaptive management under an uncertain future.

1.2 Streamflow Hydrometrics

Streamflow hydrometrics are pre-defined indicators that quantitatively characterize an aspect of streamflow. There are hundreds of hydrometrics that have been presented in the literature to describe the natural flow regime from various perspectives, induing flow magnitude, timing, frequency, duration, rate of change, low- and high-flow events (i.e. floods or droughts), and seasonality (Olden and Poff, 2003). Hydrometrics can be as simple as basic hydrograph statistics (e.g. average, minimum, maximum and standard deviation) or more complex indices such as Seasonality Index (SI, Coopersmith et al., 2014). Many hydrometric are derivative functions of the hydrograph such as the slope of the Flow Duration Curve (FDC) or the Baseflow Index (BI). There also exist combination hydrometrics that combine climate and flow information such as the runoff ratio, aridity index and streamflow elasticity, all of which are useful tools that link catchment response to seasonality and climate change.

Performing trend analyses on the time series of hydrometrics is one of the most standard approaches to identify and measure changes in streamflow over a period of interest. However, the natural drivers of streamflow variability and change (climate, ecosystem change) are complex and affect a variety of hydrological processes in multiple ways (Hicke et al., 2022). Therefore, it is not practical to fully describe the changes in streamflow with a single hydrometric, and a combination of hydrometrics is typically required to encapsulate changes in the flow regime. Subject to geographic variation in climate, geology, topography and vegetation cover, flow regimes often show remarkable regional dependency (Poff et al., 1997). Consequently, a hydrometric may be indicative for some flow regimes, but not useful for others. For example, the timing of spring peak flows are sensitive indicators of climate change in snow-dominated streams but are less useful for streams that are rain dominated or with little snow accumulation. As a result, to effectively understand changes to streamflow regime, a combination of hydrometrics are needed which often requires careful consideration and knowledge of the physical, hydrological and ecological processes in specific watersheds and regions. Furthermore, hydrometric combinations are commonly subject to information redundancy through intercorrelation, which reduces the effectiveness of trend detection (Olden and Poff, 2003). Selecting a subset of hydrometrics able to fully represent hydrological change while minimizing intercorrelation remains a challenge across larger scales.

1.3 Deep Learning in Hydrology

Deep Learning (DL) has been widely embraced in scientific research and applications in areas such as image classification and segmentation, signal de-noising, time series prediction and forecasting, and information compression. Compared with other disciplines, DL is relatively nascent in hydrological applications. However, there is considerable emerging research that explores the promise and feasibility of DL in hydrological applications that can be grouped into three main categories: 1) modelling dynamic hydrological variables, and 2) hydrological information extracted from remote sensing imagery (Shen, 2018). Two types of DL approaches have attached considerable attention in hydrology: Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN). LSTMs provide a powerful tool for time series prediction and forecasting and have been successfully applied in rainfall-runoff modelling (Kratzert et al., 2018), soil moisture prediction (Sungmin and Orth, 2021), and spatio-temporal gap filing of data (Ren et al., 2019). With regards to CNN, there has been considerable success in image segmentation and classification applied to extract vegetation (Kattenborn et al., 2021), soil (Padarian et al., 2019) and snow coverage (Wang et al., 2020) information from remote sensing products that act in support for hydrological modelling.

One of the most attractive capabilities of DL is automatic feature extraction from raw data, making it particularly useful for 'big data' analysis. Traditional machine learning algorithms are limited as they take only human-crafted features as inputs, yet representative features are often difficult to design and require considerable domain expertise (LeCun et al., 2015). In contrast, DL models are able to automatically learn useful features from raw data using a general-purpose learning procedure, which greatly reduces the time in data preparation. The limited requirement for human guidance and intervention makes DL a data driven approach that is particularly well suited for big data across multiple fields. Considering the ever-increasing volumes of data in hydrology, DL is an exciting tool to advance hydrological science.

DL networks are considered more interpretable than traditional neural networks, which hare often are referred to as 'black boxes' as their inner mechanisms are unknown and often hard to identify, quick and interpret. Consequently, this limits their utility in processes-based research such as hydrology as they are considered to have a lower confidence (Shen, 2018). In comparison, DL networks are referred to as 'grey boxes' that can be probed, investigated, and visualized to discover what patterns are recognized by the networks (Shen, 2018). For example, CNNs produce a feature map from every intermediate layers that offer useful insights of what patterns have been recognized at each layer, and associating those patterns that appear on feature maps with real-world objects in input images was found helpful to probe, investigate, and understand internal learning procedures of CNN models (LeCun et al., 2015). With improved transparency and the ability to interpret within the network, the promise of DL in hydrology is considerable across many aspects of the field.

1.4 Research Objectives

The objective of this thesis is to identify and attribute climate driven streamflow changes across WNA using traditional methods and deep learning technologies. More than 30,000 annual daily hydrographs (ADHs) form 700 undisturbed Water Survey of Canada and United States Geological Survey were gathered representing the breadth of WNA flow regimes. While there have been smaller-scale or country-specific investigations of trends in flow regimes within this domain (Lins and Slack, 1999; Zhang et al., 2001; Burn and Hag Elnur, 2002; Dery and Wood, 2005), the most recent analysis that covered the entire domain was published 17 years ago (Stewart et al., 2005). Considering the acceleration of climate change in WNA (Hicke et al., 2022) and the more frequent phase changes of large-scale elections (i.e. PDO, Li et al., 2020) there is a pressing need to revisit this region as a whole to evaluate the shifts in streamflow regimes. Moreover, the expansive nature of this data allows for the application of emerging DL models to detect streamflow changes.

This thesis consists of three distinct chapters that are written as journal manuscripts. In Chapter 2, a novel DL-based embedding technique was developed to visualize the similarity among the large number of hydrographs. Conventional methods of visualization similarly among time series (e.g. cross correlation matrices) are impractical and intractable for big data due to the high computational demand and challenge of pattern recognition and interpretation of results. In this chapter, I develop a novel method that integrates the state-of-the-art embedding technique, t-SNE, with a DL autoencoder. This approach provides and efficient and intuitive way to visualize similarity among the large number of ADHs, as the t-SNE projected the ADHs on to a 2D map where similar AHDs were grouped based on similarity in the flow regime. Furthermore, the original t-SNE was non-parametric and unable to provide parametric mapping from data space to t-SNE latent space (Maaten, 2009), which limited its ability to project new data on to existing t-NSE maps without retraining, limiting the utility in hydrological applications. To address this challenge, I used an autoencoder (an important member of the DL family) to approximate the non-linear mapping of t-SNE. This merging of t-SNE and autoencoder is a new parametric technique for embedding hydrograph data. Practically, incorporating this with clustering techniques, t-SNE map can facilitate classification and regionalization of streamflows and watersheds and benefit hydrological prediction and also be used to identify redundancy in hydrometric networks.

In Chapter 3, the streamflow changes in WMA over the last 43 years (1979 to 2021) were examined using classical trend analysis methods. Trend-Free Pre-Whitening (TFPW) Mann-Kendall test (Yue et al., 2002) was implemented with 55 annual-based hydrometrics that characterize streamflow from perspectives of flow magnitude, timing, duration, frequency, rate of changes, seasonality, and high- and low-flow events. As changes in streamflow have considerable spatial variability, the WNA domain was divided into six hydro-climate regions including Northwestern (Alaska, Yukon, Northwestern Territories, and northern BC), Canadian Rocky Mountains (CRM), U.S. Rocky Mountains (USRM), Coastal Pacific NorthWest (CPNW), Coastal Pacific Middle West (CPMW), and Southwest (Arizona, New Mexico, and southern California). Hydrometric trends detected from individual streams were synthesized on the basis of the sub-regions, and the changing patterns of streamflow for the six sub-regions were reported separately. In addition, grid-based trend analysis was performed for air temperature and precipitation collected from ERA5 reanalysis product (Hersbach et al., 2020) across WNA over the same observation period. The link between changing climate and hydrometric response for each region provided new insights into the attribution of hydrometric changes regional scales. A comparison of glaciated and non-glaciated streams in the regions of WNA was also completed. A dashboard app is provided to allow readers to further explore the dataset.

In Chapter 4, a DL-based approach was proposed to automatically extract features from ADHs and apply them to investigate changes in streamflow. DL's exceptional ability to learn features from raw data provides and opportunity to explore new concepts in hydrology and, perhaps, 'see the unseen'. Furthermore, the superior interpretability of DL networks offers new opportunities to explore the physical meaning of machine-extracted features. In this chapter, Variational AutoEncoder (VAE), a relatively new DL model, was developed and trained to extract Latent Features (LFs) from ADHs. LFs are considered machine-learning hydrological features (analogous to machine determined hydrometrics), and once extracted trend analysis was applied on the LFs. Spatial patters of trends in LFs were compared with those of the hydrometric produced in Chapter 3, and new patters of streamflow change revealed by the LFs are highlighted. Interpreting and understanding the hydrological implications of the LFs are at the core of this study. LF were related to classical hydrometric using Spearman rank correlation to establish how LF were related to the changes observed in Chapter 3. A key component of VAE, the decoder allows ADHs to be reconstructed form any given LFs, which I demonstrate is an effective way of identifying the hydrological implications of LFs and what changes are expected to occur in the hydrographs for different regions. This research is a pioneering application to design VAEs that suitable for streamflow data and is a promising DL-based approach for fast detection of streamflow change from a large number of streams.

Chapter 2

Classifying Annual Daily Hydrographs in Western North America using t-Distributed Stochastic Neighbor Embedding (t-SNE)

2.1 Introduction

Hydrologists typically acquire process knowledge from detailed place based studies and from representative experimental catchments, where hydrometric and biophysical attributes can be intensively measured over time. There are large number of global catchment observation networks, yet in many parts of the world they are in decline due to the expense in establishing, operating and maintaining their infrastructure (Laudon et al., 2017). Consequently, extrapolating process knowledge to watersheds that are hydrologically similar, yet not necessarily measured, has been a major focus of the hydrological community for the past several decades with initiatives such as the Prediction in Ungauged Basins (PUB) program (Sivapalan et al., 2003), whose goal was to predict flow quantiles at ungauged or poorly gauged basins according to the historical flow data collected at hydrologically similar basins.

Catchment classification has a long history as a means to generalize the functional behaviour that exists within watersheds, quantify their similarity, and to transfer information among them (Wagener et al., 2007). While there is no universal hydrological classification, the degree of similarity that exists is often defined from intrinsic and response characteristics of watersheds such as: climate (e.g. temperature, precipitation), watershed biophysical characteristics (e.g. geological conditions, soil type, relief, and vegetation), and the flow regime (e.g. annual hydrograph). Climate indices for classification (e.g. Köppen, Thornthwaite) are widely applied at varying time scales and have an long history identifying the intrinsic seasonality, thermal and moisture regimes of a region. Physiographic and biophysical indices such as soils, topography and geology strongly influence catchment behaviour (Buttle, 2006; Bormann, 2010), yet are not always ideal in defining process controls on catchment behaviour across scales and regions (Merz and Bloschl, 2005). Often, catchments with similar climate and physical conditions are not hydrologically similar (Oudin et al., 2010; Ali et al., 2012).

Evaluating catchment similarity based solely in terms of streamflow characteristics is popular; particularly in aquatic ecology where habitats are particularly sensitive to flow regimes (Poff et al., 1997). However, as Sawicz et al. (2011) notes, ecological studies are not typically aimed at understanding the behaviour of the catchment including the causes of a particular regime. Over time, the flow regime of a catchment is a descriptor of the seasonal behaviour of the streamflow (Haines et al., 1988) and by its nature is an integrator of a variety of hydrological processes produced by the interaction between climate and catchment physical characteristics. After decades of development, there are hundreds of indices available which quantitatively characterize five major components of flow regime: magnitude, timing, duration, frequency, and rate of change (Poff et al., 1997). Flow statistics (e.g. mean, max, and quantiles, standard deviation) at varying temporal scale are widely-used indices that reveal first-order information regarding magnitude, distribution, and variation of stream flow over a period of interest (Hall and Minns, 1999; Carey et al., 2010; Ali et al., 2012; Toth, 2013). More sophisticated indices, often explicitly reflecting specific hydrological processes, are preferred in catchment classification with respect to hydrological functions and system complexity (Sawicz et al., 2011). However, it remains a challenge to design a combination of hydrological indices that fully describe dominant hydrological characteristics of flow regimes, maximize distinctiveness among different flow regimes, as well as avoid information redundancy.

Classification based on flow statistics using clustering algorithms such as C-means and artificial neural networks (ANN) (Hall and Minns, 1999), hierarchical models (Snelder et al., 2005), and Bayesian clustering algorithm (Kennard et al., 2010; Sawicz et al., 2014), have been successfully applied for catchment classification and regionalization. The premise is to identify groups (or regions) in a way that similarity within a region is maximized whereas similarity between regions is minimized. Self-organized mapping (SOM), an unsupervised ANN machine learning technique has become increasingly appealing as it produces a low dimensional (typically two) representation of higher dimensional data that is simple to visualize. SOM preserves the topological structure of data as it transforms information from high-dimension feature space, and clusters information visually on maps where clustered points are more similar that distal points. When hydrological indices are transformed, catchments with homogeneous features are close on the 2-D map, and distance on the map can be used to visually infer similarity (Di Prinzio et al., 2011; Ley et al., 2011; Razavi and Coulibaly, 2013; Toth, 2013). Previously, SOM has been applied for catchment grouping with a moderate (∼50) number of samples (Ley et al., 2011; Toth, 2013), yet for extremely large data sets with thousands or millions of samples, computational time increases with sample size, challenging the utility of SOM application for very large data sets.

The objective of this research is to design and implement a novel method to visualize and classify streamflow regimes for a large streamflow data set focused on undisturbed rivers western North America. The classification is based on annual daily hydrographs (ADHs) from 304 sites over multiple years, providing 17110 ADHs for classification. The large nature of this data set renders traditional SOM impractical, and we therefore utilize t-distribution Stochastic Neighbor Embedding (t-SNE), an alternative machine learning algorithm proposed by Maaten (2009), to map ADHs to 2D feature space to assess flow similarity and compare this to traditional Principal Component Analyses. Furthermore, we develop an encoder neural network that allows additional data to be projected on to the t-SNE map; overcoming previous challenges with the non-parametric t-SNE technique. While this methodology only focuses on a limited region and does not attempt a universal classification, we attempt to show the novelty, flexibility and potential of this approach for future classification activities.

2.2 Methods

2.2.1 Data

Daily flow data were gathered from Western North America (WNA) stream gauges in the Referenced Hydro-metric Basin Network (RHBN) and Hydro-climate Data Network (HCDN), which are operated by Water Survey of Canada (WSC) and United States Geological Survey (USGS), respectively. Streams selected for the networks have predominantly natural flow regimes, with minimum human disturbance (e.g. significant land-use change, dams, reservoirs, and hydro-power stations) during long-term observation periods. In total, 304 HCDN and RHBN stream gauges were deemed suitable in WNA, including four Canadian provinces/territories (British Columbia, Alberta, Yukon, and Northwest Territory) and eleven American states (Washington, Oregon, California, Idaho, Nevada, Montana, Utah, Wyoming, Colorado, Arizona and New Mexico) (Fig. 2.1). The observation length varies among gauges, ranging from ten to over one hundred years. At each site, the long-term daily hydrograph was broken into the Annual Daily Hydrographs (ADHs). Each ADH contains 365 values of daily flow over a year from 1 January to 31 December (leap days excluded if applicable). Small gaps ($\leq$ 7 days) in ADHs were filled up via linear interpolation. ADHs with gaps > 7 consecutive days were excluded.

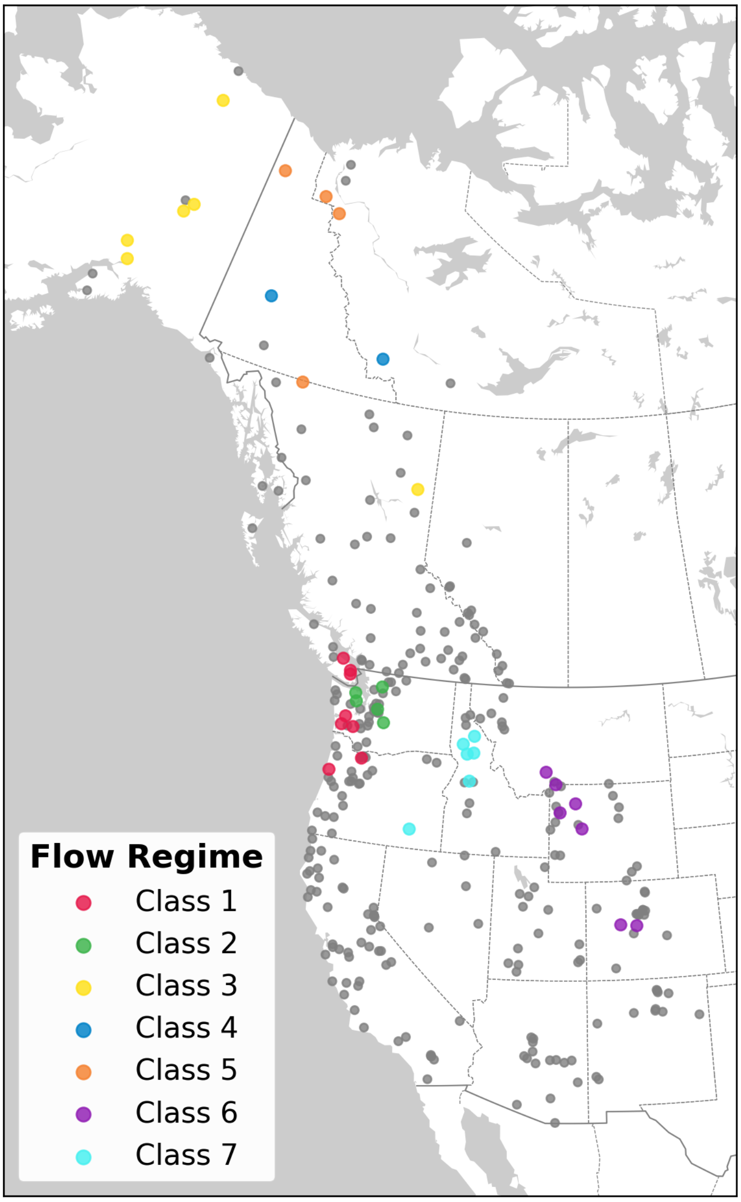


FIGURE 2.1: Selected HCDN and RHBN streamflow gauges in western North America. Those manually labeled with flow regime type are highlight with colors.

During initial data screening, ADHs with atypical shapes (e.g. sudden zig-zag patterns) were detected and excluded due to uncertainty in data quality. Furthermore, we excluded the ADHs of extreme years as they are less representative to the general flow pattern. In this case, extreme wet years are defined as ADHs with a maximum flow one hundred times higher than its long-term average, while extreme dry years are ADHs with a minimum flow 100 times smaller than its long-term average. From an initial set of 19499 ADHs, 17110 ADHs were preserved for analysis (see Fig. 2.2).

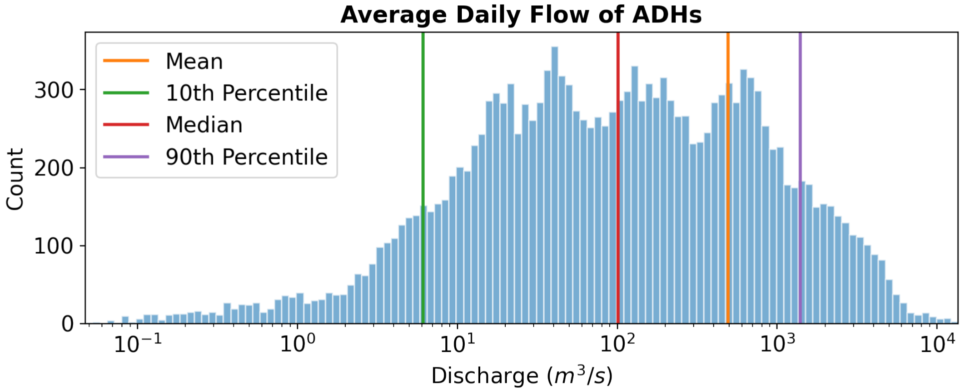


FIGURE 2.2: Distribution of average daily flow of the selected ADHs.

Selected ADHs were pre-processed prior to analysis. First, ADHs for a given stream were divided by its long-term average, which helps limit the scaling factor of watershed size and enhance comparability. Subsequently, a log transformation was applied to reduce skewness of the data as machine learning algorithms typically have improved performance on normally distributed data. A small number (10^{-6}) was added to ADHs to avoid invalid values during log transformation (see Eq. 2.1). Furthermore, min-max normalization (see Eq. 2.2) was applied to scale values of ADHs into a range of 0 and 1, which was recommended by the original paper of t-SNE (Maaten and Hinton, 2008).

$$f(x) = log(x + e^{-6}) \tag{2.1}$$

$$f(x) = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2.2}$$

where x denotes daily flow values of an ADH, as well as x_{min} and x_{max} are the minimum and maximum value of this ADH.

2.2.2 Labelled Samples

A subset of ADHs were manually labelled with flow regime types. The determination of flow regime type was based on visual assessment to the shape of ADHs, which is subjective to some degree. 2480 ADHs were labelled that represent seven distinctive flow regimes in the WNA domain (Fig. 2.3). ADHs selected for the same type typically come from streams that are geographically clustered. Class 1 is characterized by frequent rain events during winter and low flows in summer, with streams mostly located in coastal Pacific North West (PNW). Class 2 is similar to Class 1, but has a distinct snowmelt-driven spring freshet, and are mainly located in the interior PNW. Class 3 has extremely low winter flows, large spring freshet and many summer rain events. All of Class 3 gauges are located in Alaska. Class 4 exhibits a large spring freshet followed by summer storm events with gauges located in the Northwest Territories and northern Alberta. Class 5 is similar to Class 4, with later summer rain events and gauges are located in Yukon and northern British Columbia. Class 6 is characterized by dominant snowmelt freshet in spring that accounts for more than 70% of the annual discharge with gauges primarily located Canadian Rockies between British Columbia and Alberta. Class 7 has a spring freshet along with occasional winter events and gauges are most often located in Idaho. The labelled ADHs were used to evaluate the performance of t-SNE, with details provided in later sections.

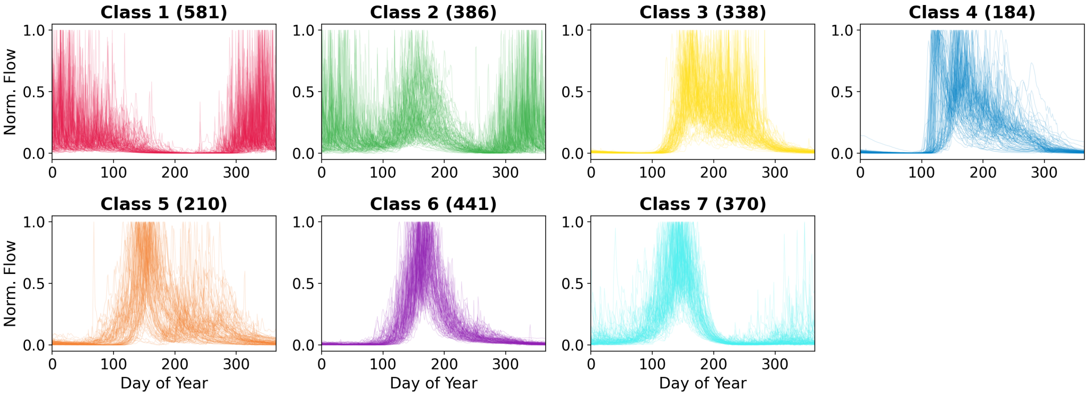


FIGURE 2.3: ADHs for the seven flow regimes normalized by min-max method. The number in the parenthesis indicates the number of ADHs in that class. Day of year starts with January 1st.

2.2.3 t-distribution Stochastic Neighbor Embedding (t-SNE)

t-distribution Stochastic Neighbor Embedding (t-SNE; Maaten and Hinton, 2008) is a state-of-the-art technique for dimensionality reduction and high-dimensional data visualization. It is a variant of the SNE that was originally proposed by

Hinton and Roweis (2002). SNE represents similarity between datapoints using conditional probabilities that are converted from pairwise Euclidean distances (Eq. 2.3).

$$p_{i|j} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\|x_i - x_k\|^2/2\sigma_i^2)} \tag{2.3}$$

where x_i and x_j denote datapoints in the original feature space, $\|x_i - x_j\|^2$ is the Euclidean distance, $p_{i|j}$ is the conditional probability between x_i and x_j, and σ_i is the variance of the Gaussian distribution centered at x_i. The value of σ_i primarily depend on the data density at x_i, which varies among datapoints. Details on how to determine the value of σ_i can be found in the paper of Maaten and Hinton (2008). The conditional probabilities between the low-dimensional counterpart y_i and y_j is indicated as $q_{i|j}$. Through minimizing the differences between $p_{i|j}$ and $q_{i|j}$, SNE attempts to copy the local and global structure of datapoints from original to new feature space. Kullback-Leibler Divergence (KLD), a measure of difference between probabilities, is employed by SNE as the loss function (see Eq. 2.4).

$$KL(P \parallel Q) = \sum_i \sum_j p_{ij} \cdot log\frac{p_{ij}}{q_{ij}} \tag{2.4}$$

Compared with original SNE, t-SNE has two major adjustments: 1) the use of joint probabilities (Eq. 2.5) instead of conditional probabilities to represent similarities, and 2) the use of a Student t-distribution instead of Gaussian distribution to compute similarities between two datapoints in low-dimensional space (Eq. 2.6).

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2N} \tag{2.5}$$

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k\neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{2.6}$$

where N is the total number of datapoints, where y_i and y_j denote datapoints in the embedded feature space, and p_{ij} and q_{ij} denote the joint probability in high- and low-dimensional space. The first adjustment simplifies the form of loss function improving training efficiency. The second adjustment effectively alleviated the "crowding problem", a common challenge faced by SNE and many other dimensionality reduction techniques (Goodfellow et al., 2016). Datapoints that lie moderately distant in high-dimensional space tend to be crushed together in the embedded space, which prevents gaps from forming between natural clusters. Using a heavy-tailed distribution (i.e. Student t-distribution) to calculate pairwise similarities for low-dimensional representation effectively alleviates the crowding problem and preserves the local and global structure of datapoints in the embedded space. In all trials of Maaten and Hinton (2008), t-SNE produced considerably better visualization than other embedding techniques, including SNE, Sammon mapping, curvilinear components analysis, Isomap, maximum variance unfolding, locally linear embedding, and Laplacian Eigenmaps.

Datapoint arrangement on t-SNE maps are sensitive to two parameters: 1) perplexity and 2) learning rate. The perplexity is considered a smooth measure of the effective number of neighbors and is used to determine σ_i of the Gaussian distribution for high-dimensional datapoints (Maaten and Hinton, 2008). Intuitively speaking, it controls the "roundness" of the arrangement of the datapoints in the embedded space. Conventionally, perplexity is chosen from a range between 5 and 50, with Maaten and Hinton (2008) recommending a perplexity of 30. Learning rate (η) is an important parameter for optimization that determines the final convergence of loss function. In this study, we follow the process of Maaten and Hinton (2008) and choose perplexity as 30 and η as 100.

Results from the t-SNE map were compared with Principle Component Analysis (PCA), likely the most popular linear embedding technique and evaluated based on the separability of ADHs with distinct flow patterns. A subset of ADHs were labelled with flow regime type as detailed earlier, and then projected on the 2D map using t-SNE and PCA respectively. K-Nearest Neighbors (KNN, Goldberger et al., 2005) was used to classify datapoints on both embedded maps, with k set to 30. Favorable embedding techniques should arrange the ADHs of different types into separable clusters on 2D map and allow accurate classification with simple classifiers (e.g. KNN). Here, we employed classification accuracy as a quantitative indicator for separability of datapoints, and better embedding technique should result in higher accuracy. Tools for t-SNE, PCA, and KNN are all available in *scikit-learn* package in Python (Pedregosa et al., 2011).

As a non-parametric method, t-SNE does not produce the mapping function between high- and low-dimensional data representation. It is impossible to project additional data on to an existing t-SNE map, which has been recognized as a major shortcoming of t-SNE technique (Maaten and Hinton, 2008). One solution is to merge new data to the original dataset and re-run t-SNE, yet it is computationally inefficient for large datasets. In addition, datapoints in the original dataset could be displaced on the new t-SNE map as inclusion of new data alters the similarity matrix (i.e. pairwise joint probabilities). The location inconsistency of embedded datapoint is highly unfavourable. To address this, Maaten (2009) proposed an alternative solution that build an parametric t-SNE through incorporating an autoencoder.

2.2.4 Autoencoder

An autoencoder is a neural network that is trained to copy inputs to outputs through the use of an encoder and decoder. The encoder function $\mathbf{h} = f(\mathbf{x})$ converts input data $\mathbf{x}$ to latent features $\mathbf{h}$, while the decoder function $\mathbf{r} = g(\mathbf{h})$ reconstruct data from latent features to its original format. As a lossy technique, autoencoders are not trained to copy perfectly, but to transfer the most salient information of input data to latent features (with less dimensions) while ignoring noise. This approach is widespread in dimensional reduction and feature learning (Goodfellow et al., 2016).

Encoder networks are often considered a universal function approximators. They are trained to approximate complicated, non-linear functions that map high-dimensional data to low-dimensional representation. Here, we use an encoder network to approximate the mapping between ADHs and t-SNE 2D data. Consequently, newly collected ADHs can be projected on the existing t-SNE map using the trained encoder, and there is no need to re-run t-SNE with the entire dataset. The objective is to minimize the distances between the datapoints projected by the t-SNE and encoder. Mean Absolute Error (MAE) was employed as the loss function.

The performance of the encoder network largely depends on the choice of model hyperparameters, which are referred to as untrainable parameters as they do not change during the training procedure. The network architecture (i.e. number of layers and nodes) and activation function are critical hyperparameters for the encoder. In an iterative manner, we tested a variety of architectures and activation functions in order to search for the optimal model configuration (see Table A1.1 in Appendix A). ADH samples are split into two subsets: training (75%) and testing (25%). The dimension of input layer is 365 (match to the values in ADHs), and the output layer has two dimensions. Network architecture defines the depth (i.e. number of layers) and width (i.e. number of nodes) of the encoder network. Our baseline architecture is a three-layer fully-connected network, with 512, 256, and 128 nodes for each layer. Activation functions bring non-linearity to neural networks, and enable the encoder model to approximate complicated, non-linear functions. The choice of activation function affects the optimization process (Goodfellow et al., 2016), and we selected the Rectified Linear Unit (ReLU, as $f(x) = \max(0, x)$) as the default function for the baseline model. A modified version of ReLU, called leaky ReLU (as $f(x) = \max(a * x, x)$), has been recommended as providing additional network optimization benefits (Maas et al., 2013), and a number were tested with tuned a ranging from 0.02 to 0.4. A dropout layer (Srivastava et al., 2014) was added after the first dense layer to avoid overfitting, and the dropout rate is tuned within a range between 0.1 and 0.2. The optimizer was chosen as Adam (Kingma and Ba, 2015) for all tested encoders. It employed a default learning rate of 0.001. Each model was trained 1000 epochs during the training procedure. The encoder networks were built using *Keras* (2.4.3 version), a Python package for deep learning (Chollet et al., 2015).

2.3 Results

2.3.1 t-SNE vs PCA

Following the normalization of the ADHs, a PCA and t-SNE were implemented on the labelled data (Fig. 2.4). Although the first two PCs explained 69.8% of the total variance, the PCA suffered from the "crowding issue", with points clustered within a relatively narrow region on the 2D map. In contrast, t-SNE had points that were more evenly dispersed across the 2D map, yet with distinct clusters. Classes 1, 2 and 7 were notably separate on both PCA and t-SNE maps, with class 6 more separable on the t-SNE map. Classes 3-5 were mixed together on both maps, yet they form an individual cluster on the t-SNE map that is isolated from the other clusters. All three of these classes were dominated by gauges in the northern part of the domain (Alaska, Yukon, Northwest Territories, and northern British Columbia), highlighting their similarity (Fig. 2.3). Consequently, we merged Classes 3-5 into one (as Class 5) for subsequent analyses.

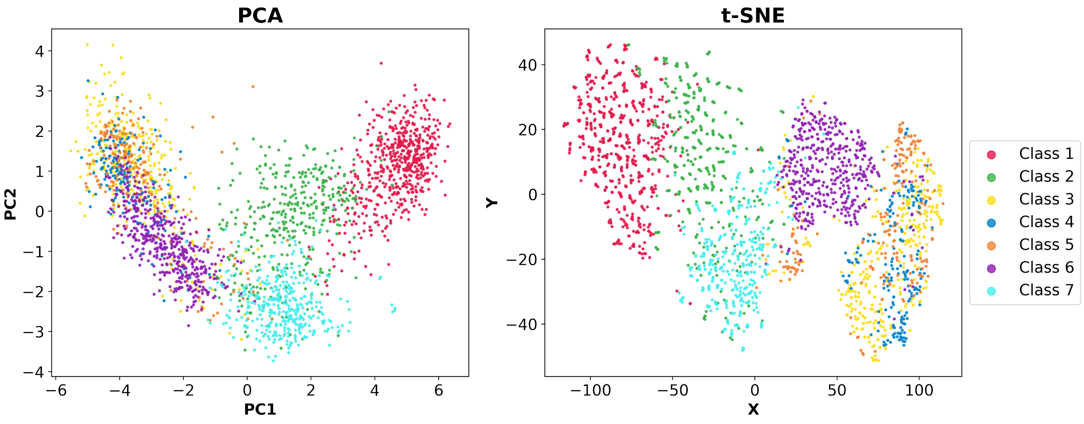


Figure 2.4: 2D map of PCA and t-SNE with labelled ADHs of seven flow regimes. Color of data points indicate flow regime type.

Based on the ADH arrangement on 2D map and KNN classification accuracy, t-SNE was superior to PCA. The PCA and t-SNE transformed data were classified using KNN, with a classification accuracy of 84.5% (PCA) and 90.3% (t-SNE), indicating the t-SNE map had enhanced separability among the flow regimes. A confusion matrix (Table 2.1) indicated that misclassification most often occurred among ADHs from Class 2 (the interior PNW) and Class 7 (predominantly Idaho), which are geographically proximal. This is reasonable as both flow regimes were characterized by a large spring freshet and winter storm events, yet flows in Class 7 typically had a more dominant freshet and less winter events than Class 2. In certain years dominated by either greater/less snow and rain in the winter resulted in ADHs resembling the other class. This highlights the challenge of subjectively labelling ADHs for classification, and suggests that refining the selection of labelled AHDs for these two classes may be warranted.

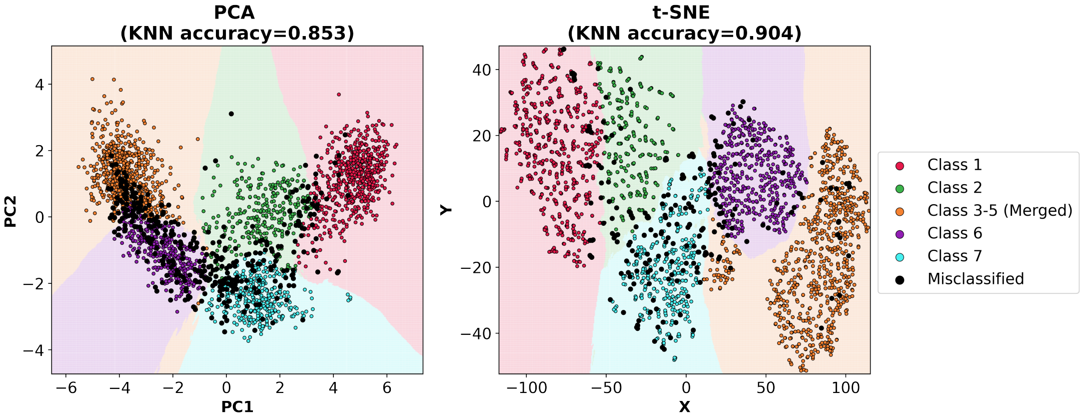


FIGURE 2.5: KNN-classified data points on PCA and t-SNE map. Black points represent misclassified ADHs.

TABLE 2.1: Confusion matrix of KNN classification with t-SNE datapoints

	Class 1	Class 2	Class 5	Class 6	Class 7
Class 1	0.996	0.004	0.000	0.000	0.000
Class 2	0.043	0.823	0.000	0.003	0.131
Class 5	0.000	0.000	0.989	0.011	0.000
Class 6	0.000	0.000	0.000	1.000	0.000
Class 7	0.000	0.019	0.000	0.000	0.981

2.3.2 Global t-SNE Map

A global t-SNE map was constructed with all ADHs, with labelled ADHs highlighted in color based on their flow regime (Fig. 2.6). ADHs from the same flow regime remained clustered on the map, yet the absolute location of labelled ADHs has changed from the previous map (Fig. 2.4). However, the topological relations among the clusters of flow regimes were largely preserved. For instance, the clusters for Class 1 and 2, which have similar ADH shape, remained close on the global t-SNE map, while those that were distinct (e.g. Class 1 and 6) are widely spaced across the map.

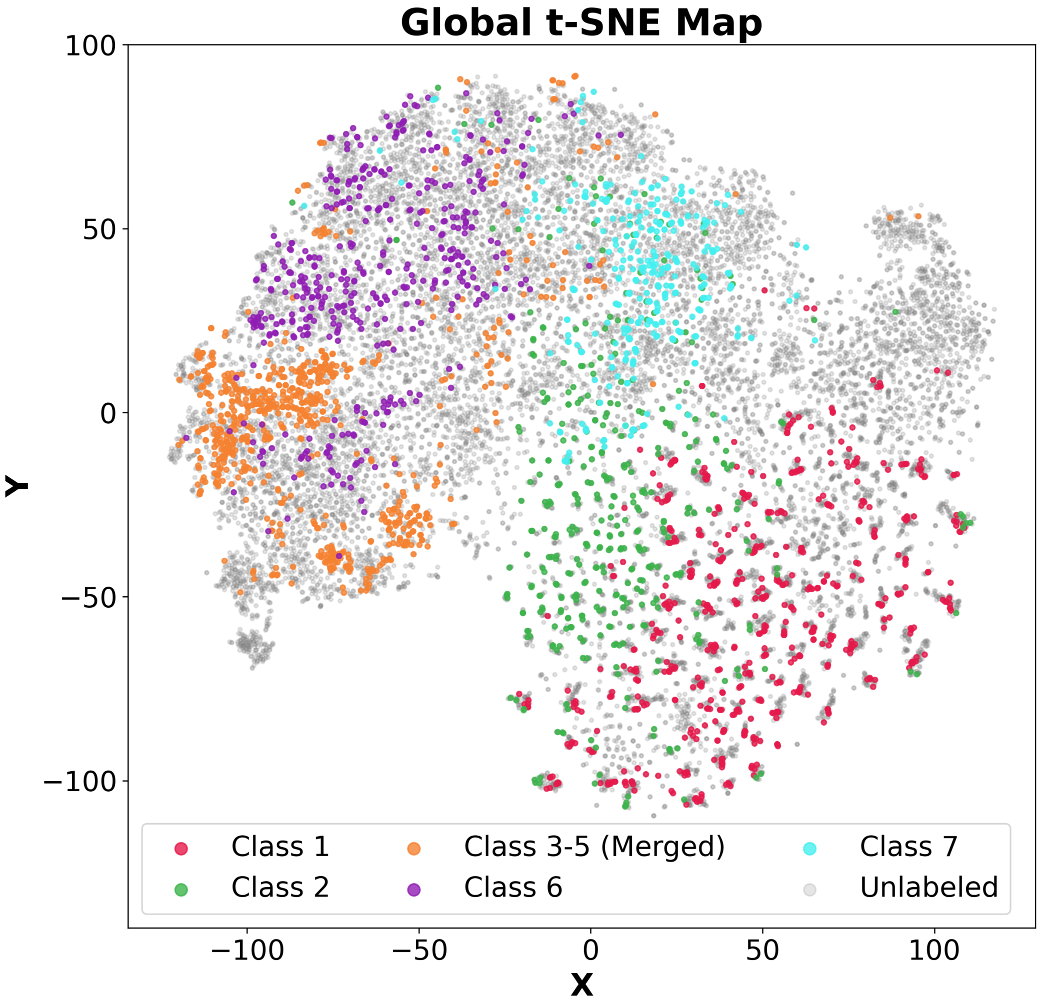


FIGURE 2.6: t-SNE map with all ADH samples. ADHs labeled with flow regime type are highlighted with colors.

A large number of compact clusters on the t-SNE map were observed for Class 1 and 2 as the ADHs in these clusters were highly correlated (Fig. 2.7). Both classes are featured by high-frequency rainfall events during wet season. The timing of rainfall occurrences are usually random, resulting in large variability in ADH shape. Therefore, the ADHs of those two classes fall into a broader zone on the t-SNE map in comparison with those of snow-dominated regime (e.g. Class 6 and 7). Approximately 50 of those clusters were scrutinized, and they result from same-year ADHs from a number of geographically proximate watersheds located near the Pacific coast (i.e. southern BC, Washington, Oregon, and northern California). This suggests that climate for a given year drives similar hydrological responses across these watersheds, which have highly similar responses. There are 24 streams whose ADHs consistently occur in these clusters.

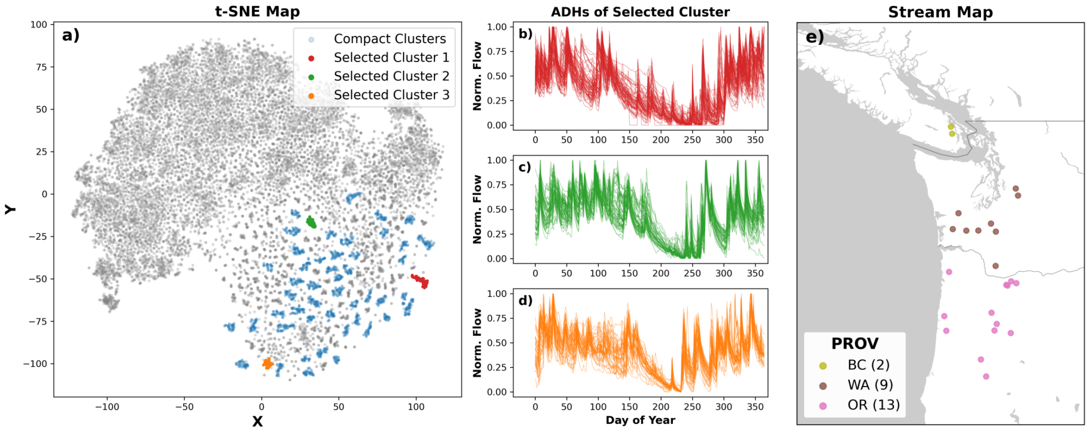


FIGURE 2.7: Small, compact clusters of ADHs are mostly observed in the division of Class 1 and 2 (a). ADHs of the three selected clusters exhibit strong correlation within clusters (b, c, d). Streams whose ADHs are frequently included in the compact clusters are all located in coastal PNW region (e).

2.3.3 Encoder

A number of encoders were tested, and the encoder with best MAE on the testing dataset employed a nine-layer architecture with activation functions of LeakyReLU (alpha=0.07) (see Encoder 50 in Table A1.1). The Dropout layer slightly improved loss and remarkably reduced the gap in MAE between the training and testing set. This trained encoder achieved a MAE of 1.07 on training set, and 3.81 on testing set. The average displacement between the projected ADH points and their encoded counterparts on the t-SNE map was 5.93 for the testing set, which was small compared to the extent of the map (Fig. 2.8). The points formed a near-circular shape distribution on t-SNE map (see Fig. 2.6), so the map extent was measured by a radius to the map centroid. The radius that covers 98% data points on t-SNE map was 115.9. The ratio of average displacement to map radius is only 0.05, suggesting the errors of encoder projection are limited. For each flow regime class, the within-class distance is calculated as the average distance of datapoints to the centroid of the class. The within-class distance for the five flow regime types range from 22.6 to 38.5, with an average of 32.4. The average displacement is less than 1/5 of the average within-class distance, also indicating the errors of encoder projection under an acceptable level. Furthermore, the average displacement without outliers (> 95 percentile) is 4.36 for testing set.

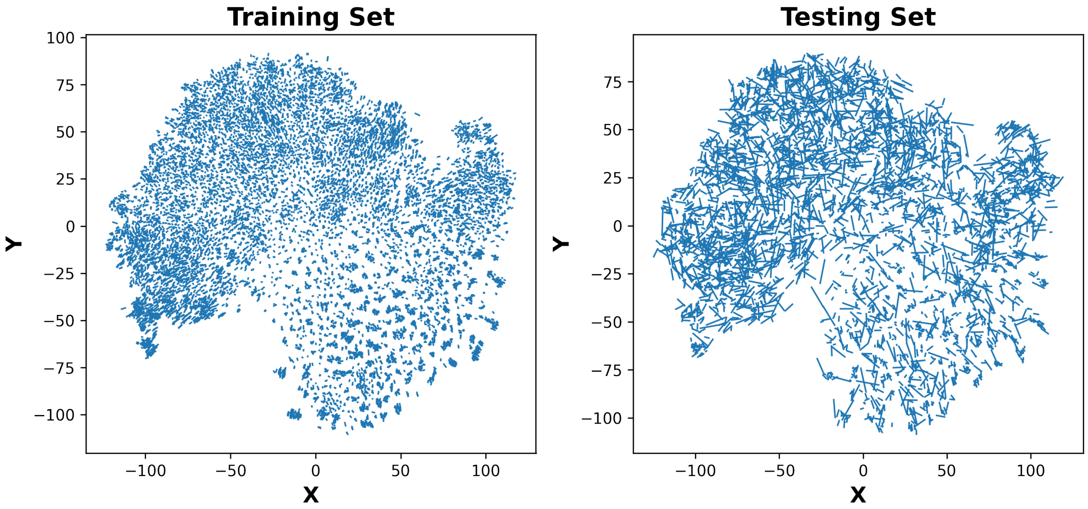


FIGURE 2.8: Lines on map indicate the displacement between original t-SNE points and their counterpart projected by encoder. The displacements for training set are very small, while those for testing set are relatively large.

To illustrate the potential use of an encoder and the t-SNE map for classification, an ADH with an unknown flow regime type was randomly selected and projected onto the t-SNE map using the optimal encoder (Fig. 2.9). Its ten nearest neighbours were identified and ADHs plotted with the unknown ADH (Fig. 2.9). They are found sharing a great similarity in the shape of ADH. The NSE between the target ADH and its nearest neighbours are higher than 0.9. This ADH is projected into a zone, where many datapoints were labeled with Class 6. Based on this procedure, the ADH is presumed to belong in Class 6; a flow regime that resembles the snow-dominated regime of the Canadian Rockies. The implications of this procedure are discussed below.

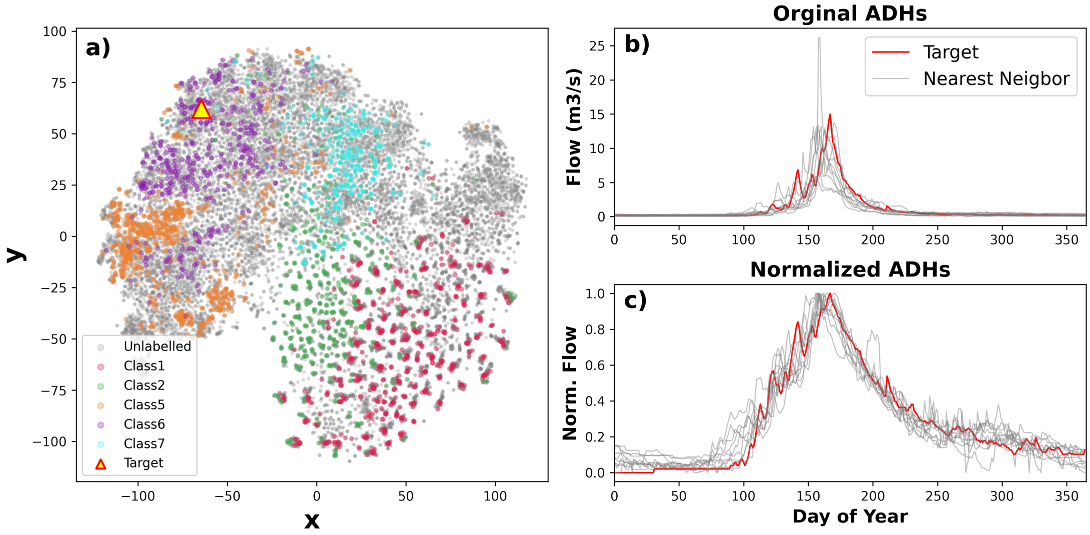


FIGURE 2.9: A randomly selected ADH on t-SNE map projected by Encoder (a). Original (b) and normalized ADHs (c) of ten nearest neighbors.

2.4 Discussion

The t-SNE map provides an novel and intuitive way to visualize similarities for large sets of stream hydrographs, as ADHs with similar shapes remain close on the 2D map. Here, we propose that the distance between points on the 2D map can be used as a similarity metric among watersheds, or for a given watershed the space it occupies on the map can be used to infer its relative variability in flows. In terms of grouping watersheds, t-SNE is superior to PCA with respect to separability of ADHs from different flow regimes. Furthermore, t-SNE is particularly suited to large data sets, requires less computational power, and is more interpretable than conventional visualization tools (i.e. pairwise similarity matrix). While we have only used data from western North America, t-SNE can be more broadly applied to larger or more constrained data sets.

If new hydrographs are obtained, ADHs can be projected on the t-SNE map with the trained encoder, allowing hydrographs to be quickly associated with like counterparts and information, such as flow regime type, seasonal pattern, and dominant hydrological processes can be estimated according to its nearest neighbors on t-SNE map. In this way, this approach can be used for comparative analysis, flow regime classification and regionalization and potentially for change detection.

2.4.1 A Novel Similarity Metric

To confirm the validity of t-SNE distance as a metric of similarity, the pairwise t-SNE distances between ADHs on the 2-D map are compared with the counterparts of cross correlations (Xcorr), which is a conventional metric used to measure similarity between time series. Xcorr is calculated for every pair of ADHs. For every single ADH, the average of Xcorr for all other ADHs is 0.20, while that for its 10 nearest neighbor is 0.87. It manifest that ADHs close on t-SNE map indeed share considerable similarity of flow regime pattern. Regression analysis between t-SNE distance and Xcorr indicates a significant relationship (p-value < 0.01) with a Spearman R of -0.80. Nash-Sutcliffe Efficiency (NSE) is another widely-used metric that measure the consistency between hydrographs. The relationship between t-SNE distance and NSE is also statistically significant and with Spearman R of -0.75.

Cross Correlation Matrix (CCM) is an alternative method to visualize pairwise similarity between hydrographs, yet is impractical for large datasets. Our dataset produces a 17110-by-17110 CCM, with more than 140 million entries (after removing twins), making it computationally expensive to generate and difficult to recognize and interpret patterns at this scale. On the other hand, the t-SNE map provides an intuitive and efficient way to visualize similarity of ADHs, which is simply indicated by their distance on the map.

2.4.2 Misclassification of Mixed Regimes

Misclassification is a ubiquitous feature of machine learning algorithms, and Class 2 had the highest misclassification rate in our procedure. This is unsurprising as Class 2 is a mixed regime with both snowmelt driven freshet and high flows in winter, is a relative superposition of Class 7 and Class 1, and lies between them on the 2-D map. Depending upon winter temperatures, the ADHs in Class 2 tend to shift towards Class 1 in warm years and Class 7 in cold years, which is expected as warm anomalies bring rain (which prevails in the PNW) and cold anomalies enhance snow accumulation and melt (such as in the Canadian Rockies). We presume that watersheds that are more sensitive to climate anomalies are also more subject to misclassification.

2.4.3 Labeling Strategy

The performance of the t-SNE map is based on the separability of ADHs among different flow regimes, and the accuracy of the KNN classification provides a quantitative measure of this separability. Data quality of labelled ADHs is critical as it directly links to classification accuracy. While our strategy allowed a large number of samples to be labelled in a practical time frame, there is the potential for mislabeling ADHs for atypical years. The general nature of hydrological variability can impart a large variance on ADHs, and that in some years hydrographs have flow patterns that are unrepresentative of their labelled class. This inherent variability also explains why some samples from each class are apart from the majority on the t-SNE map.

Manually labelling ADHs is time consuming and there is considerable subjectivity and process knowledge that is required for meaningful classification. To remove subjectivity, Generative Adversarial Networks (GAN) have been used to artificially generate samples that closely mimic real ones (Goodfellow et al., 2014). GAN samples preserve the main patterns of the training data and introduce some random variation, and have considerable potential to create high-quality labelled samples of ADHs. Fed with a limited number of ADHs, GAN can generate an infinite number of samples for each flow regime, enlarging the sample size and reducing the influence of human subjectivity. While not used in this work, we suggest GAN is a promising tool for improving classification.

2.4.4 Optimal Encoder Selection

An effective encoder is a critical component of this approach as it determines the reliability of the mapping function between ADHs and the t-SNE data points and allows insertion of new data onto the t-SNE map. In this work, we tested 55 encoder models with various network architectures and activation functions (see Table A1.1) before selecting the optimal encoder (i.e. with minimum MAE) for ADH dataset. Here, we demonstrate the sensitivity of encoders to network depth and activation function type.

We found even for models with exactly the same configuration (i.e. hyperparameters), they often ended up with different MAE loss during training procedure, which is likely attributed to the randomness of initial state of dense layers. Therefore, for each model configuration, 30 identical models were built and trained independently. The median of their final MAE losses was used to represent the model performance, and the variance was indicated by the error bars in Fig. 2.10.

In many deep learning applications, deeper networks prevail due to their strong capability in recognizing and processing complicated patterns of data (Goodfellow et al., 2016). However, in our case, the encoder's performance did not consistently improve with the depth of network. A clear reduction in MAE was observed when increasing the number of layers from one to four, but further deepening of the layers show very slow improvement (Fig. 2.10). The relative few number of layers was likely due to the relative simplicity of ADHs compared with photographic images, which are the most common application of DL research.

Another method to enhance encoder performance was to widen the layers of the network, as increasing the nodes of the initial layer from 128 to 1024 provided marked improvement in MAE (Fig. 2.10). However, further expansion to 2048 nodes only marginally improved encoder's performance, suggesting wider layers are unnecessary. In the trials, we found the nine-layer network with an initial layer of 2048 nodes (N2048-L9, see Encoder 39 in Table A1.1) produced the lowest MAE of testing set, so this architecture is employed by the final encoder.

Using LeakyReLU as the activation function instead of ordinary ReLU substantially improved the encoder's performance. Decreasing the values of a from 0.4 to 0.02, we found the MAE of testing set continuously declined and reached to the minimum at $a = 0.07$, but with a slight increase below that (Fig. 2.10). With an appropriate a, LeakyReLU reduced the MAE by 10%. Our final encoder employed a LeakyReLU with $a = 0.07$ as the activation function.

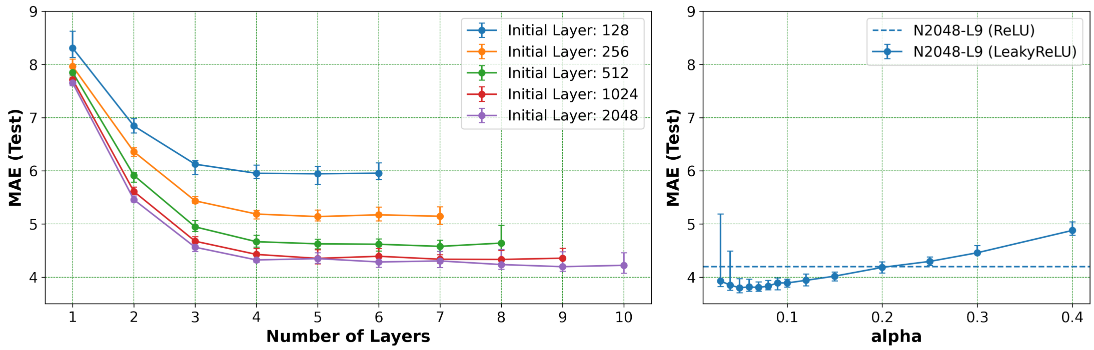


Figure 2.10: Changes of encoders performance with width and depth of network (left) and alpha of LeakyReLU activation function (right).

Gaps of loss between the training and testing set are very common in supervised learning, as models always tend to perform less reliable on "unseen" data. However, huge gaps are often considered a sign of overfitting. In this case, there is a relatively large gap between the MAE of training and testing set, so a Dropout layer (Srivastava et al., 2014) was inserted behind the top Dense layer in selected the network in order to improve model generalization. During training, Dropout deactivate a number of randomly selected nodes (neurons) at each iteration. As a result, it prevents models from depending exclusively on certain neurons and therefore improve generalization. In this case, we found that Dropout effectively narrowed the gap of MAE between training and testing set. However, the reduction is mostly led by increasing loss of training set, while the loss of testing set largely remain the same. This indicated that model generalization had little improvement. Presumably, a more effective way to enhanced encoder generalization is to increase diversity and sample size of ADHs datasets.

2.4.5 Potential use of t-SNE maps

As a clustering technique, t-SNE is a powerful technique that has had very limited application to hydrological data sets (Mazher, 2020; Liu et al., 2021), yet there are many potential uses as 'big data' and machine learning emerge in hydrology. First, the t-SNE map provides an intuitive way to visualize the similarity of ADHs and other large time series data sets, and the encoder improves its practical use by turning the t-SNE from a non-parametric to a parametric method. It overcomes a major challenge of t-SNE by allowing new data to be projected on an existing map as it becomes available. The approach of information visualization can benefit a broad range of research that seeks to establish similarity among time series such and need not be restricted to streamflow hydrographs. For example, watershed classification, regionalization and streamflow change detection are all potential future areas of research using this methodology. t-SNE can also be used to infer similarity in other time series and signals to assess natural groupings and patterns.

Perhaps the most obvious use of t-SNE is in watershed classification as the parametric t-SNE technique can effectively identify natural groupings of watersheds and place new data within these groupings. While there is considerable research over the past few decades on hydrological classification (Wagener et al., 2007), t-SNE is a novel data-driven approach that allows fast identification of homogeneous and similar regimes, and process inference can be quickly transferred to their nearest neighbours. If the location is close enough on the t-SNE map, it would be possible to reconstruct hydrographs of ungauged watersheds according to their nearest neighbours (Patil and Stieglitz, 2012). Another potential utilization of t-SNE is identifying information redundancy, which is important in designing hydrometric and other monitoring networks (Coulibaly et al., 2013). For example, a number of highly homogeneous watersheds on the t-SNE map from the PNW suggest high consistency in the hydrographs and potential information redundancy in the network. However, the validity of information redundancy detected by t-SNE should be further verified with entropy-based approaches (Singh, 1997; Mishra and Coulibaly, 2009).

2.5 Conclusion

Here, we demonstrate the potential of t-SNE as an approach to compare similarity among hydrographs that is particularly useful for large data sets. By converting annual daily hydrographs to 2-D representation, their degree of similarity is indicated by distance on the map. t-SNE distance can be used as novel similarity metric, supplementing other comparative metrics such as Xcorr and Nash-Sutcliffe efficiency. In this application on reference watersheds in western North America, t-SNE outperformed PCA analysis for dimensional reduction, suggesting its potential in classifying and regionalizing streamflows. A deep learning encoder network was developed and trained to project new data onto existing maps to identify hydrological counterparts, overcoming previous challenges of the non-parametric t-SNE approach. While subjectivity in classification limits deep learning algorithms in some circumstances, we believe that t-SNE is underutilized in hydrological applications and that it has considerable potential for extremely large data sets.

It provides an intuitive way to compare similarity among hydrographs that is particularly useful for "big data" of streamflows. Converting ADHs to 2-D representation, t-SNE map allows hydrologists to visualize the similarity of a large number of ADHs, and the degree of similarity is indicated by their distance on the map. t-SNE distance could serve as novel similarity metric for hydrographs, as supplementary to Xcorr and NSE. t-SNE is considered outperforming the linear dimensionality reduction technique (i.e. PCA), as samples from different flow regime types become more separable on t-SNE map than on PCA map. Thus, t-SNE map could facilitate classifying and regionalizing streamflows.

It allows to project new hydrographs of interest to an existing map, so hydrologists could quickly identify hydrologically homogeneous counterparts from archive databases. It facilitates the process of transferring knowledge and understanding obtained from well-studied sites to new sites, and therefore increase the values of existing datasets and historical research to hydrological systems. We believe it could benefit some PUB programs.

In this case, t-SNE map reveals information redundancy in current hydrometric network in coastal Pacific Northwest region. Thus, it is considered a useful tool to evaluate the efficiency of hydrometric network. We hope this paper could raise the attention of hydrological community to t-SNE, so colleagues may find more possible applications for this relatively new technique in hydrological sciences.

Chapter 3

Identifying the streamflow changes in western North America from 1979 to 2021 - Part 1: Insights from hydrometric trend analysis

3.1 Introduction

Changes in the flow regimes of rivers across the entirety of Western North America (WNA) have been widely reported and are expected to accelerate as a result of climate change (Groisman et al., 2000; DeBeer et al., 2016; Williams et al., 2020). For example, transitions from snowmelt towards rainfall-driven regimes are not uncommon (Burn et al., 2010; Burn and Whitfield, 2017), yet as WNA crosses several hydroclimate zones, detecting and quantifying trends in streamflow and attributing them to specific climate and other factors remains a challenge (Dethier et al., 2020). In addition, identifying integrated and specific metrics of change is complicated by the dozens of widely used indicators that describe flow regimes (Poff et al., 1997), all of which are subject to change with time. Most common analyses of change focus on annual peaks or a small number of specified metrics (e.g. low flows, timing of peak, etc). Finally, studies have taken a range of approaches to rationalize data to enhance statistical performance based on political, geographical and statistical clustering based on flow characteristics (Hall and Minns, 1999; McDonnell and Woods, 2004).

Across WNA, from the southern mountains of New Mexico to the Arctic coast, the processes that generate annual and seasonal flows are distinct, and the influence of low-frequency climate oscillations and shifts on streamflow varies among regions and across scales (Whitfield et al., 2010). A complex picture emerges attempting to describe and synthesize change over WNA, yet throughout this domain, there have been numerous studies that have described changes and ascribed their cause. For example, in the subarctic and arctic, increasing winter low flows are widely reported while other observed changes include increasing mean annual flow, earlier spring freshet, and enhanced inter-annual variability (Zhang et al., 2001; Burn et al., 2004; Dery et al., 2009b; Burn et al., 2010). Some of the most widely observed trends are increasing winter (Jan-Mar) baseflows and related low flow hydrometrics. St. Jacques and Sauchyn (2009) detected upward winter flow trends in the Northwest Territories for the 1978-2007 period and (Walvoord and Striegl, 2007) suggested that groundwater contribution to streamflow on average increased by 0.9% per year across the Yukon River Basin for the period 1949-2005. Increases in low flows are typically attributed to thawing permafrost, which enhances percolation and supports deeper flow paths that can contribute greater groundwater fluxes throughout the year (Walvoord and Striegl, 2007). Other patterns for many northern streams include increased mean annual flows and earlier onset of spring freshet (St. Jacques and Sauchyn, 2009; Burn et al., 2004). An intensifying hydrological cycle was used to explain enhanced inter-annual variability in streamflow from 1964-2007 by Dery et al. (2009b).

Earlier onset of spring freshet is another feature widely reported throughout WNA, particularly for streams in the Canadian Rocky Mountains (Zhang et al., 2001; Stewart et al., 2005; Dery et al., 2009b; Burn et al., 2010). Examples include significant increase in spring flow during 1967-1997 in southern BC, whereas late summer and autumn flows significantly declined (Zhang et al., 2001). Similarly, Burn and Hag Elnur (2002) reported significant decreasing trends in the annual maximum flow at majority of streams in southern BC during 1957-1997. Barhney et al. (2017) also reported declines in August flow at majority of glacierized streams in Canadian Rocky Mountains from 1975-2012, which was considered a consequence of reducing contribution from glacier and perennial snow patches. In contrast, Hernandez-Henriquez et al. (2017) indicated positive detectable trends over recent time periods, and for 1986-2015 showed fall increases for glacierized rivers around Vancouver Island and southern coastal BC. In recent work, Moore et al. (2020) suggest wastage flux contributions from net glacial mass loss recently passed peak water, with declines in August flows expected to continue the coming decades in terms of magnitude and significance. Further south in the north central Rocky Mountains of the United States (i.e. Idaho, Montana, and Wyoming), Leppi et al. (2012) also reported declining August flow, where 89% of the streams exhibited declining slopes from 1950-2008. They considered that the warming air temperature along with intensifying evapotranspiration during spring and summer were the primary driver for declining late summer flow. Declining patterns of streamflow were detected at most gauging stations in southern Alberta over the last century (St. Jacques et al., 2010). Rood et al. (2005) reported the streamflow at 35 headwater catchments in Rocky Mountains (on both Canada and US side) declined an average of 20% from 1910 to 2002.

Declines in summer flow, particularly late summer, were also observed in the Pacific North West (PNW). Kormos et al. (2016) showed that August and September flow declined at majority of PNW stream gauges, with an average decline of 22% over a 65-year period from 1948-2013. In addition, 7-day minimum flow during summer declined an average of 27%. Significant negative trends in September flow was also reported by Chang et al. (2012) for most streams in the PNW. Furthermore, Luce and Holden (2009) found 25th percentile annual flow significantly declined at most PNW streams, and nearly half of the streams dropped by more than 29% from 1948-2006. In contrast, March flow significantly increased in the majority of PNW streams (Chang et al., 2012). Stewart et al. (2005) reported remarkable increase in monthly fractional flows in March and April in western US, and Kormos et al. (2016) observed that the centroid of flow mass had advanced an average of 7.8 days over 65 years as flows were redistributed between summer and spring.

The Pacific Decadal Oscillation (PDO) strongly influences streamflow in WNA through its influence on air temperature and precipitation (Woo and Thorne, 2003; Gobena and Gan, 2006; St. Jacques et al., 2014; Bawden et al., 2015). Under the negative PDO phase, catchments in WNA tend to be cooler and wetter, while positive PDO phase leads to warmer and drier conditions. Historically, PDO phase switched every 20-35 years and four complete phases were observed in the last century 1900-1924 (cool), 1925-1945 (warm), 1946-1976 (cool), and 1977-1997 (warm). However, this cycle was broken in last two decades, when phase shifting of PDO became much more frequent: 1998-2002 (cool), 2003-2007 (warm), 2008-2013 (cool), and 2014-present (warm) (Li et al., 2020). Warm-phase PDO is considered an important driver for changes in streamflow in WNA over last three decades, such as increased winter flow in Arctic (Hodgkins, 2009), and earlier snowmelt and decreased streamflow in northwestern North America (Stewart et al., 2005; Rood et al., 2005; Dery et al., 2009a; St. Jacques et al., 2010). Other large scale pressure features such as the El Nino-Southern Oscillation (ENSO) and Pacific-North America (PNA) pattern influence flows as El Nino and positive PNA often leads to a warm, dry climate and lower streamflow, while La Nina and negative phase PNA typically to create a cool and wetter condition with higher streamflow in WNA (Woo and Thorne, 2003; Gobena and Gan, 2006).

Changing streamflow regimes in WNA have profound socioeconomic and ecological consequences. For example, hydropower operation in the Colorado river basin has declined remarkably as severe aridity during the last several decades substantially diminished water storage in reservoirs (Sabo et al., 2010), and in 2021, the United States declared an emergency water shortage for the Colorado River for the first time. Reduced water supply from rivers has already harmed agricultural and relevant industries for southwestern states, leading to multi billion-dollar losses every year (MacDonald, 2010). Water stress is considered the primary challenge to socioeconomic and environmental sustainability in southwestern USA in 21st century (MacDonald, 2010). In addition to agriculture, shifts in flow regime also adversely influence fisheries at streams in PNW and Rocky Mountains. Increasing magnitude of peak flows during spawning and incubation period has led to high mortality rate of salmon (Battin et al., 2007). This is in addition to higher stream temperature during summer, which was jointly driven by elevated air temperature and less cold water melting from glaciers, threatens endangered populations of many native fish species (Leppi et al., 2012). In contrast, Arctic rivers which are expanding their hydroelectric potential, the possibility of increased autumn and winter flows could promote electricity generation (Cherry et al., 2017). While these are just a few select examples, changes to river flow regimes have broad impacts, and quantifying and understanding the drivers of change is critical for future design, management and adaptation.

The objective of this study is to provide a comprehensive overview of streamflow changes across WNA domain during the latest decades from 1979 to 2021. Trend analysis were performed on a selection of hydrometrics aiming to examine changes from various aspects. This study focused on not only hydrometric trends at individual streams but also regional-scale flow regime shifting. Trending patterns of hydrometrics were synthesized into six hydro-climate regions, where streams exhibit distinctive flow regimes. Detected changes in streamflows were associated with regional changing climates. Comparative analysis was conducted between glacial and non-glacial watersheds for three relatively cold hydro-climate regions, in order to identify the impact of glaciers on streamflows response to climate changes. This information update and advance previous work by focusing on the most recently available data, by examining a wider range of flow regime parameters than are typically evaluated, and by providing an easily navigable dashboard to explore patterns and trends in this large data set.

3.2 Methods

3.2.1 Data

Daily flow data were collected for stream gauges in the Referenced Hydrometric Basin Network (RHBN) and Hydro-Climate Data Network (HCDN), which are operated by Water Survey of Canada (WSC) and United States Geological Survey (USGS), respectively. Both networks select streams with a natural flow regime with minimal land use change and are intended for long-term hydrometric analysis. Stream gauging stations in RHBN and HCDN are widely used in previous studies regarding changes in streamflow and relevant hydrological processes (Zhang et al., 2001; Burn and Hag Elnur, 2002; Stewart et al., 2005; Burn and Whitfield, 2017).

We define WNA to include four Canadian provinces/territories (British Columbia, Alberta, Yukon, and Northwest Territory) and eleven American states (Washington, Oregon, California, Idaho, Nevada, Montana, Utah, Wyoming, Colorado, Arizona and New Mexico). Long-term flow data were split into Annual Daily Hydrographs (ADHs). Each ADH includes 365 values of daily flow, from 1 January to 31 December (leap days excluded if applicable). Strict criteria were applied to select high-quality, complete ADHs. ADHs with gaps larger than 7 days were excluded. Gaps smaller than or equal to 7 days were filled using linear interpolation. This study focuses on the 43-year period from 1979 to 2021 and only stations having at least 30 valid ADHs during this period were included. In total, 126 RHBN and 430 HCDN stream gauging stations were included in this study (see Table A2.1 in Appendix B).

Monthly air temperature and precipitation were obtained from the ERA5 climate reanalysis product (Hersbach et al., 2020) from 1979 to 2021 with complete spatial and temporal coverage of the WNA domain at 0.1°. We conducted trend analysis on this data across the WNA domain over the study period (see Appendix B) to identify possible drivers of changes in streamflow.

Watersheds containing glaciers were identified using data of global glacier coverage from the Global Land Ice Measurements from Space data (GLIMS, `http://glims.colorado.edu/glacierdata`). The area of glacier coverage in each watershed was obtained by clipping glacier coverage with the watershed boundary. We considered watersheds glaciated when glacier area was $> 0.5\%$ of the total basin area; and in total 53 watersheds were identified throughout the domain.

3.2.2 Hydrometric Calculation

TABLE 3.1: Description of selected hydrometrics

Symbol	Name	Description	Category
Q_{mean}	Mean	mean of annual daily flows	Magnitude
Q_{median}	Median	median of annual daily flows	Magnitude
Q_{std}	STD	standard deviation of annual daily flows	Others
Q_{range}	Range	range of annual daily flows	Others
Q_{skew}	Skewness	skewness of annual daily flows	Others
Q_{10p}	10th Percentile	10th percentile of annual daily flows	Magnitude
Q_{25p}	25th Percentile	25th percentile of annual daily flows	Magnitude
Q_{75p}	75th Percentile	75th percentile of annual daily flows	Magnitude
Q_{90p}	90th Percentile	90th percentile of annual daily flows	Magnitude
Q_{min7d}	7d Min Flow	minimum of 7-day average flow	Magnitude

Continued on next page

TABLE 3.1: Description of selected hydrometrics

Symbol	Name	Description	Category
Q_{max}	Max Flow	maximum of annual daily flows	Magnitude
N_{high}	# High Flow Events	number of high flow events	Event
D_{high}	Duration of High Flow	duration of high flow	Event
N_{low}	# Low Flow Events	number of low flow events	Event
D_{low}	Duration of Low Flow	duration of low flow	Event
T_{cen}	Day of Centroid	day of centroid of flow mass	Timing
T_{25p}	Day of 25th Perc	day of cumulative flow exceeding 25% of annual total flow	Timing
T_{50p}	Day of 50th Perc	day of cumulative flow exceeding 50% of annual total flow	Timing
T_{75p}	Day of 75th Perc	day of cumulative flow exceeding 75% of annual total flow	Timing
T_{sprmax}	Day of Spring Freshet Peak	day of maximum of 14-day average flow in spring window (from Feb 1st to May 31st)	Timing
Q_{jan}	January	January flow	Monthly
Q_{feb}	February	February flow	Monthly
Q_{mar}	March	March flow	Monthly
Q_{apr}	April	April flow	Monthly
Q_{may}	May	May flow	Monthly
Q_{jun}	June	June flow	Monthly
Q_{jul}	July	July flow	Monthly
Q_{aug}	August	August flow	Monthly
Q_{sep}	September	September flow	Monthly
Q_{oct}	October	October flow	Monthly
Q_{nov}	November	November flow	Monthly
Q_{dec}	December	December flow	Monthly
F_{jan}	Jan (%)	January fractional flow	Monthly
F_{feb}	Feb (%)	February fractional flow	Monthly
F_{mar}	Mar (%)	March fractional flow	Monthly
F_{apr}	Apr (%)	April fractional flow	Monthly
F_{may}	May (%)	May fractional flow	Monthly
F_{jun}	Jun (%)	June fractional flow	Monthly
F_{jul}	Jul (%)	July fractional flow	Monthly
F_{aug}	Aug (%)	August fractional flow	Monthly
F_{sep}	Sep (%)	September fractional flow	Monthly
F_{oct}	Oct (%)	October fractional flow	Monthly
F_{nov}	Nov (%)	November fractional flow	Monthly
F_{dec}	Dec (%)	December fractional flow	Monthly

Continued on next page

TABLE 3.1: Description of selected hydrometrics

Symbol	Name	Description	Category
Q_{spr}	Spring Flow	sum of monthly flow from March to May	Seasonal
F_{spr}	Spring Fraction Flow	ratio of spring flow to annual total flow	Seasonal
Q_{smr}	Summer Flow	sum of monthly flow from June to August	Seasonal
F_{smr}	Summer Fraction Flow	ratio of summer flow to annual total flow	Seasonal
Q_{aut}	Autumn Flow	sum of monthly flow from September to November	Seasonal
F_{aut}	Autumn Fractional Flow	ratio of autumn flow to annual total flow	Seasonal
Q_{win}	Winter Flow	sum of monthly flow from December to February	Seasonal
F_{win}	Winter Fractional Flow	ratio of winter flow to annual total flow	Seasonal
Q_{lsmr}	Late Summer Flow	sum of monhtly flow from July to September	Seasonal
F_{lsmr}	Late Summer Fractional Flow	ratio of late summer flow to annual total flow	Seasonal
SI	Seasonal Index	seasonal index (Coopersmith et al., 2014)	Others

55 hydrometric indices were applied to characterize the ADHs. We classify hydrometrics into six groups: 1) magnitude, 2) timing, 3) event, 4) monthly, 5) seasonal, and 6) others (see Table 3.1). For N_{high} and D_{high}, high flow events share a similar definition with Peak-Over-Threshold events proposed by Burn and Whitfield (2017). A threshold for high flow here is taken as the 75th percentile of the long-term hydrograph (rather than ADHs). Segments above the threshold were taken as high flow events (Fig.3.1). N_{high} is the count of over-threshold segments in an ADH, and D_{high} is the sum of length of the over-threshold segments. Likewise, low flow events are the under-threshold segments in a hydrograph, and the threshold for low flow is the 25th percentile of the long-term hydrograph at a site. N_{low} and D_{low} indicate the number of and the total duration of the low flow events in an ADH.

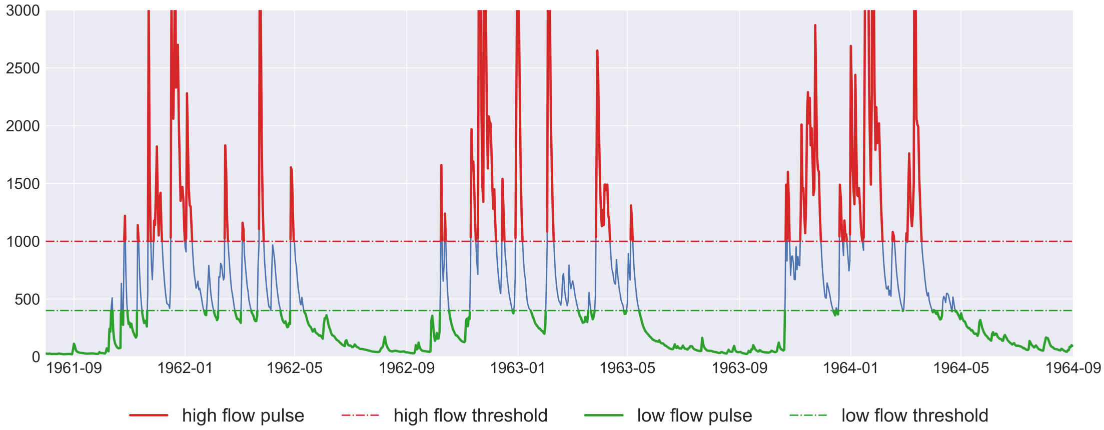


FIGURE 3.1: Example of high and low flow pulses.

Monthly hydrometrics applied include the monthly mean flows and monthly fractional flows. Monthly fractional flow is the percentage of monthly flow to annual total flow. Monthly flow of several continuous months can be combined together to form a seasonal flow, yet this varies by region. Seasonality Index (SI) proposed by Coopersmith et al. (2014) was adopted to examine the changes in seasonality and intra-annual variability of streamflow.

3.2.3 Trend Analysis

Mann-Kendall (MK) test is one of the most prevailing, if not the most prevailing tool for trend analysis in meteorology and hydrology (e.g. Zhang et al., 2001; Burn and Hag Elnur, 2002; St. Jacques et al., 2014) for detecting monotonic linear trends from long-term time series. As a non-parametric approach, MK has no requirement for data distribution. MK score (S) compares the value at one time step with a later one ($j > i$, see Eq.3.1). A positive MK score indicates values are generally increasing trend, while a negative score indicates time series values decreasing over time. The Z-score determines the significance level of a trend (Eq.3.4).

$$S = sgn(x_j - x_i) = \begin{cases} 1, & \text{if } x_j > x_i \\ 0, & \text{if } x_j = x_i \\ -1, & \text{if } x_j < xi \end{cases} \tag{3.1}$$

$$E[S] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i) \tag{3.2}$$

$$Var[S] = \frac{1}{18}[n(n-1)(2n+5) - \sum_{k=1}^{p} q_k(q_k-1)(2q_k+5)] \tag{3.3}$$

$$Z_{MK} = \begin{cases} \frac{E[S]-1}{\sqrt{Var(S)}}, & E[S] > 0 \\ 0, & E[S] = 0 \\ \frac{E[S]+1}{\sqrt{Var(S)}}, & E[S] < 0 \end{cases} \tag{3.4}$$

where $E[S]$ and $Var[S]$ is the mean and variation of MK scores. In Eq.3.3, p is the total number of tie groups in the data, and q_k is the number of data points in the k-th tie group. Null hypothesis (i.e. no monotonic trend) is rejected if $|Z_{MK}| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is $100 \cdot (1-\alpha/2)$-th percentile of standard normal distribution and α is a user-defined threshold of p-value. Here, α was chosen as 0.1. In addition to significance level, slope of trend line was calculated using Sen's slope estimator (Eq. 3.5), which is also known as Theil-Sen estimator (Sen, 1968).

$$Sen's\ slope = Med(\frac{x_j - x_i}{j-i}),\ j > i \tag{3.5}$$

where x_i and x_j denote the value at time step i and j respectively. Magnitude of changes in hydrometrics are measured by net change and the rate of change. Net change is the absolute difference between the start and end point on the trend line (i.e. Sen's slope). The rate of change is the net change divided by the long-term average of the hydrometric.

Serial correlation increases the chance of detecting significant trend from a time series (Hamed and Rao, 1998) and prewhitening time series is typically recommended before implementing MK test (Burn and Hag Elnur, 2002). In this case, Trend-Free PreWhitening (TFPW; Yue et al., 2002) was adopted in this case, which consists of three steps: 1) removal of Sen's slope from the original time series, 2) removal of lag-1 autocorrelation from the detrended time series, and 3) adding back Sen's slope to the prewhitened time series (see Eq. 3.6-3.8).

$$x_t^{'} = x_t - \beta \cdot t \tag{3.6}$$

$$y_t^{'} = x_t^{'} - \rho \cdot x_{t-1}^{'} \tag{3.7}$$

$$y_t = y_t^{'} + \beta \cdot t \tag{3.8}$$

where x_t, $x_t^{'}$, $y_t^{'}$, and y_t represent the original, detrend, prewhiten, and TFPW times series, respectively. β is the Sen's slope; ρ is the estimated lag-1 autocorrelation coefficient; t denotes the time step. More details regarding TFPW can be found in the paper of Yue et al. (2002). Functions regarding MK test, Sen's slope, and TFPW are all available in the Python package of *pymannkedall* (Hussain and Mahmud, 2019).

3.2.4 Regional Analysis

Streams in WNA domain were assigned into six hydro-climate zones based on geographic location and shape of hydrographs: 1) Northwest, 2) Canadian Rocky Mountains (CRM), 3) US Rocky Mountains (USRM), 4) Coastal Pacific North West (CPNW), 5) Coastal Pacific Middle West (CPMW), and 6) Southwest. We acknowledge that this grouping has limitations as it is not based on objective criteria (e.g. Dethier et al., 2020), yet classifying and grouping watersheds based on geographical boundaries, hydrometric indicators, climate or other geophysical parameters are also subjective and beyond the scope of this change detection work. The grouping of streams benefits from t-SNE map developed in Chapter 2 that greatly facilitate finding and identifying streams with similar flow patterns. Each region contains at least 50 stream gauging stations (Fig.3.2). Field significance of changes in a hydrometric indicator within these regions was assessed using a bootstrap procedure with 10000 iterations. The time series of the gauging stations in a hydro-climate zone were permuted iteratively. In each iteration, we run MK test with the permuted time series, and counted the number of significant trends (N_{sig}). After 10000 iterations, it formed a distribution of N_{sig}. The critical value related to the regional significance level ($p < 0.1$) can be estimated as 90th percentile of the distribution. If the actual number of significant trends is higher than the critical value, the hydrometric changes in the region are considered to be field significant. More details regarding this bootstrap method can be found in the paper of Renard et al. (2008).

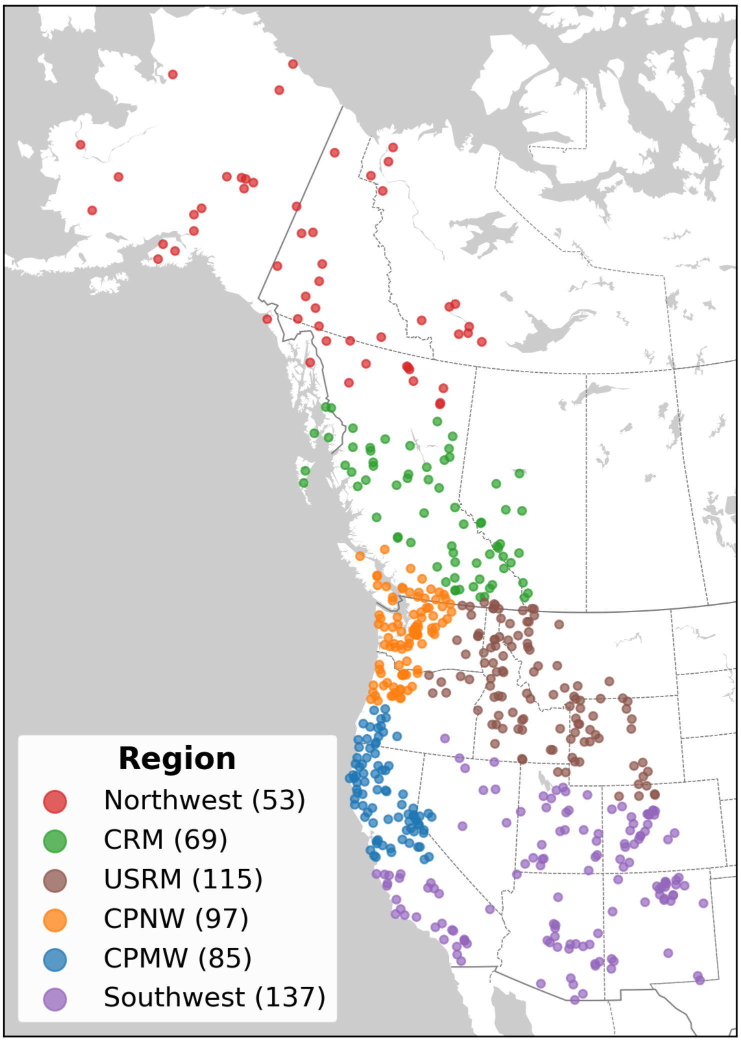


FIGURE 3.2: Stream gauging stations in six hydroclimate regions. The numbers in the legend indicate how many streams the corresponding region includes.

3.3 Result and Discussion

Across the WNA domain, there is considerable spatial heterogeneity in changing hydrometric indicators with regionally distinct patterns. Changes for individual streams were evaluated on the basis of direction (positive or negative), slope (i.e. magnitude of change), and significance level of trend line derived from TFMK and Sen's slope. Changes at regional scale were evaluated on the basis of field significance, percentage of positive and negative trends, and average of net change or rate of change for streams within every hydro-climate region. Due to the expansive nature of the analysis, not all hydrometrics and trends are discussed in this paper. However, we provide a companion dashboard application (`https://app-trend-map.herokuapp.com`) that has been developed to allow viewers to inspect all outputs for all sites, including the trend maps for all target hydrometrics along with the MK-test outcomes and other associated significance metrics.

We consider Q_{mean} as the single most important hydrometric representing the average amount of water availability/flow within a watershed. For the entire WNA domain, there is a general divide at approximately the 46th parallel north (near the southern border of Oregon and Montana). On the north side, more than three quarters of streams showed increasing trends in Q_{mean}, while on the south side, most streams ($> 80\%$) exhibited decreasing trends (Fig. 3.3). This general pattern across WNA suggests overall wetting in the north with a more pronounced drying in the south. In the following sections, we outline the major changes across the domain and provide context with regards to regional climate and previous research.

Although trends of most hydrometrics present considerable spatial heterogeneity, declines in late summer flows predominate across the entire WNA domain. Nearly 80% of the WNA streams exhibited negative trends in Q_{lsmr}, with one third statistically significant. On average, Q_{lsmr} declined by 18% from 1979 to 2021 in WNA domain. This drying pattern was stronger in western USA than Canada. The average decline of Q_{lsmr} for the 430 USA streams was 22.6%, while that for the 126 Canadian streams was only 6.6%. Given high demand of water for domestic, industrial or agricultural use, declines in late summer flows likely further challenge water security.

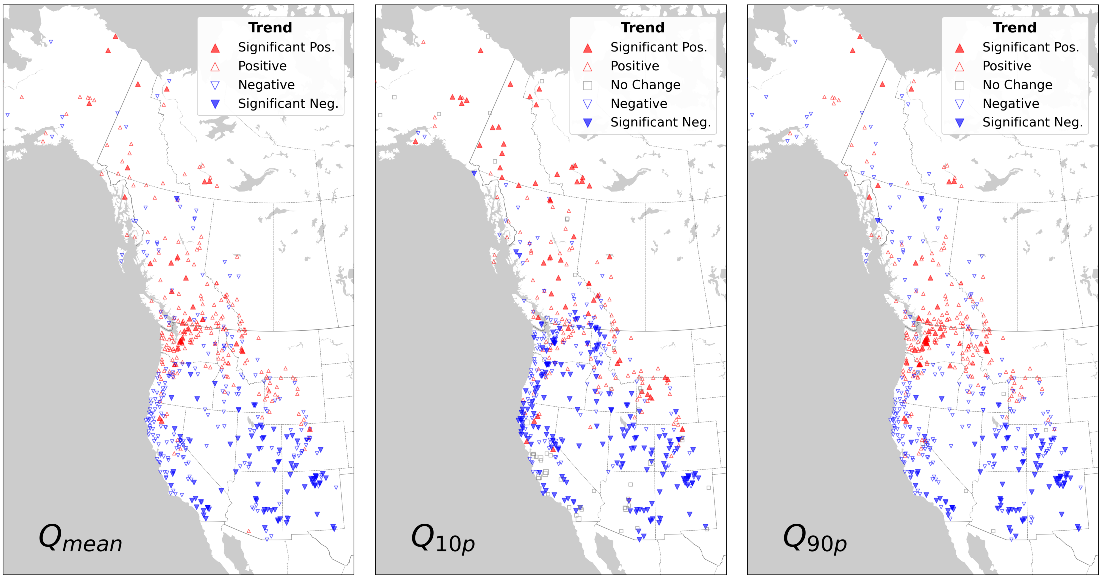


FIGURE 3.3: Map of trends in Q_{mean} (left), Q_{10p} (middle) and Q_{10p} (right). Positive trends are represented by red upward triangles, while negative trends by blue downward triangles. Triangles are solid if the corresponding trend is statistically significant (p-value < 0.1).

3.3.1 Northwest

Streamflows in the Northwest domain exhibited widespread and significant increasing trends over the observation period, and increases were particularly strong for those cold-season low-flow hydrometrics. Significant positive trends were observed for approximately one quarter of the streams for Q_{mean}, yet near half the streams had increases in low-flow hydrometrics, including Q_{10p}, Q_{25p} and Q_{min7d}. These low flows typically occurred in winter; and except for June-August, all monthly flows had a field significant increase that was most pronounced from December-April. Compared with Burn and Hag Elnur (2002) who noted increasing monthly flow from February to May, our analysis indicated a longer period of increase (nine months from September to May), suggesting enhanced streamflow increases over the last four decades. Similar patterns were reported in the paper of Durocher et al. (2019), who observed an increasing streamflow magnitude at a number of WNA rivers flowing to Arctic Oceans that was not detected in records before 2000 (McClelland et al., 2006). St. Jacques and Sauchyn (2009) reported upward trends in winter flow at gauges in the Northwest Territories, and ascribed this increase to thawing permafrost and enhanced subsurface flow and baseflow. Increasing groundwater contribution driven by permafrost thaw was also observed in Yukon River Basin (Walvoord and Striegl, 2007) and interior Alaska (Jones and Rinehart, 2010). Cold-season flows, defined here as the sum of monthly flows from September to May, increased by an average of 24% at the 53 selected streams across Northwest domain from 1979 to 2021. Although much less significant, the increase in Q_{75p} and Q_{90p} were also field significant, which was surprising as increases in high flow metrics have seldom been reported in previous studies for northern regions.

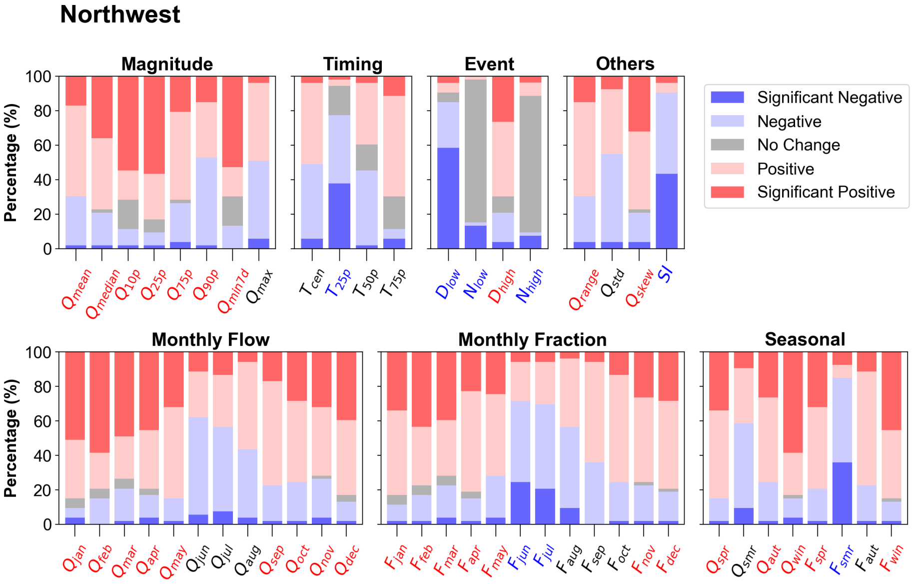


FIGURE 3.4: Percentage of the five trend types at Northwest streams for the hydrometrics. If the change is field significant, hydrometric symbol is coloured with red (positive), blue (negative), or purple (both).

Earlier spring freshet of snow-dominated streams in WNA have been previously reported. Aziz and Burn (2006) found earlier onset of spring freshet at streams across Mackenzie River Basin. However, the advance of spring freshet was relatively weak in Alaska, and some streams even exhibited delayed onset of spring pulse (Stewart et al., 2005). In our analysis, the early onset of thaw and freshet was identified according to the timing of spring peak flow (T_{sprmax}). The majority of streams in Northwest domain showed an advancing trend in T_{sprmax}, and eleven of them were statistically significant, with an average advance of 16 days.

Nearly 90% of the streams in Northwest domain exhibited a negative trend in F_{smr}, half of which were statistically significant. On average, summer flow contribution to annual total flow dropped from 56% to 49% over the observation period. Widespread decrease in F_{smr} suggest that summer flows increased slower than cold-season flows and therefore accounted for a smaller proportion of annual total flow. Increasing volumes of water in low-flow months (i.e. cold season) substantially weakened the seasonality of flow regime at Northwestern rivers, lowering SI across the domain.

14 of 53 streams in the Northwest domain had glacier coverage, and Wilcoxon rank-sum test was applied to assess potential glacier influence on late summer flows (i.e. Q_{jul}, Q_{aug}, Q_{sep} and Q_{lsmr}). Significant differences were observed between glacierized and non-glacierized watersheds in the change rate of Q_{lsmr}, with the unexpected pattern of late summer flows slightly decreased at glacial watersheds while increased at non-glacial watersheds (Fig. 3.5). Streamflow at non-glacial watersheds increased through all the three months, while at glacial ones it mildly declined in July and August.

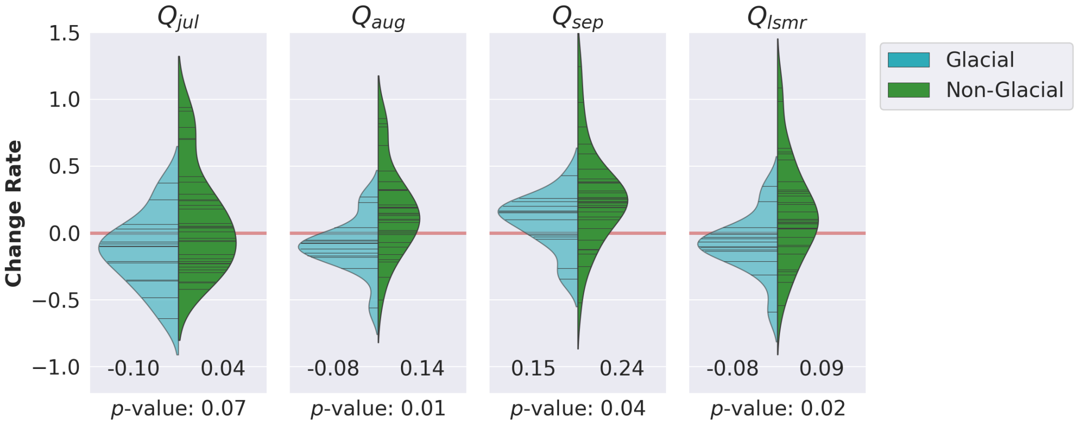


FIGURE 3.5: Comparison of rate of change of late summer flow between glacial and non-glacial streams in Northwest domain. The two numbers inside the plot indicate the median changing rate for the glacial and non-glacial group, respectively.

3.3.2 Coastal Pacific North West (CPNW)

Changes in Q_{mean} and Q_{median} suggested that the CPNW has become a wetter area over last four decades. Most CPNW streams exhibited a positive trend in Q_{mean} and Q_{median}, with an average increase of ∼12% for both hydrometrics, although only a small portion of those trends proved to be statistically significant. High flows increased slightly faster than average flows, and in general Q_{90p} increased 15% over the 43-year period, and the positive trend in Q_{90p} proved to be field significant. Luce and Holden (2009) previously reported declining trends in annual mean, median, 75th-percentile flow in PNW that were not observed in this analysis. However, it is important to note that their analysis covered a longer and earlier period from 1947-2006 and included streams from not only coastal states of Washington and Oregon but interior states of Idaho and Montana, which were assigned to USRM domain in this study.

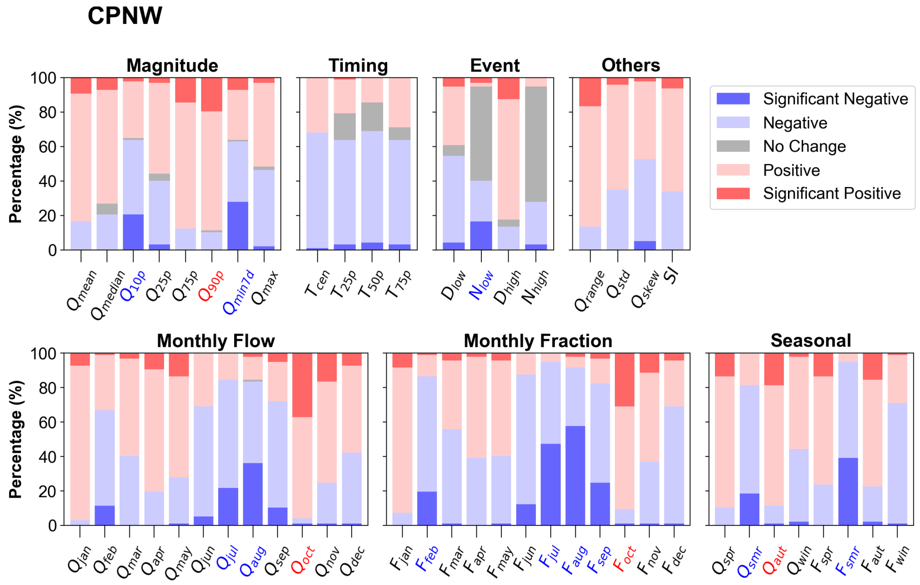


FIGURE 3.6: Percentage of the five trend types at CPNW streams.

In contrast, low flows in CPNW tended to decline. Q_{10p} and Q_{min7d} declined significantly at more than 20% of the streams, and the negative trends proved field significant. As low flows typically occurred during summer, negative trends prevailed across CPNW in terms of summer flows. Q_{jul}, Q_{aug}, and Q_{smr} all decreased an average of $\sim$20% over last 43 years. Kormos et al. (2016) also reported declining Q_{min7d} and Q_{smr} in PNW, and they attributed decreasing low flows to reduced summer precipitation. According to the ERA5 reanalysis product, significant reduction in summer precipitation occurred across the CPNW domain over last four decades (see Fig. A2.2 in Appendix B). This decline was particularly strong in July, and the decreasing rate of the monthly precipitation was up to 1.2 mm per year.

One of the most distinct changes in CPNW was the increase in autumn flows. More than 80% of the streams exhibited a positive trend, and one quarter were statistically significant. On average, Q_{aut} increased 22% over last four decades across CPNW. The increases in Q_{oct} were even more widespread (at 96% of the streams) and greater (rose by 43% from 1979 to 2021). ERA5 data suggest a substantial rise in October precipitation in CPNW during the same period that may in part explain the increased autumn flows.

There are 15 glacial watersheds identified in CPNW. Late summer flows decreased significantly slower at glacial than non-glacial watersheds (Fig. 3.7). The average decrease of Q_{lsmr} for glacial and non-glacial streams were 7% and 17%, respectively. The difference was most significant in July and gradually reduced through September. This result provides some evidence of enhanced for glacial contributions to streamflow in the late summer that is expected from enhance wastage (Moore et al., 2020).

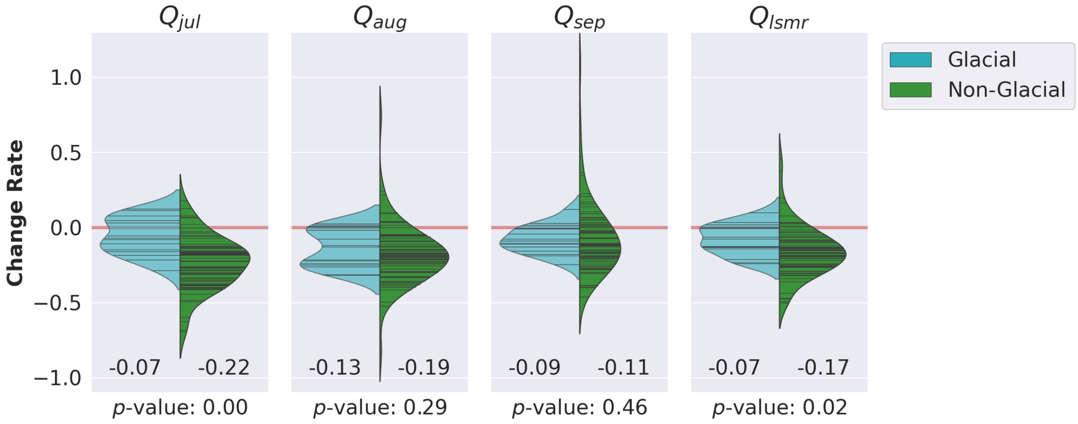


FIGURE 3.7: Comparison of rate of change of late summer flow between glacial and non-glacial streams in CPNW.

3.3.3 Coastal Pacific Middle West (CPMW)

Streamflow changes in CPMW domain largely resemble to its northern neighbours but experienced stronger decrease in low flows during late summer. CPMW was predominated by declining trends of low flows (i.e. Q_{10p}, Q_{25p}, Q_{min7d}), and around half of those trends were statistically significant. The duration of low-flow periods (D_{low}) has been substantially prolonged, with an average extension of 34 days for the CPMW streams. As a typical dry period in CPMW, the three late summer months (i.e. July, August, and September) exhibited widespread declines in streamflow, with field significant negative trends for Q_{jul}, Q_{aug} and Q_{sep}. In total, Q_{lsmr} decreased at more than 80% of the CPMW streams, with an average reduction of 25% from 1979 to 2021.

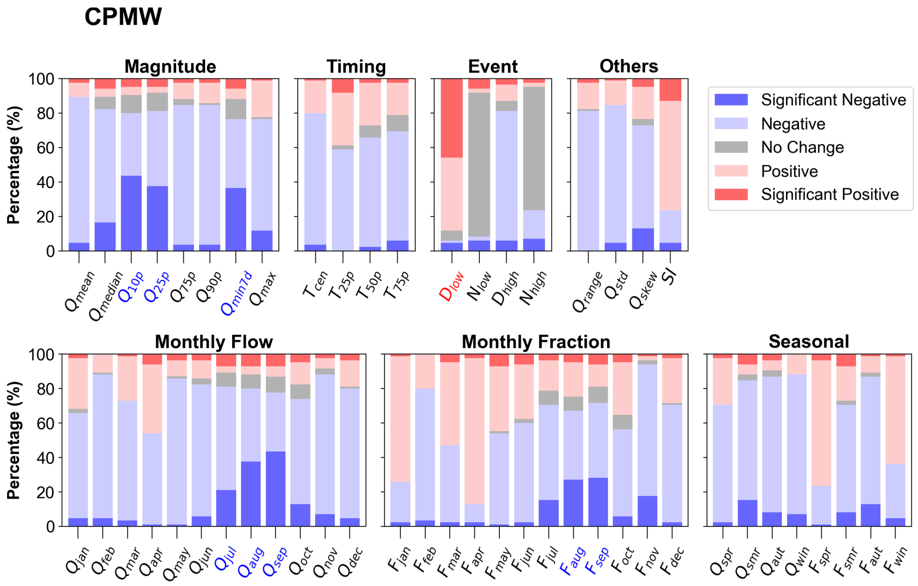


FIGURE 3.8: Percentage of the five trend types at CPMW streams.

3.3.4 Canadian Rocky Mountains (CRM)

The majority of streams in CRM domain had a positive trend in Q_{mean}, yet only 10% were statistically significant. Comparatively, more significant positive trends were observed for low flows. The increase in Q_{10p}, Q_{25p} and Q_{min7d} were all field significant, and Q_{min7d} increased an average of 8% over last 43 years. For streams in CRM domain, low flows often occur during winter, so accordingly winter flows in CRM significantly increased.

Increasing spring flow driven by earlier snowmelt has been reported in previous studies (Zhang et al., 2001; Aziz and Burn, 2006; Rood et al., 2008) and was also observed in our analysis. Almost all of the streams exhibited a positive trend in Q_{spr}, and one fifth of them were statistically significant. Across CRM domain, Q_{spr} increased an average of 18% over last 43 years with the strongest increase observed in May. Both monthly total and fractional flow in May (Q_{may} and F_{may}) showed field significant increases. The increased flows in early season advanced the timing of hydrometrics in the year. Negative trends were widely observed at the CRM for all timing hydrometrics, yet only the trend for T_{50p} was field significant. On average, the centroid of ADHs advanced 3 days from 1979 to 2021. This observation agrees with the previous study of Dery et al. (2009b) that reported earlier onset of spring freshet at snow-dominated streams across western Canada.

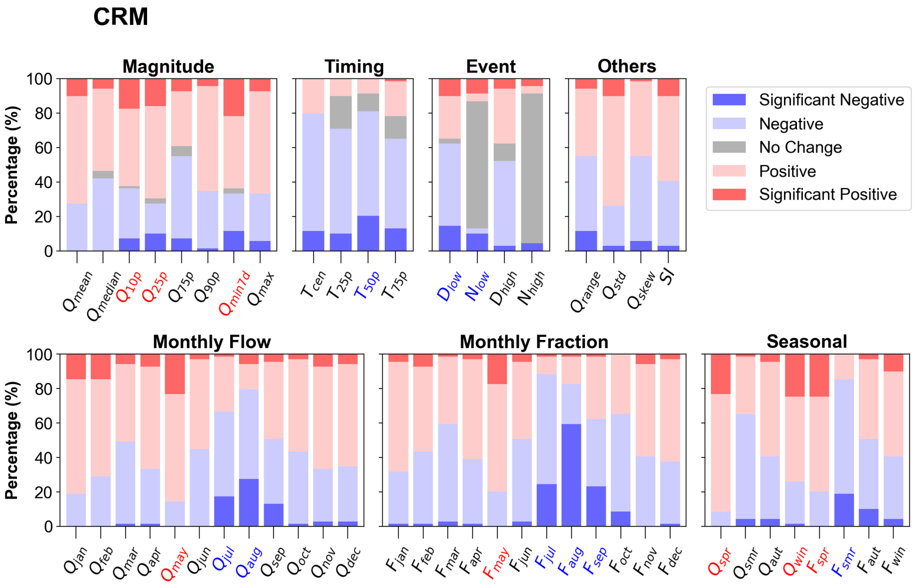


FIGURE 3.9: Percentage of the five trend types at CRM streams.

In contrast, monthly flow for July and August decreased significantly over the observation period, with an average reduction of 7% and 17%, respectively. Most of statistically significant negative trends in Q_{aug} were observed in southern BC. Reduction in fractional flows were even more remarkable. Nearly 60% of the CRM streams showed significant decrease in F_{aug}, and the negative trend for F_{jul} and F_{sep} proved field significant. On average, the contribution of late summer flow to annual total (i.e. F_{lsmr}) at CRM streams dropped from 30% to 25% over the observation period. Trend analysis on the ERA5 dataset (Fig. A2.2) indicated a reduction in summer precipitation across the CRM domain, which was considered a likely contributor to drying summer streamflow. Noteworthy, the declines in Q_{lsmr} were significantly slower at glacial watersheds (Fig.3.10). The 21 glacial watersheds located CRM domain had an average decrease of Q_{lsmr} of 4%, much lower than that for non-glacial ones (15%). This manifested that glacier was important buffer against drought for local streams.

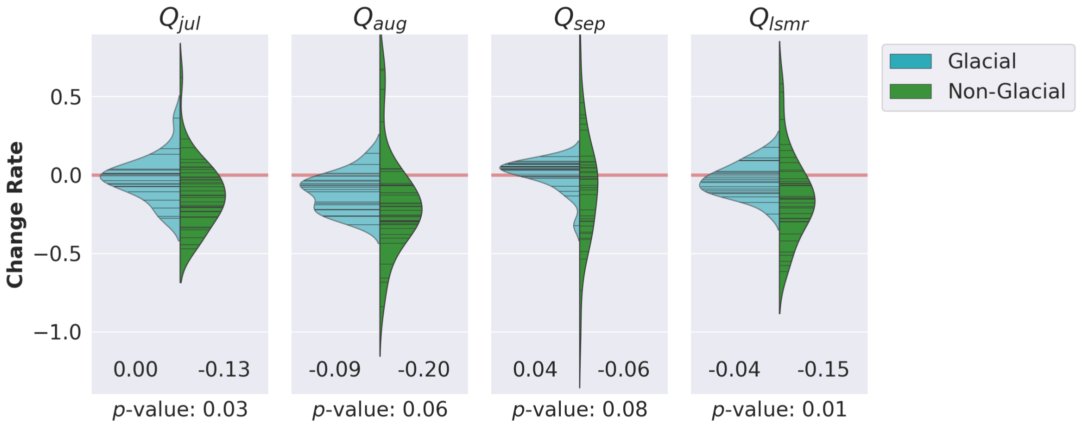


FIGURE 3.10: Comparison of rate of change of late summer flow between glacial and non-glacial streams in CRM.

3.3.5 US Rocky Mountains (USRM)

Across USRM domain, positive and negative trends in Q_{mean} were approximately equal in number, and few changes were statistically significant. Previously reported declines in Q_{mean} in USRM from 1910 to 2002 (Rood et al., 2005) were no longer significant during more recent decades. Changes of spring and late summer flow in USRM shared considerable similarity with CRM. Spring flows also increased and were field significant. On average, Q_{spr} increased by 9% over the observation period, yet the rate of increase was less than half of CRM. Negative trends in the timing of centroid were widely observed at USRM streams, and on average T_{cen} advanced 4 days over the last four decades, signaling an earlier onset of freshet. Earlier spring freshet was one of the most distinct hydrometric patterns in western USA, which has also been reported in previous studies (Cayan et al., 2001; Stewart et al., 2005).

Streamflows in August and September had field significant negative trends across the USRM. On average, Q_{aug} and Q_{sep} declined by 21% and 25% from 1979 to 2021, respectively. Furthermore, a large portion of USRM streams had a significant decline in the late summer fractional flows (F_{jul}, F_{aug}, and F_{sep}). On average, the contribution of late summer flows declined from 16% to 12% over the observation period. Declining summer flows in this domain were some of the strongest changes in WNA. These declines have been attributed to reduced summer precipitation and enhanced evapotranspiration associated with warming air temperatures along with the influence of PDO (Rood et al., 2005; Kormos et al., 2016). In this study, trend analysis of ERA5 climate data showed significant increase in air temperature and reduction in precipitation during summer period across the USRM domain (see Fig. A2.1 and A2.2). According to Rood et al.

(2005), earlier spring freshet likely resulted in drier soil condition during summer, and therefore could further reduce summer flows.

In contrast to CRM with significant increase in low flows, trends of low-flow hydrometrics at USRM streams suggested significant drying patterns. Field significant negative trends were observed for Q_{10p}, Q_{25p}, and Q_{min7d}, while the duration and frequency of drought events (D_{low} and N_{low}) significantly increased. However, a disparity in the trending direction between eastern and western slopes was observed (see Q_{10p} in Fig. 3.3). Most drying trends of low flows mostly appeared at streams in western slopes (i.e. Idaho), while significant positive trends in Q_{10p}, Q_{25p}, and Q_{min7d} were observed at eastern-slope streams (i.e. Montana and Wyoming).

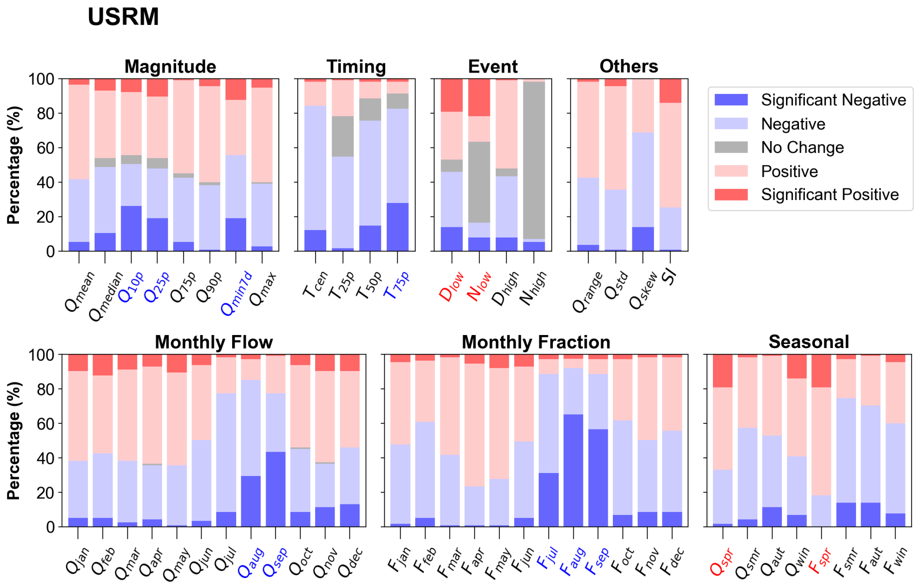


FIGURE 3.11: Percentage of the five trend types at USRM streams.

3.3.6 Southwest

Across the entire Southwest domain, declining flows were observed and nearly all of the magnitude-related hydrometrics exhibited field significant decreases. Negative trends in Q_{mean} were detected at almost all streams (only three exceptions), and more than half of them were statistically significant (Fig. 3.3). Over the last four decades, Q_{mean} at the selected 137 streams in Southwest domain dramatically declined an average of 45%. Furthermore, significant declines also observed for both high and low flows, with an average of 36% and 49% decrease for Q_{10p} and Q_{90p}, respectively. The drying patterns were seasonally consistent, as all monthly flows exhibited field significant negative trending patterns. The average decrease of the monthly flows ranged from 20% to 42%, with the greatest declines observed in April, May, June, and July, while the declines in winter was relatively small. In addition, the duration and frequency of dry pulses (i.e. D_{low} and N_{low}) increased markedly across the Southwest and on average D_{low} prolonged 79 days over the observation period.

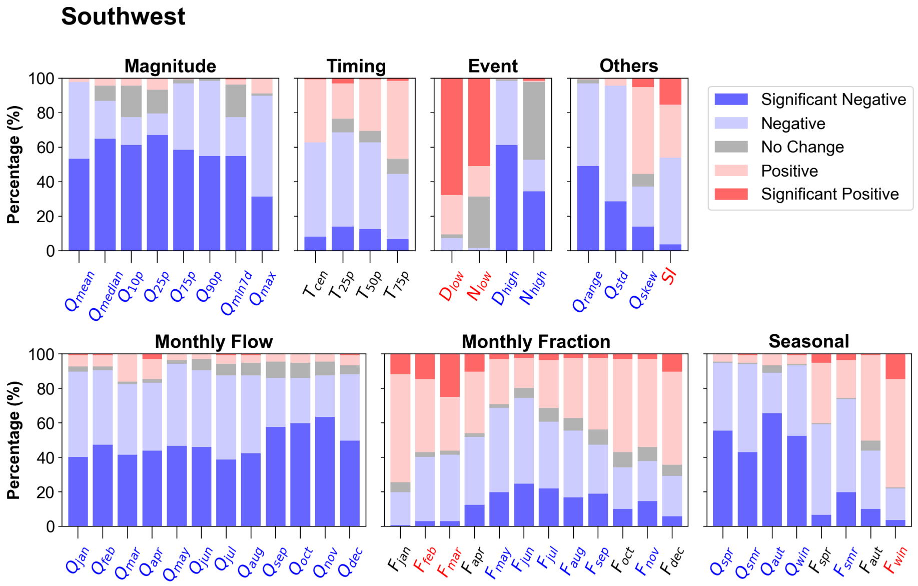


FIGURE 3.12: Percentage of the five trend types at Southwest streams.

Changes in streamflow are largely attributed to a widespread and well documented reduction in precipitation. Based on the ERA5 data, reduction in annual precipitation was widespread in the interior states (i.e. Utah, Colorado, Arizona and New Mexico), with an average reduction of 200 mm from 1979 to 2021. Most of this drying was observed in spring and summer. The widespread drying in this region had been widely reported (MacDonald, 2010; Cook et al., 2015; Williams et al., 2020), and our analysis showed that the drying patterns remained very strong in the most recent years.

3.3.7 Dashboard Application

The companion dashboard application (`https://app-trend-map.herokuapp.com`), which may take up to 30 seconds to load, allows the visualization of all analysis presented within the paper and additional functionalities. At the top of the dashboard, the region can be selected within WNA, and in fact the entire Canadian and United States RHBN and HDCN data set can be explored. Watersheds can be filtered by glacial coverage and individual watersheds can be selected. In terms of tests, the select hydrometric indicator can be chosen from a dropdown list. Five different iterations of the Mann-Kendall test can be chosen, and the significance level set as 0.01, 0.05 or 0.10. There are various outputs to visualize the results, including the yearly trend line, the daily trend line bar for the mean hydrograph and metrics of net change. Statistical values are presented and colour are used to indicate significant positive (red) and negative (blue) trends. At the bottom left, summary changes for the select regions can be visualized.

3.4 Conclusion

Changes in streamflow measured via a wide variety of hydrometric indicators suggest considerable shifts across WNA from 1979 to 2021, although degree and direction of shifting varied among the six hydro-climate regions. A general divide between wetting and drying trending pattern was detected roughly at 46th parallel north. Increasing trends in annual flows predominated in the north side, while decreasing trends prevailed in the south side. Streams in Northwest domain was featured by significant increases of low flow during cold season (from September to May) and therefore weaker seasonality. Forward shifts of streamflows from summer to spring observed in CRM domain implied that earlier warming climate had advanced snowmelt-driven spring freshets. Similarly, significant increases of spring flow were also observed in USRM domain, along with field significant negative trending patterns of low flow. The Southwest domain had a comprehensive drying pattern, as streamflows dramatically declined throughout the entire year and at all flow percentiles. Streamflows in CPMW domain also suffered from widespread negative trends, and declines were particularly strong for summer low flows. In CPNW domain, streams exhibited the significant increase in high flows and autumn flows, which was rarely reported in previous studies. Except for the Northwest, streams in all other five hydro-climate regions have large declines in late summer flows, which is considered the most outstanding and extensive changes of streamflows in WNA. Noteworthy, glacial watersheds usually exhibited higher resilience, as late summer flows declined significantly slower than glacial than non-glacial ones. However, this buffer effect of glacier against droughts is expected to be gradually weakened along with accelerated shrinkage in glacier cover and mass driven by warming climates. At monthly and seasonal scales, streamflow changes often highly correspond with the changing pattern of local precipitation, such as increased October flows in CPNW and declined late summer flows in western USA. However, this correlation with precipitation was weaker in cold regions. For example, the dramatic increase of streamflows in Northwest during cold season did not coincide with evident increases of precipitation. This supports the previous assertions that changes in the cryosphere (e.g. permafrost thaw) are influencing the magnitude and seasonality of flows in cold regions. In conclusion, this study provides an overview of the documented changes of streamflows across WNA domain during latest four decades. We refrain from commenting on the water security impacts of these changes, yet believe this information is useful to help understand the direction and drivers of change to inform water resource management.

Chapter 4

Identifying the streamflow changes in western North America from 1979 to 2021 - Part 2: Insights from machine-learned features

4.1 Introduction

In the face of a rapidly changing climate, identifying and contextualizing changes to streamflow has broad societal and scientific importance. There are numerous methods to describe and characterize the streamflow hydrograph, and the term “hydrometrics” encompasses a broad suite of human-designed indicators that quantitatively characterize streamflow. Hydrometrics can be as simple as the basic statistics of flow records (i.e. average, minimum, maximum, percentile, and standard deviation) or complex indices such as seasonality index (Coopersmith et al., 2014), slope of the flow duration curve, and streamflow elasticity (Sawicz et al., 2011). There are several hundred hydrometrics available that describe the characteristics of flow regimes, including seasonal pattern of flows, timing, frequency and duration of floods and droughts, daily, seasonal, and annual flow variability, and rate of changes (Olden and Poff, 2003). How these hydrometrics change with time is complex as there are multiple drivers of change including shifting land use, river regulation, and complex climate drivers over multiple timescales. Historically, the most common method to detect changes in streamflow is to perform the Mann-Kendall test (or its variants) on time series of hydrometrics (Lins and Slack, 1999; Zhang et al., 2001; Burn and Hag Elnur, 2002), where statistically significant trends are considered evidence for changes in streamflow. This approach however is not without critique (Wang et al., 2020).

Climate change is having a profound and comprehensive influence on a variety of hydrological processes (DeBeer et al., 2016). How these changes are manifested in streamflow is complex and often unclear due to numerous confounding feedback within the hydrological cycle. Consequently, individual hydrometrics are inadequate to fully describe changes in streamflow. When performing trend analyses, there is typically a chosen set of hydrometrics that sufficiently represent key process and changes for given streams. However, flow regimes are regionally diverse and vary geographically based on climate, geology, topography and vegetation cover (Poff et al., 1997). Therefore, some hydrometrics may be indicative for some flow regimes but not applicable for others. For example, the timing and magnitude of spring freshet is a sensitive indicators for snow-dominated streams (Barnett et al., 2005) but are less useful in regions with little snow accumulation. As such, the hydrometric indicators used to assess trends must be designed specifically for different geographic regions or flow regimes. Furthermore, hydrometric sets are often subject to information redundancy (Olden and Poff, 2003) that potentially reduces their effectiveness and efficiency for trend analysis. Practically, this provides challenges to select and design hydrometric suites suitable for change detection, and there remains considerable subjectivity and domain expertise required when evaluating how streamflow regimes are changing.

Deep Learning (DL) has achieved rapid and considerable success in many fields of science and engineering, and its potential and capabilities in hydrology have been explored and discussed in recent work (Shen, 2018; Reichstein et al., 2019; Beven, 2020). DL improves on previous machine learning algorithms due to its exceptional capability for self-extracting features from raw data (LeCun et al., 2015), which saves considerable labour and time for manually designing features and model pretraining. In the family of DL algorithms, autoencoder is the prevailing method for dimensionality reduction and feature learning (Goodfellow et al., 2016). Kingma and Welling (2014) proposed a modified version, called Variational Auto-Encoder (VAE), which improves generalization and is adaptive and robust to "unseen" samples. VAE transforms input data into Latent Features (LFs) that are a low-dimension representation, yet preserve the most salient features of the input data. In this way, LFs are regarded as machine-learned features from raw data. While VAE provides exceptional feature learning capabilities, there are few studies to explore its utility in hydrological applications.

The objective of this study is to use machine learned features (LFs) to identify streamflow changes in Western North America (WNA) over the last four decades (1979-2021). We obtain Annual Daily Hydrograph (ADHs) from more than 500 natural streams (i.e. minimally disrupted by human activities over observation period) across WNA domain and use VAE to extract a number of LFs to characterized the machine-learned features describing the flow regimes in WNA. Trend analysis is performed on the extracted LFs for each stream with sufficient flow records and are compared with results from classical hydrometrics on the same data set presented in Chapter 3. The coherence and differences between the changes in LFs and classical hydrometrics is scrutinized to identify changes not represented in typical hydrometric sets. Finally, we seek to identify the physical meaning and hydrological implications of the extracted LFs, and provide guidance of future applications of this approach.

4.2 Data

Daily streamflow data were obtained from the Referenced Hydro-metric Basin Network (RHBN) and Hydro-climate Data Network (HCDN), which are operated by Water Survey of Canada (WSC) and USGS, respectively. Both networks include streams of with natural flow regimes that are minimally affected by human activities (e.g. significant land-use change, dams, reservoirs, and hydro-power stations) across the observation period. They have been widely used to detect changes of streamflow across North America (Lins and Slack, 1999; Zhang et al., 2001; Burn and Hag Elnur, 2002; Stewart et al., 2005). Long-term daily flow data were broken into Annual Daily Hydrographs (ADHs). Each ADH contains 365 values of daily flow within one year from 1 January to 31 December (leap days excluded if applicable). Small gaps ($\leq$7 days) were filled using linear interpolation, while ADHs with large gaps (>7 days) were excluded from analysis. ADHs were collected from streams across western North America (WNA), covering four Canadian Provinces/Territories (British Columbia, Alberta, Yukon, and Northwest Territory) and eleven American states (Washington, Oregon, California, Idaho, Nevada, Montana, Utah, Wyoming, Colorado, Arizona and New Mexico). In total, 30510 ADHs from nearly 700 stream gauges in the WNA domain were included in this dataset. Data are divided into three subsets in a ratio of 0.6, 0.2, and 0.2 for training, validating, and testing VAE models respectively. Furthermore, a subset of streams with long, continuous flow records were selected for trend analysis of LFs. Those streams must have at least 30 full-record ADHs within 43-year period from 1979 to 2021. In total, 557 streams were qualified (see Fig.4.1) and are the same data set presented in Chapter 3.

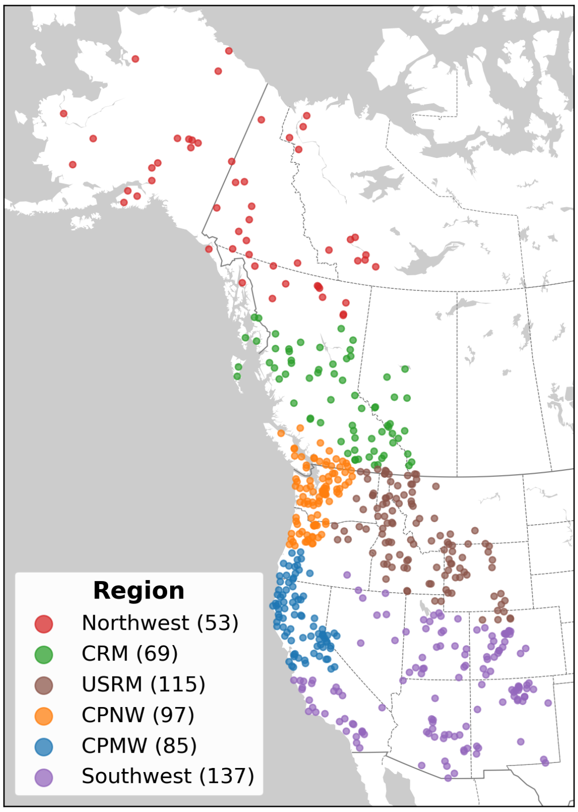


FIGURE 4.1: Streamflow gauges of HCDN and RHBN selected for trend analysis

Flow data often has a highly skewed distribution, which can deteriorate the performance of deep-learning models, so a site-specific normalization method (Eq. 4.1) was applied.

$$y_i = \max[1, \sqrt{x_i}/P_{99.8\%}(\sqrt{x_1}, \sqrt{x_2}, \sqrt{x_3}, ..., \sqrt{x_n})] \tag{4.1}$$

where $x_i, x_2, x_3, ..., x_n$ denote the daily flow records, $P_{99\%}$ indicates the 99-th percentile of long-term streamflow data, and y_i is the normalized values of daily flow. Square-root transformation was applied to alleviate the skewness of distribution in the flow data. In this case, VAE can only generate values within a range between 0 and 1, as a sigmoid function was employed as the activation function for the last layer of decoder. Therefore, it was necessary to re-scale values of input ADHs in the 0-1 range. A high percentile ($P_{99\%}$) was employed as the upper bound in the rescale equation instead of the maximum ($P_{100\%}$) to eliminate the most extreme events in the analysis.

4.3 Methods

4.3.1 Variational AutoEncoder (VAE)

Autoencoder is a distinct type of neural network that attempts to copy input to output information through fewer variables and consists of two sub-networks: encoder and decoder. An encoder function $z = f(x)$ converts input data (x) into a new lower-dimension representation (z) called Latent Features (LFs). In the next step, a decoder function $\hat{x} = d(z)$ reconstructs data to the original format $(\hat{x})$ based on the LFs. An autoencoder is trained to maximally transfer the most salient features of input data into LFs as well as filter out insignificant variations (noise). Autoencoder is recognized as extremely effective for dimensionality reduction and information compression. It is more powerful than linear embedding techniques (e.g. PCA) due to the strong capability of learning non-linear structures of datasets (Goodfellow et al., 2016).

Proposed by Kingma and Welling (2014), VAE is considered an excellent combination of deep learning and Bayesian inference. It employs approximate inference $q(x|z)$ and $p_{model}(x|z)$ as the encoder and decoder as opposed to $z = f(x)$ and $\hat{x} = d(z)$. Instead of converting input data to fixed codes, VAE attempts to learn the probability distribution of z in latent space. Conventionally, z is presumed to form a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ and the VAE encoder is trained to learn the two primary parameters: mean (μ) and variance (σ). Then, the VAE decoder randomly samples z from the distribution defined by the learned parameters μ and σ and converts them back to the original format of inputs. The VAE workflow is illustrated in Fig. 4.2.

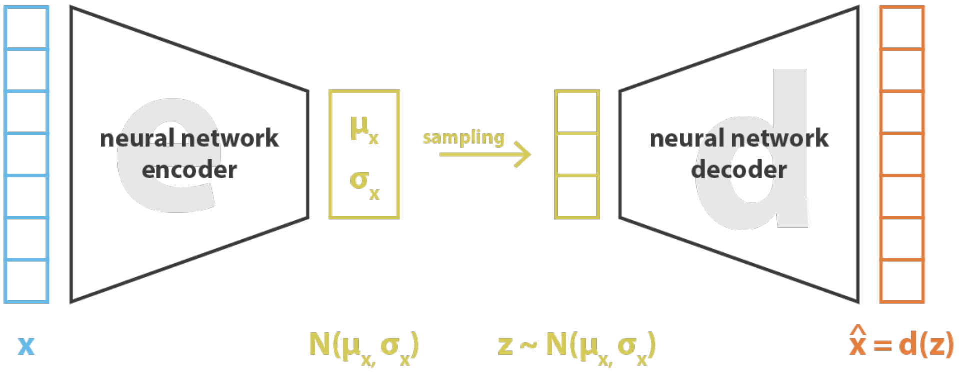


FIGURE 4.2: Diagram of VAE model, from Rocca (2019). x and $\hat{x}$ denote input and output data. μ_x and σ_x indicate the mean and variance for a Gaussian distribution $\mathcal{N}$ for latent variables z.

Another feature of VAE is the two-term loss function (Eq. 4.2) that consists of both reconstruction loss $\mathbb{E}_{z\sim q(z|x)} \log p_{model}(x|z)$ and regularization loss $D_{KL}(q(z|x)||p_{model}(z)$. Reconstruction loss is the cross entropy between $q(z|x)$ and $p_{model}(x|z))$ that forces VAE to reconstruct data as close to the original input as possible. Regularization loss measures the KL divergence between $q(z|x)$ and $p(z)$, and $p(z)$ denotes the prior distribution of z. This helps VAE to learn well structured latent spaces, therefore reducing overfitting to training data and enhance model generalization (Chollet, 2017). Additional mathematical details regarding VAE can be found in the paper of Kingma and Welling (2014).

$$\mathcal{L}(q) = \mathbb{E}_{z\sim q(z|x)} \log p_{model}(x|z) - D_{KL}(q(z|x)||p_{model}(z)) \tag{4.2}$$

Building and Testing the VAE

In this study, *Keras* (Chollet et al., 2015), a Python package for deep learning research, was used to build VAE models. Success of VAE largely depends on whether LFs capture salient features from the input ADH dataset. In order to search for the VAE that is most suitable for ADHs in WNA, a number of hyper-parameters were tuned, including the depth (i.e. # layers) and width (i.e. # nodes) of encoder and decoder network, and the dimensions of latent space. The baseline model was built with one-layer encoder and decoder with 128 nodes, and the dimension of latent space was set as six. Typically, the encoder adopted a downward-triangle structure, while the decoder was upward-triangle (Fig. 4.2). In the encoder network, the number of nodes of new layer is always half of its previous one, while the nodes of last layer must be no less than 32. Encoders were tested with an initial layer of 128, 256, 512, 1024, and 2048 nodes. Conversely, decoders were built up from bottom to top. In decoder network, a new layer was put on top of the existing layers, and is half size of the lower one. Decoders were tested with a last layer of 128, 256, 512, 1024, and 2048 nodes. Different combination of encoder and decoder were tested to search for the optimal model architecture. As recommended in Chollet (2017), Rectified Linear Unit (ReLU) was used as the activation function for all layers except for the last one, which employed Sigmoid. Adam (Kingma and Ba, 2015) was chosen for the optimizer. Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970), as a supplementary metric, was utilized to evaluate VAE performance through measuring goodness-of-fit between original and reconstructed ADHs.

4.3.2 Trend Analysis

MK test is the prevailing statistical test for trend analysis due to its simplicity of use and minimum requirement for time series data (i.e. non-parametric), yet is limited as it can only detect monotonic trends in the time series. The original MK test (Mann, 1945; Kendall, 1975) is subject to serial correlation in time series, and a number of modified versions have been proposed in previous studies that allow more robust trend detection in hydrological time series (Hamed and Rao, 1998; Yue and Wang, 2002; Yue et al., 2002; Yue and Wang, 2004). One of the variations of MK test, Trend-Free Mann-Kendall test (TFMK; Yue and Wang, 2002), was employed in this study to examine the potential trends in time series of LFs. The trend line of the time series was fitted using Sen's slope (Sen, 1968), which is often used to estimate the magnitude of changes. LFs were extracted from ADHs, so they are treated as annual-based "hydrometrics". For every stream, the extracted LFs from ADHs are arranged in yearly order to form a time series of LFs for trend analysis. In order to make this study comparable to Chapter 3, the same working scheme was adopted for trend analysis.

4.3.3 Correlation with Hydrometrics

An important aspect of this study is to interpret LFs extracted from streamflow data with respect to classical hydrometric indicators; which have well documented rationales. 52 hydrometrics were selected that represent various aspects of streamflow characteristics including magnitude, timing, duration, frequency, and rate of change. Spearman's rank test was used to calculate the correlation between LFs and these hydrometrics, which is most appropriate considering the highly skewed nature of most hydrometrics.

4.4 Results

4.4.1 VAE Model Selection

The performance of the VAE was largely dependent upon the decoder architecture and less sensitive to the choice of the encoder network. (Fig.4.3). Wider decoder networks always resulted in smaller training loss, yet this was not the case for the test loss. Test loss exhibited marked improvement from 128 to 512 layers, yet further widening of decoder deteriorated VAE performance. We presumed deeper networks were more capable of approximating complicated inferences, yet in this case, VAE did not exhibited consistent improvement with the depth of the decoder network. There was a significant improvement in training loss when the number of layers was increased from one to three (Fig.4.4), yet improvement became subtle with four layers and even led to higher training loss when additional layers were appended. This degrading effect is more apparent on the testing set, as loss typically increased if the decoder network grew deeper than three layers (Fig. 4.4). After many iterations, a three-layer decoder with a final layer of 512 nodes (512-D3) was considered optimal for this ADH dataset. The encoder had less of an influence on VAE performance as training losses remained largely invariant based on encoder architecture (Fig. 4.3). One-layer encoders resulted in higher testing losses than deeper counterparts, suggesting multiple layers for the encoder. To achieve a balance of power between encoder and decoder, a symmetric architecture is typically preferred by autoencoder. A three layer encoder (512-E5) was chosen for VAE as a mirror of the selected decoder.

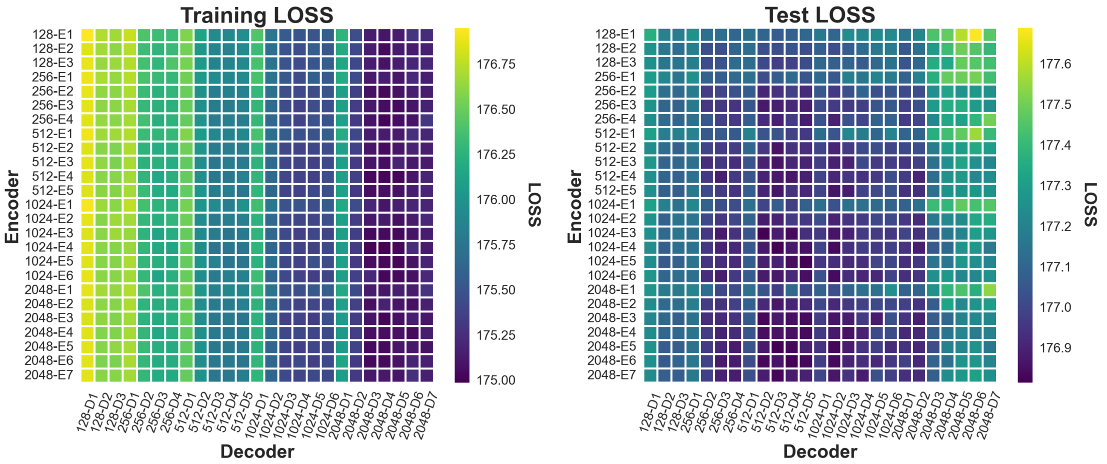


FIGURE 4.3: Comparison of VAE network architecture. The number before dash (-) indicates the number of nodes of the initial layer for an encoder (E) and of the last layer for a decoder (D). The number after E or D indicates the number of layers of the network. For example, 128-E3 represent a three-layer encoder with an initial layer of 128 nodes, denoted as [128, 64, 32], and 128-D3 is a three-layer decoder with a last layer of 128 nodes (i.e. [32, 64, 128]).

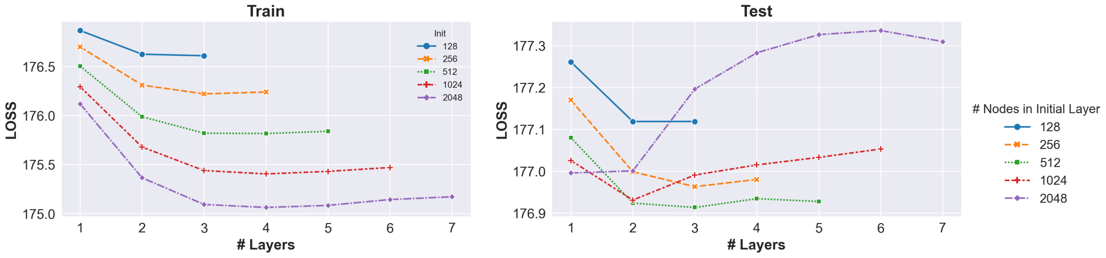


FIGURE 4.4: Comparison of VAE decoder with different number of layers.

After the VAE architecture was established, the next step was to obtain the optimal number of LFs (between 2 and 20) that described the ADHs. Both training and testing loss declined sharply when the number of LFs grew from two to eight, yet further additional LFs did not dramatically improve performance (Fig.4.5). This suggests that eight LFs were sufficient to capture most features of WNA flow regimes and the selected VAE model employed eight LFs between the encoder and decoder.

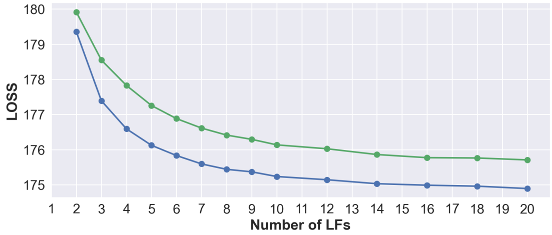


FIGURE 4.5: Comparison of VAE performance with different number of LFs.

4.4.2 Site Performance of VAE

For all sites, VAE performed best for high-latitude mountainous streams and had poor performance for dry regions in the southwestern USA (Fig. 4.6). Performance of VAE was evaluated by site-average NSE, which measured the goodness-of-fit between the original and reconstructed ADHs for a given stream. Higher NSE values indicate the VAE has captured the salient features of the site ADHs. Streamflow regimes dominated by snowmelt have a lower degree of natural complexity than rain-dominated ones, so embedding dimension is more likely to succeed (Aksamit and Whitfield, 2019). For example, in WNA snow-dominated streams, hydrographs are typically dominated by freshet during the melt season that accounts for a large proportion of the annual flow. While freshet occurs at different times across WNA, and changes over short and long time scales, its seasonal and regular nature enhances the ability of VAE to simulate ADHs. In contrast, streams in the southwestern USA (i.e. southern California, Arizona, and New Mexico) are characterized by intermittent and short-lived events that are result in very poor NSEs. The episodic nature of these flows makes it difficult for the VAE to capture, comprehend and reconstruct ADHs in this region. This issue has been previously recognized, as the low predictability of streams in the dry southwest USA has been attributed to large inter-annual variability of climate and low gauge density (Patil and Stieglitz, 2012). While also dominated by rainfall, streams in the coastal pacific have higher NSE largely because precipitation is more seasonal; generally occurring in winter. Furthermore, due to high gauge density, training samples include numerous ADHs that increase the chance of VAE recognizing and extracting the salient features of the flow regime.

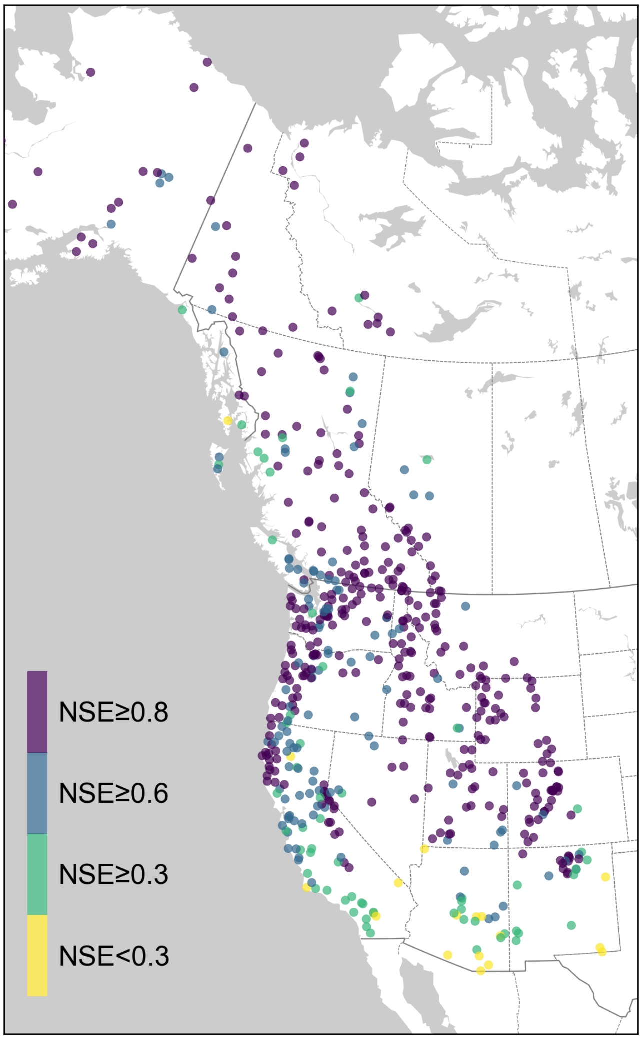
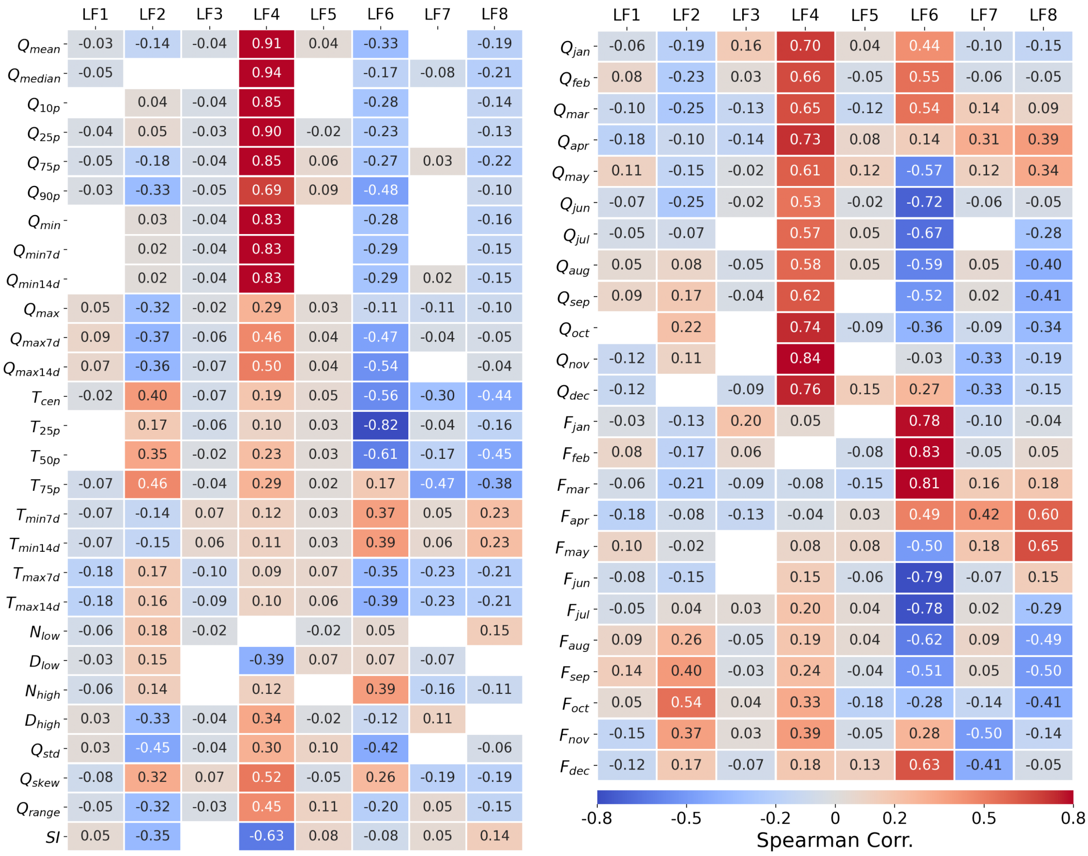


Figure 4.6: Site-average NSE map.

4.4.3 Interpretation of LFs

Assessing what streamflow characteristics have been identified as LFs by the VAE model is critical to understanding their value in hydrological analysis. We used two approaches to interpret LFs: 1) Spearman rank correlation between LFs and hydrometrics (Fig. 4.7), and 2) Reconstructed Hydrograph Evolution (RHE; Fig. 4.8). RHEs provide an approach to evaluate the influence of each LF on the temporal distribution of streamflow across a year and are generated for each of the eight LFs individually. For each target LF, values are spread across a range from the 1st to 99th percentile of the corresponding distribution formed by all ADHs, while other LFs are held constant at their median value. Following this, the artificially adjusted LFs were fed to the VAE decoder to generate a series of synthetic ADHs exhibiting the evolution of the hydrograph with the values of the target LF. Both approaches reveal that LFs are commonly associated with different hydrometrics and hydrograph components. However, the relation between LFs and the hydrograph varies, with some LFs exhibiting a stronger connection with hydrometrics and the general shape of the ADHs than others. Among the LFs, LF4 and LF6 are most closely linked to classical hydrometrics.



	LF1	LF2	LF3	LF4	LF5	LF6	LF7	LF8
Q_{mean}	-0.03	-0.14	-0.04	0.91	0.04	-0.33		-0.19
Q_{median}	-0.05			0.94		-0.17	-0.08	-0.21
Q_{10p}		0.04	-0.04	0.85		-0.28		-0.14
Q_{25p}	-0.04	0.05	-0.03	0.90	-0.02	-0.23		-0.13
Q_{75p}	-0.05	-0.18	-0.04	0.85	0.06	-0.27	0.03	-0.22
Q_{90p}	-0.03	-0.33	-0.05	0.69	0.09	-0.48		-0.10
Q_{min}		0.03	-0.04	0.83		-0.28		-0.16
Q_{min7d}		0.02	-0.04	0.83		-0.29		-0.15
Q_{min14d}		0.02	-0.04	0.83		-0.29	0.02	-0.15
Q_{max}	0.05	-0.32	-0.02	0.29	0.03	-0.11	-0.11	-0.10
Q_{max7d}	0.09	-0.37	-0.06	0.46	0.04	-0.47	-0.04	-0.05
Q_{max14d}	0.07	-0.36	-0.07	0.50	0.04	-0.54		-0.04
T_{cen}	-0.02	0.40	-0.07	0.19	0.05	-0.56	-0.30	-0.44
T_{25p}		0.17	-0.06	0.10	0.03	-0.82	-0.04	-0.16
T_{50p}		0.35	-0.02	0.23	0.03	-0.61	-0.17	-0.45
T_{75p}	-0.07	0.46	-0.04	0.29	0.02	0.17	-0.47	-0.38
T_{min7d}	-0.07	-0.14	0.07	0.12	0.03	0.37	0.05	0.23
T_{min14d}	-0.07	-0.15	0.06	0.11	0.03	0.39	0.06	0.23
T_{max7d}	-0.18	0.17	-0.10	0.09	0.07	-0.35	-0.23	-0.21
T_{max14d}	-0.18	0.16	-0.09	0.10	0.06	-0.39	-0.23	-0.21
N_{low}	-0.06	0.18	-0.02		-0.02	0.05		0.15
D_{low}	-0.03	0.15		-0.39	0.07	0.07	-0.07	
N_{high}	-0.06	0.14		0.12		0.39	-0.16	-0.11
D_{high}	0.03	-0.33	-0.04	0.34	-0.02	-0.12	0.11	
Q_{std}	0.03	-0.45	-0.04	0.30	0.10	-0.42		-0.06
Q_{skew}	-0.08	0.32	0.07	0.52	-0.05	0.26	-0.19	-0.19
Q_{range}	-0.05	-0.32	-0.03	0.45	0.11	-0.20	0.05	-0.15
SI	0.05	-0.35		-0.63	0.08	-0.08	0.05	0.14

	LF1	LF2	LF3	LF4	LF5	LF6	LF7	LF8
Q_{jan}	-0.06	-0.19	0.16	0.70	0.04	0.44	-0.10	-0.15
Q_{feb}	0.08	-0.23	0.03	0.66	-0.05	0.55	-0.06	-0.05
Q_{mar}	-0.10	-0.25	-0.13	0.65	-0.12	0.54	0.14	0.09
Q_{apr}	-0.18	-0.10	-0.14	0.73	0.08	0.14	0.31	0.39
Q_{may}	0.11	-0.15	-0.02	0.61	0.12	-0.57	0.12	0.34
Q_{jun}	-0.07	-0.25	-0.02	0.53	-0.02	-0.72	-0.06	-0.05
Q_{jul}	-0.05	-0.07		0.57	0.05	-0.67		-0.28
Q_{aug}	0.05	0.08	-0.05	0.58	0.05	-0.59	0.05	-0.40
Q_{sep}	0.09	0.17	-0.04	0.62		-0.52	0.02	-0.41
Q_{oct}		0.22		0.74	-0.09	-0.36	-0.09	-0.34
Q_{nov}	-0.12	0.11		0.84		-0.03	-0.33	-0.19
Q_{dec}	-0.12		-0.09	0.76	0.15	0.27	-0.33	-0.15
F_{jan}	-0.03	-0.13	0.20	0.05		0.78	-0.10	-0.04
F_{feb}	0.08	-0.17	0.06		-0.08	0.83	-0.05	0.05
F_{mar}	-0.06	-0.21	-0.09	-0.08	-0.15	0.81	0.16	0.18
F_{apr}	-0.18	-0.08	-0.13	-0.04	0.03	0.49	0.42	0.60
F_{may}	0.10	-0.02		0.08	0.08	-0.50	0.18	0.65
F_{jun}	-0.08	-0.15		0.15	-0.06	-0.79	-0.07	0.15
F_{jul}	-0.05	0.04	0.03	0.20	0.04	-0.78	0.02	-0.29
F_{aug}	0.09	0.26	-0.05	0.19	0.04	-0.62	0.09	-0.49
F_{sep}	0.14	0.40	-0.03	0.24	-0.04	-0.51	0.05	-0.50
F_{oct}	0.05	0.54	0.04	0.33	-0.18	-0.28	-0.14	-0.41
F_{nov}	-0.15	0.37	0.03	0.39	-0.05	0.28	-0.50	-0.14
F_{dec}	-0.12	0.17	-0.07	0.18	0.13	0.63	-0.41	-0.05

FIGURE 4.7: Spearman's correlation matrix between hydrometrics and LFs. Only show correlation that prove statistically significant ($p < 0.01$)

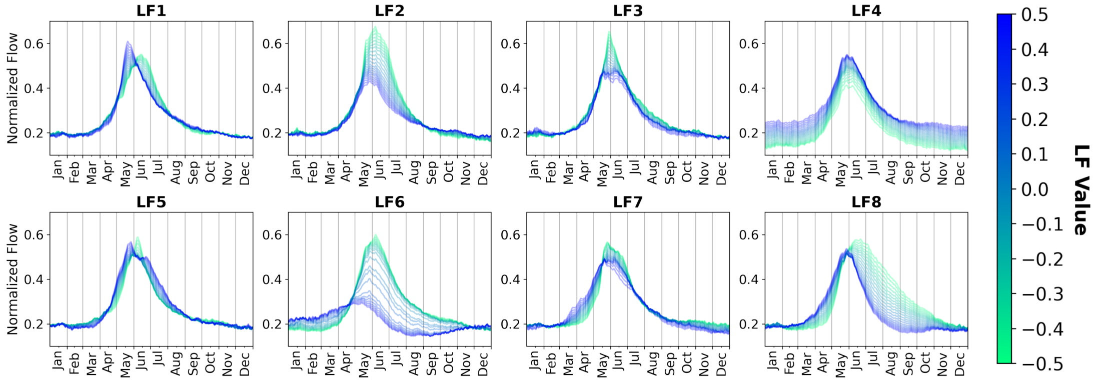


Figure 4.8: Evolution of ADHs with changing LFs.

LF6 effectively divides streams into snow- and rain-dominated regime across WNA. In the coastal Pacific and US southwest, rainfall dominated regimes have a positive value of LF6 (Fig. 4.9). In contrast, mountainous and high latitude streams dominated by snowmelt freshet, the most distinct feature of these streams, generally have negative values of LF6. By lowering the value of LF6 in the RHE, there was a progressive decline in spring freshet and it acted as an effective binary classifier of snow versus rain dominated regimes in WNA. As the strength of the correlation between LFs and hydrometrics varies by flow regime, we assigned streams to a snow or rain-dominated group based on their LF6 value. Following this, Spearman rank correlation was conducted for these three groups (all, snow and rain) separately.

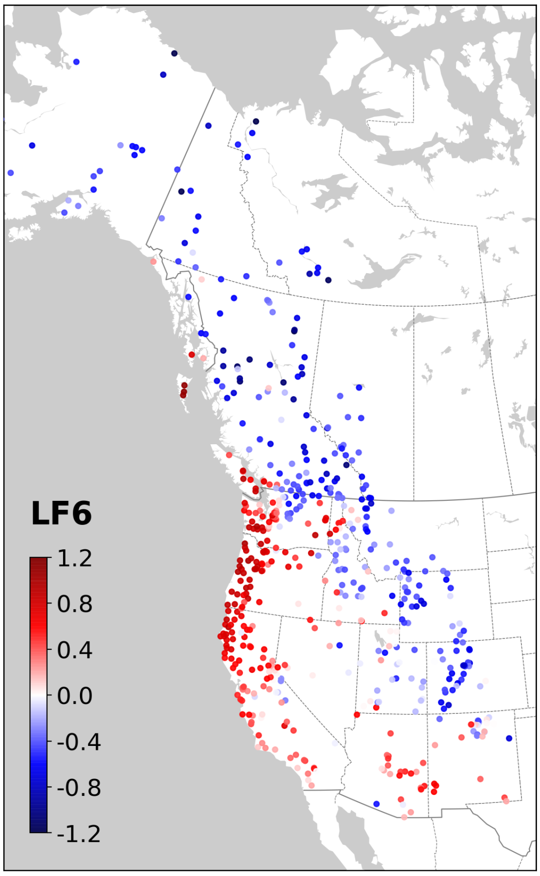


FIGURE 4.9: Map of site average of LF6.

LF4 is a broad indicator of streamflow quantity, exhibiting statistically significant correlations with almost all magnitude-related hydrometrics (e.g. flow percentiles and monthly flows), yet the strength of correlation with high flows is weaker. For streams grouped as rainfall dominated (LF6>0), Spearman's correlation of LF4 with Q_{min} is substantially lower than that with Q_{max}. For streams in the snow group (LF6<0), high flows most often occur during spring and early summer, so there were weaker relations between LF4 and corresponding monthly flows (from May to September).

LF2 is related to LF4 in that for snow-dominated streams, it is associated with the magnitude of peak flows, and has very high correlation (>0.6) with Q_{max7d} and Q_{max14d}. The RHE of LF2 indicates that increasing the value of LF2 will boost the dominant peak flows of the reconstructed ADHs (Fig. 4.10). As peak flow in snow dominated streams typically occurs from May to July, LF2 had strong correlations with monthly flows during this period (Q_{may}, Q_{jun}, and Q_{jul}). In contrast, streams in the rain group lacked any relation between LF2 and peak flows. Overall, LF2 represents the balance of seasonal flow distribution between winter and late summer. LF2 is negatively correlated with winter fractions (F_{jan}, F_{feb}, and F_{mar}), while positively related to late summer fractions (F_{aug}, F_{sep}, and F_{oct}). In WNA, rain-dominated streams typically feature higher winter flows compared with summer. Increasing the value of LF2 narrows the gap between winter and summer and result in comparatively uniform streamflow throughout a year (Fig. 4.10). This resultant low seasonality explains the strong negative correlation between LF2 and SI, which is not observed for streams in snow group. Finally, the ability to shift the flow distribution between winter and late summer suggests LF2 is influential to the timing of flow mass center and cumulative flow percentiles of ADHs, shown by the strong negative correlations with the timing hydrometrics of T_{cen}, T_{25p}, T_{50p}, and T_{75p}.

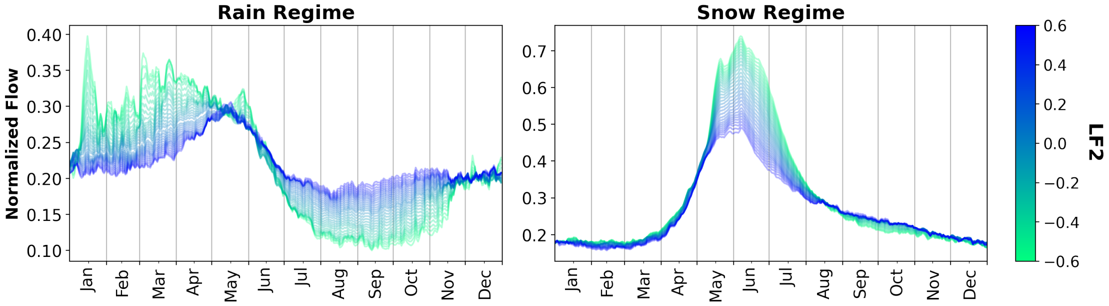


FIGURE 4.10: Control of LF2 on the shape of snow- and rain-dominated ADHs.

In terms of timing hydrometrics, LF8 had a strong correlation with T_{cen} and T_{50p} for snow-dominated streams, which were absent for the rain group. Northern streams in general had a lower LF8 value compared with southern streams, as freshet typically occurred later with increasing latitude. For example, streams in Alaska and Yukon had the lowest values of LF8 of the snow-dominated streams. Low LF8 values were also observed for streams in southern BC that have substantive glacial coverage. LF8 is positively correlated with spring streamflow (March, April and May) and negatively correlated with late summer flows (July through October). Applying the RHE, freshet shifts from spring to summer when LF8 is lowered whereas increasing LF8 advances freshet and increases early season flows. (Fig. 4.8).

LF7 is most strongly correlated to late autumn monthly flows (Q_{nov} and Q_{dec}) for rain dominated streams, and has only a very weak association with the snow group hydrometrics. Coastal Pacific streamflow regimes are most often rainfall dominated during winter, with northern streams (in Southern BC, Washington and northern Oregon) having distinct high-autumn flows compared with more southerly streams (southern Oregon, California). LF7 has a clear ascending gradient from north to south (Fig. 4.11) Similar gradients were observed for F_{nov} and F_{dec} in a descending order, which explains the strong negative correlation.

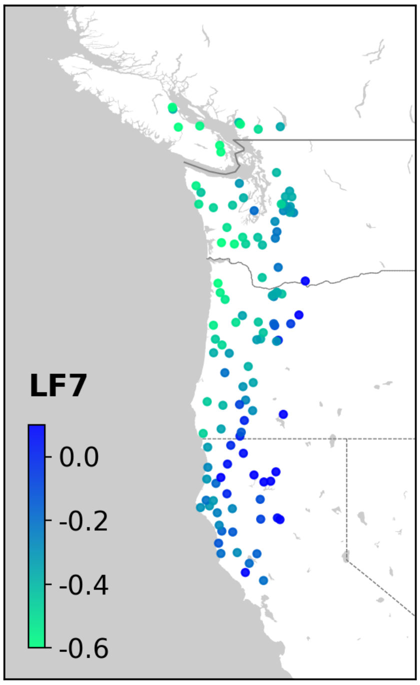


FIGURE 4.11: Map of site-average LF7 of coastal Pacific streams.

There were certain LFs that had weak correlations with hydrometrics and uncertain hydrological meaning. LF1 and LF3 have correlation coefficients > 0.4 with only one or two hydrometric indicators (F_{Feb} and F_{Jan}), which are limited to streams in the rain group. LF5 had the most ambiguous meaning as it had no correlation greater than > 0.4 for any hydrometric in either group. The RHE also suggests that the shape of the reconstructed ADH was insensitive to changes in the LF5 value (Fig. 4.8).

4.4.4 Trend Analysis on LFs

As each LF has a yearly value, we performed trend analyses for each LF at each site for the 43 years of data in an analogous approach to hydrometric trend analysis performed in Chapter 3. The changing pattern of streamflow revealed by the LFs were in considerable agreement compared with classical hydrometrics, suggesting the LF are hydrologically meaningful and may be an effective approach for change detection (Fig. 4.12).

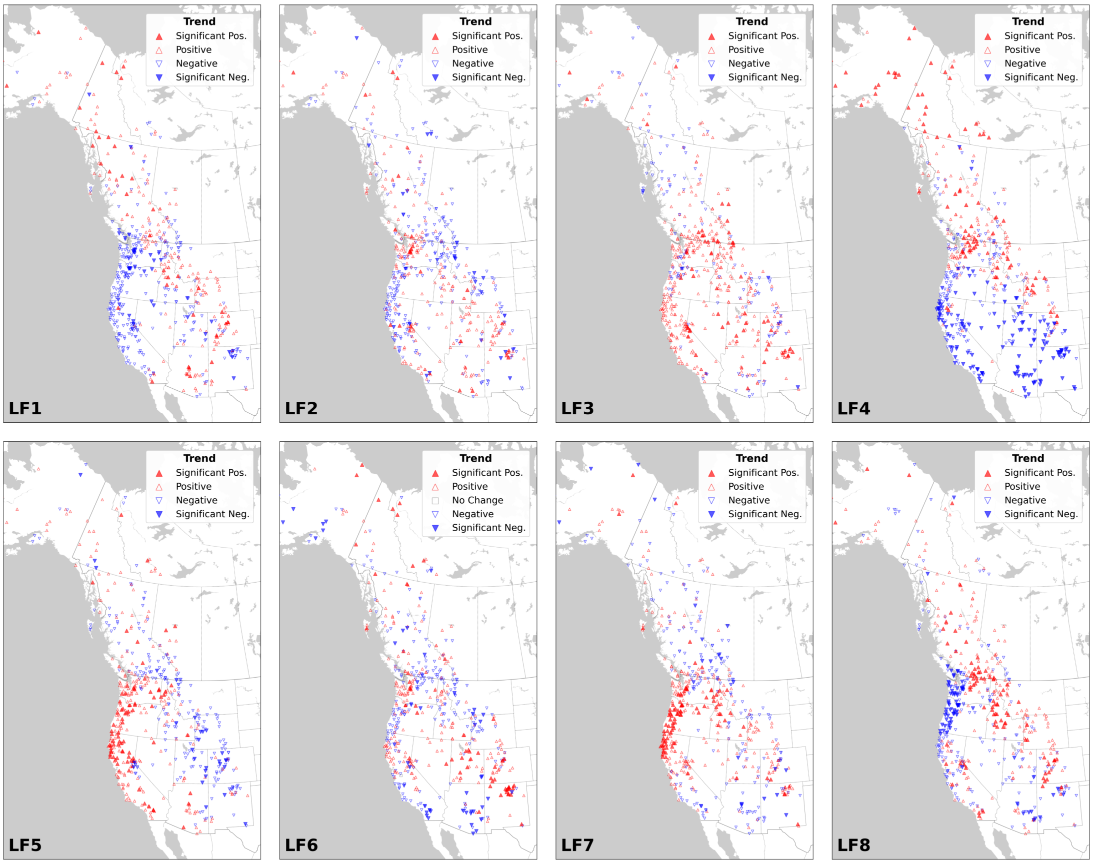


FIGURE 4.12: Trend map of LFs in WNA.

The trend in LF4 exhibited considerable similarity with Q_{mean} (Fig.4.13) across the domain. Significant declines were widely observed in the southwestern USA, where severe recent drought has been widely reported (Cook et al., 2015; Williams et al., 2020), reinforcing the utility of LF4 as an indicator of streamflow quantity and concurring with the hydrometric analysis presented in Chapter 3. Furthermore, where streamflow increases were observed in the Northwest, Coastal Pacific Northwest and US Rocky Mountains, LF4 also increased and was highly consistent with Q_{mean} and Q_{median}.

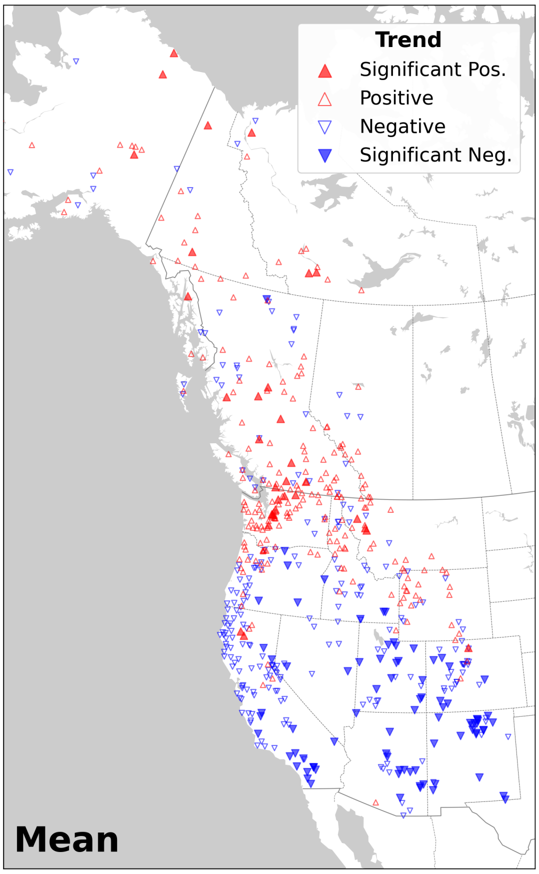


FIGURE 4.13: Trend map of annual mean flow (Q_{mean}) in WNA.

Positive trends in LF6 were apparent in the southern portion of Rocky Mountains (Utah, Colorado, and northern New Mexico) and CPNW, many of which are statistically significant. As an indicator of snow-to-rain transition (SRT), the increases in LF6 suggests these streams are becoming increasingly dominated by rain. This weakening snowmelt signal has been previously reported in mountains areas of WNA (Regonda and Rajagopalan, 2005; Mote et al., 2005). Representing the integrated impacts of SRT is complex, and typically requires a number of hydrometrics that include timing of onset and peak of spring freshet, centroid of flow mass, seasonal fractional flows, and date of snowmelt (Whitfield and Cannon, 2000; Stewart et al., 2005; Dudley et al., 2017). However, LF6 was able to represent SRT through a single metric.

For central coastal Pacific sites in the United States, there were widespread significant increases in LF7 (Fig. 4.12). While similar, streams above and below this region are distinct in that they do not have characteristically higher flows in November and December. As noted, there is a strong negative correlation between

LF7 and Q_{nov} and Q_{dec}, and this result suggests that northern coastal Pacific sites are gradually shifting in their flow regime similar to streams to the south, with less flows during early winter.

Earlier snowmelt and spring freshet have been widely reported across the mountainous areas of WNA (Barnett et al., 2005; Regonda and Rajagopalan, 2005). A strong positive trend in LF8 was identified in the U.S. Rocky Mountains, particularly in Western Idaho. This increase indicates the earlier arrival of freshet as it is negatively correlated to the timing metrics, which again is confirmed through the results in Chapter 3. As negatively related to the timing of snowmelt-driven freshet, increasing LF8 indicates earlier arrival of spring freshet at those streams. This result is supported by previous findings of snowmelt-driven streamflows shifting toward earlier dates across interior PNW (Regonda and Rajagopalan, 2005).

4.5 Discussion

4.5.1 Evolution of ADH based on LF trends

Understanding the temporal evolution of streamflow hydrographs is critical for water resource management, and we present a new approach to estimate the progressive evolution of streamflow over time based on trends in LFs. First, we selected streams that had statistically significant ($p < 0.1$) trends for at least one LF and whose goodness-of-fit ($NSE > 0.7$) was high between the original and reconstructed ADHs to ensure that LFs preserved the salient features of the hydrograph. Then, LFs with statistically significant slopes had the 43 years of LF data replaced by the value on the trend line (the Sen's slope). For LFs without a significant trend, the long-term LF median over the observation period was used. Following this, the synthetic time series of LFs were entered into the VAE decoder to reconstruct ADHs. As synthetic time series deliberately exclude non-significant fluctuation while preserving the first-order changes of shifting LFs, the reconstructed ADHs highlight the primary expected changes in the ADH while minimizing variation.

Based on this procedure, streams in the southern Rocky Mountains (i.e. Utah, Colorado, New Mexico) show RHEs that have substantial reduction during the spring and summer season (Fig.4.14). All of these streams had significant increases in LF4, with some having upward trends in LF6. The trends in the LFs suggest notable reduction of spring freshet and weakened snowmelt-driven signals; a reflection of the severe drought and warming temperatures over the last several decades (Williams et al., 2020).

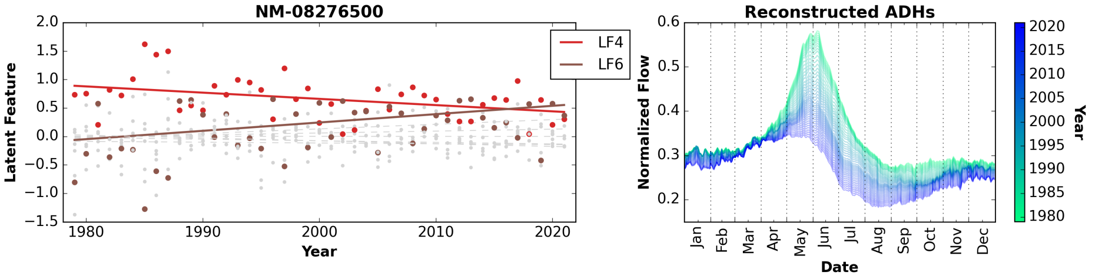


FIGURE 4.14: Trend line of LFs from 1979 to 2021 (left) and reconstructed ADHs with modified LFs (right). LFs with statistically significant trend are presented in colors, while those non-significant trend show in grey.

On the western slopes of the US Rocky Mountains, significant increases in LF8 capture the distinct patterns of advancing spring freshet on the RHE (Fig. 4.15). On the Eastern slopes, there is an increase of peak flows as a result of negative trends in LF2 (Fig. 4.16). Both earlier onset of freshet and increased peak flows were identified within the classic hydrometrics outlined in Chapter 3.

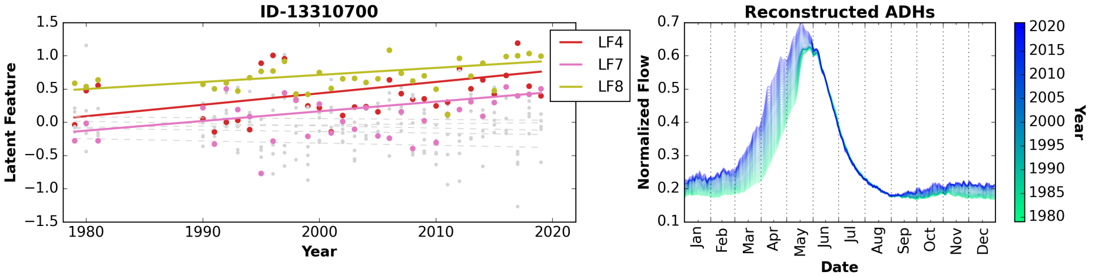


FIGURE 4.15: Trend line of LFs from 1979 to 2021 (left) and reconstructed ADHs with modified LFs (right) at ID-13310700. Statistically significant trend lines are colored, while non-significant ones are in grey.

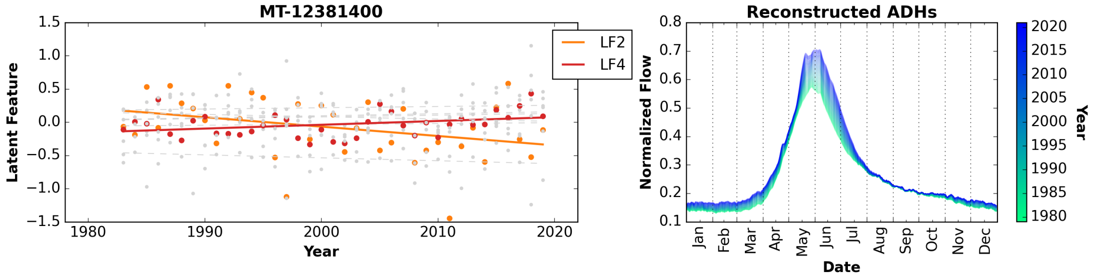


FIGURE 4.16: Trend line of LFs from 1979 to 2021 (left) and Reconstructed ADHs with modified LFs (right) at MT-12381400. Statistically significant trend lines are colored, while non-significant ones are in grey.

RHE also revealed patterns that were not identified in the classical hydrometric analysis. A diminishing spring freshet signal was observed on the RHE at a number of streams with considerable glacial coverage in the CPNW. Most of these streams had significant negative trends in LF8 and a positive trend in LF6; both of which are related to spring flow. Coopersmith et al. (2014) found that reduced snowpacks in the PNW led to lower spring peak flows, yet no significant negative trends were detected in spring monthly flows in the Chapter 3 analysis. It is possible that snowmelt-driven changes were overwhelmed by high winter rainfall events that contribute a large portion of streamflow in PNW. In contrast, the VAE-extracted features are associated with different components of streamflow and are particularly indicative of low and average flows, highlighting their sensitivities to changes in low-frequency signals. The decline in spring peaks on the RHE does suggest that reduced snowpacks in the PNW affecting spring flow.

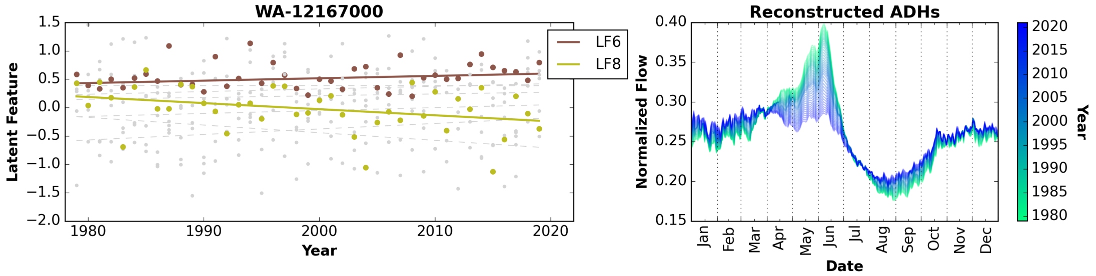


FIGURE 4.17: Trend line of LFs from 1979 to 2021 (left) and Reconstructed ADHs with modified LFs (right) at WA-12167000. Statistically significant trend lines are colored, while non-significant ones are in grey.

The decoder's ability to reconstruct ADHs from LFs provides an opportunity to project future flow regimes assuming that the trends in LF are stationary. For example, if we assume that the LF will progress along the Sen's slope, LFs can be extrapolated and the decoder can generate projected future ADHs. An example of this procedure for a stream in the Northwest Territories provided in Figure 4.18, where the RHE over the 43 year observation period indicates a late-summer secondary peak in flows. ADHs built with extrapolated LFs suggest this new peak will become more prominent over the next 20 years and there will be a sharper recession of spring freshet can be expected. While forcing hydrological models with climate projections scenarios is the most prevailing and physically based approach to prediction future streamflow (Burn et al., 2010), the data-driven technique provides a simple and fast ways to project future regimes based on current statistically significant changes in LFs.

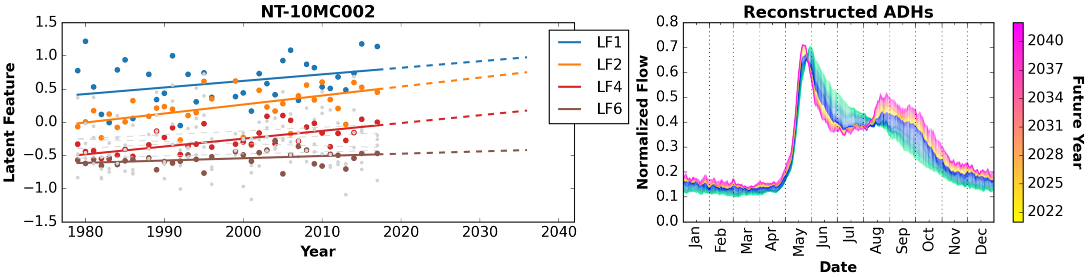


FIGURE 4.18: Trend line of LFs from 1979 to 2041 (left) and reconstructed ADHs with modified LFs (right) at NT-10MC002. Extended trend line is in dashed line. ADHs built with observed LFs are in green-blue colors, and those with extrapolated values are in yellow-magenta colors.

4.5.2 Potential Use in Flow Regime Classification

A subset samples of ADHs were labelled using the t-SNE map developed in Chapter 2 that represent seven different flow regimes in WNA: three snow-dominated (Class 1-3), one mixed (Class 4), and two rain-dominated regimes (Class 5 and 6) (see Fig. 4.19). Results indicate that samples from each of the flow regimes for a distinct joint distribution in latent feature space. ADHs from rain-dominated classes typically had a positive LF6, while those from the snow-dominated classes had a negative value. As a mixed flow regime, ADHs in Class 4 exhibited both winter rainfall events and distinct spring freshet, and as a consequence their LF6 values remained close to zero. Among the four snow-regime classes (also including Class 4), the timing of the spring freshet became gradually later, and their LF8 progressed from low to high values. In comparison, the ADHs in Class 3 and Class 4 (mostly from southern Idaho) have a more compressed high-flow signal and flows that occur earlier, also resulting in a larger LF8 value. With respect to the magnitude of peak flow, LF2 is an effective indicator in snow dominated regimes. Class 1 and 2 include ADHs from high elevation and high latitudes, and typically exhibit an extended period of freshet-driven flows. Although having a similar ADH shape, Class 1 has a higher LF4 than Class 2. LF4 is an indicator of the total streamflow magnitude, which is greater overall in Class 1. In contrast, Class 6, which consists mostly of ADHs from Southern California, is the driest flow regime and has the lowest LF4. Between the two rainfall-dominated classes (Class 5 and 6), the most distinct difference is that Class 5 had higher flow during early winter than Class 6. This is reflected in LF7, which is negatively related to November and December flows for rainfall-dominated streams, with LF7 much lower in Class 5 than Class 6.

A parallel coordinates plot indicate that ADHs from different flow regimes have unique patterns in LFs (Fig. 4.19). Embedded data, if sufficiently preserving the salient features from the original dataset, are often favorable because they greatly improve the efficiency of classification (Goodfellow et al., 2016). As low-dimension representation of ADHs, LFs could potentially be used for a broader range of hydrological classification on larger streamflow datasets.

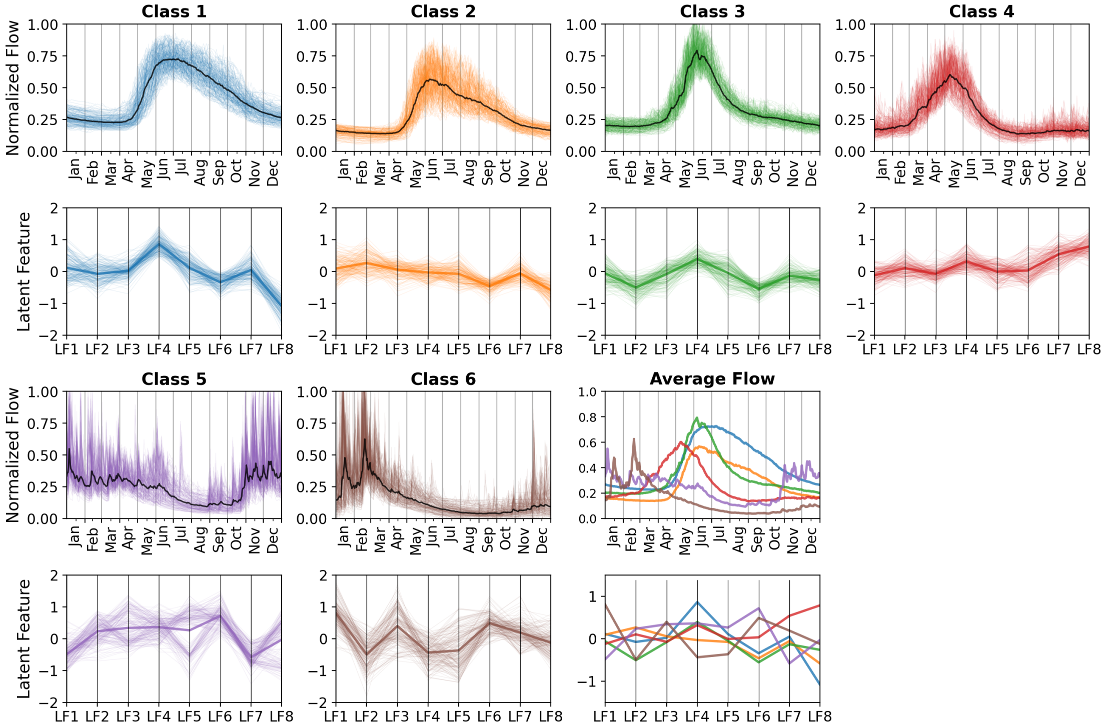


FIGURE 4.19: ADHs and parallel coordinates plot of six flow regimes.

4.5.3 Limitation of VAE

Here we present a data driven methodology and a DL algorithm to learn and extract the most salient features from ADHs from Western North American flow regimes. This VAE approach stands in contrast to manually designing and inspecting classical hydrometrics to characterize changes in streamflow regimes. Novel machine-learned features (LFs) provide an opportunity to characterize changes in flow regimes from a new perspective that is not limited to previous human-designed features. Applying a decoder, it is possible to reconstruct ADHs with any given LFs, so detected and expected changes of LFs can be directly projected on the shape of the AHDs. However, despite significant changes in LFs that can be detected, we believe it is most appropriate to use LFs as a general indicator of shifting flow directions as opposed to quantitative measures of changing magnitude. As the transformation from AHDs to LFs is highly non-linear, there is a lack of firm connection between LFs and specific hydrological processes. Our current interpretation of LFs, which is based on Spearman's correlation with hydrometrics and visual examination of RHEs, is plausible but not adequate to establish connections between LFs and changes in hydrological processes resulting in streamflow shifts. We encourage future efforts to bridge this gap.

The VAE provided several useful LF indicators, most notably LF6 that easily distinguished snow and rain-dominated flow regimes. However, this indicator is specific to this data set, the streams in Western North America. Retraining with new data is necessary to further explore and expand this VAE approach.

The VAE excelled in learning the general ADH pattern, yet typically failed to copy high, short-lived events. As a consequence, the current VAE model is unsuitable for characterizing events such as intermittent or sudden floods. However, it is possible to enhance the learning capability of VAE to extremes in ADHs by making the loss function more sensitive to errors in high/low values. Another potential approach is to choose a different normalization with a weaker distortion of high values than the square root. Overall, it is possible to adjust the normalization and loss function of the VAE based on different research objectives.

Finally, it is important to note that LF features are determined objectively. Hydrometric indicators are often designed with very specific human or societal implications (e.g. lenghts of low flows, high flows, etc.), and as a consequence their utility in an operational way may be limited.

4.6 Conclusion

In this work, we present an approach using deep learning to identify features of streamflow change in Western North America. Latent features were extracted using VAE models; five of which had direct hydrological implications. Furthermore, LFs were found to be closely associated with a variety of classical hydrometric indicators such as flow quantity, seasonal distribution, timing and magnitude of freshet, and snow-to-rain transition. Trends in LFs over the 43 years period show direct agreement with associated hydrometric trends outlined in Chapter 3. Through utilization of the LFs, the VAE decoder provides an intuitive way to visualize the evolution of streamflow over time, as temporal changes in the LFs are projected on the shape of the ADH. This also provides a mechanism to estimate future changes in streamflow regimes by extrapolating LF trends forward in time. Furthermore, as ADHs from different flow regimes form distinct joint distributions in latent feature space, LFs have considerable potential for streamflow classification and flow regime regionalization. As data-driven approaches become more common in hydrological analyses, we believe that VAE is a useful tool to learn features of raw flow data that may be absent in traditional hydrometric analyses.

Chapter 5

Conclusion

Over the past century, changes to streamflow timing and volume have been widely observed across Western North America (WNA). These changes are driven in part by natural climate variability and anthropogenic climate change. This thesis advances on previous work by identifying the most recently streamflow changes in WNA using a comprehensive suite of classical hydrometric methods along with novel Deep Learning (DL) based approaches. The period of interest was the most recent 43 years (1979 to 2021) and included more than 500 streams in Canada and the United States. Both reference hydrometric and DL approaches suggest streamflow regimes have significantly shifted at most WNA streams over the past four decades, yet changes in magnitude, timing and direction exhibit considerable spatial dependency.

Results in Chapter 3 indicate that the most consistent change across the WNA is a decline in streamflow during late summer. Statistically significant negative trends were identified for August and September at almost all middle and low-latitude streams. Glaciers provided some buffer against these declining summer flows, even when their areal coverage was relatively small. An approximate dividing line of at $46°N$ latitude existed, where streams north of this line had increased overall water availability, while streams south most often dried. As changing hydrometrics have considerable spatial dependency, streams were divided into six regions for analysis. Streams in the Northwest domain had consistent increase in winter (and therefore total) flow; an observation previously reported in the literature and ascribed to degrading permafrost. In the Canadian Rocky Mountains, the proportion of spring flow increased whereas summer flows diminished. Declines in summer flows were also observed in the Coastal Pacific Midwest along with prolonged dry periods. In the Coastal Pacific Northwest, a distinct change was an increase in autumn flows, which is related to substantial increases in October precipitation during the study period. Across the Southwest, streamflows exhibited a dramatic decline in water availability and field significant negative trends were observed for all magnitude-related hydrometrics (e.g. monthly flows and flow percentiles). On average, annual flows declined by 45% over the last 43 years across this region, underscoring the water emergency in this region.

DL techniques provide a powerful tool to extract and process information from streamflow data. In this thesis, two DL-based models, parametric t-SNE and VAE have been developed and trained. The t-SNE (Chapter 2) proves and intuitive approach to visualize similarity among a large number of hydrographs on a 2-D map, and was able to outperform conventional clustering techniques such as PCA and correlation matrices. The t-SNE was able to identify and classify hydrographs rapidly, and with a DL-based encoder network was able to project new samples of ADHs on the existing map; a new parametric approach. As a diagnostic tool, t-SNE allows the rapid identification of an ADH flow regime of a new hydrograph based on their neighbours hydrological and geographical identities. Potential applications of this approach are hydrograph reconstruction for ungauged watersheds or identifying information redundancy in monitoring networks.

The DL VAE model (Chapter 4) was developed and trained to automatically learn and extract latent features (LFs) from raw streamflow data, which were then used for trend analysis. This provided an opportunity to assess change in WNA flow regimes form an alternate DL perspective. This DL approach has two primary advantages to classical hydrometric analysis: 1) LFs are free from information redundancy and present distinct aspects of streamflow characteristics, and 2) LF can represent complex changing patterns efficiently, overcoming the challenge of selecting from dozens of classical hydrometric indicators.

Understanding and interpreting LFs was achieved by: 1) relating them to classical hydrometrics that have well understood meaning using Spearman's rank correlation coefficient, and 2) tuning the LFs and observing the shape evolution of reconstructed ADHs. In terms of the relation between LFs and hydrometrics, some were more interpretable than others. For example, LF6 was closely associated with streamflow seasonal distribution and represented snow-to-rain transition, which would typically require multiple hydrometrics to describe. LF4 was an indicator of magnitude-related hydrometrics and describe mean annual flows and all flow percentiles, yet correlation with high flows were weaker. Some LF were only interpretable for certain flow regimes (e.g. rain vs snow). LF2 was highly correlated with the magnitude of spring freshet and LF8 freshet timing, yet only for snow dominated streams.

Trend analysis of the LFs revealed patterns that were consistent with the hydrometric analysis. Positive trends in LF4 at streams in the Northwest domain and negative trends in the Southwest supported the increases/decreases observed in those regions and was a sensitive indicator of streamflow amount. A forward shifting of streamflow observed in the Rocky Mountains was supported by the positive trends in LF8. In addition, LFs revealed changes that were obscured from traditional hydrometrics. Strong positive trends in LF8 at streams in the southern Rocky Mountains imply snow-to-rain transitions of the streamflow regime; possibly related to declining winter snow accumulation. Furthermore, significant increases in LF7 for the Pacific coastal regions suggest that the relatively northern streams (in BC, Washington and Oregon) have been shifting towards a flow regime closer to those in northern California that have less flow in early winter.

With the VAE decoder, a new method was proposed to reconstruct future ADHs by using statistically significant changes in LFs. By tuning LFs along their trendline (i.e. Sen's slope) provided an intuitive way to project changes in the hydrograph based on the trends in the machine learned features. While this approach has limitations in that it assumes stationarity in the change magnitude and direction, it provides a useful data-driven approach to visualize change.

5.1 Limitations

As with all DL and comparative approaches, there were subjective choices that were made that influence the interpretation of the analysis. In Chapter 3, streams were divided into six regions with respect to their geographic location. While this was a pragmatic approach that allowed the evaluation of field-significance, there are alternate ways of grouping streams as evidence in Chapters 2 and 4. To further assess the significance of change in a region, alternate clustering and classification regimes would result in different outcomes, yet the overall hydrometric analysis on given streams would remain unchanged.

Despite its potential for classification, the t-SNE in Chapter 2 is based purely on ADH shape and ignores underlying hydrological processes and biophysical setting. In addition, the t-SNE model was developed and trained exclusively on WNA streams, and to apply this further it must be retrained for 'unseen' flow regimes.

The VAE model in Chapter 4 is poor at recognizing features related to short lived intermittent events. In most cases, ADHs reconstructed with LFs could copy the general flow pattern, yet fits were poor for catchments with transient rainfall-driven events (typically in the Southwest). This limits the VAE model, and its poor performance for extreme events may be attributed to the choice of data normalization which alternate methods may improve. In addition, the physical meaning and linkage between LF and hydrometrics is dependent on flow regimes, and the grouping of rain and snow dominated groups based on LF6 was in part subjective.

5.2 Recommendation for Future Study

This thesis provides a comprehensive and up-to-date overview of streamflow changes in WNA and a general framework for streamflow change detection that could be applicable across other geographical regions. North America, Europe, Australia, and China are all prospective candidates, where national- or continental-scale stream gauging network have been well established and operated for decades and where this approach can be applied.

Climate change and variability will continue to reshape streamflow around the world, and extending the observation period with newly collected streamflow data, the framework and supplementary dashboard app will provide consistent updating of trends and patterns in the domains of interest. In addition, it is possible to add new hydrometrics and hydro-climate indices to the dashboard to help water resource management decisions.

While this thesis has highlighted hotspots of streamflow change across WNA, there is still uncertainty in the underlying mechanisms that are driving this change and in how watersheds respond to change. While ERA5 data was used to assess links with changes in precipitation and temperature, there is considerable nuance in understanding the drivers and response to change that are beyond the scope of this thesis. For example, geology, soil, permafrost topography and vegetation all strongly influence hydrometric response, and process-based mechanisms remain critical for understanding future hydrological regimes.

As a member in the family of generative models, VAE is able to create an unlimited number of synthetic hydrographs that closely resemble training samples and their variation. The generation of large statistically meaningful synthetic data sets can benefit a range of hydrological modeling applications. In addition, more research on LFs and their joint distribution will help generate flow regimes that can be used for machine learning models that require labelled samples for training and testing purpose, alleviating the need for large empirical data sets.

Appendix A

Chapter 2 Supplement

TABLE A1.1: Model configuration of tested encoders

Model	Architecture	Activation	MSE (Train)	MSE (Test)
Encoder 1	[128]	ReLU	5.811	8.307
Encoder 2	[128, 64]	ReLU	2.714	6.843
Encoder 3	[128, 64, 32]	ReLU	2.170	6.124
Encoder 4	[128, 64, 32, 16]	ReLU	2.114	5.953
Encoder 5	[128, 64, 32, 16, 8]	ReLU	2.168	5.944
Encoder 6	[128, 64, 32, 16, 8, 4]	ReLU	2.182	5.955
Encoder 7	[256]	ReLU	4.226	7.962
Encoder 8	[256, 128]	ReLU	1.755	6.358
Encoder 9	[256, 128, 64]	ReLU	1.215	5.435
Encoder 10	[256, 128, 64, 32]	ReLU	1.137	5.187
Encoder 11	[256, 128, 64, 32, 16]	ReLU	1.212	5.139
Encoder 12	[256, 128, 64, 32, 16, 8]	ReLU	1.236	5.173
Encoder 13	[256, 128, 64, 32, 16, 8, 4]	ReLU	1.242	5.145
Encoder 14	[512]	ReLU	3.080	7.849
Encoder 15	[512, 256]	ReLU	1.214	5.913
Encoder 16	[512, 256, 128]	ReLU	0.846	4.946
Encoder 17	[512, 256, 128, 64]	ReLU	0.839	4.667
Encoder 18	[512, 256, 128, 64, 32]	ReLU	0.845	4.627
Encoder 19	[512, 256, 128, 64, 32, 16]	ReLU	0.862	4.619
Encoder 20	[512, 256, 128, 64, 32, 16, 8]	ReLU	0.854	4.579
Encoder 21	[512, 256, 128, 64, 32, 16, 8, 4]	ReLU	0.894	4.641
Encoder 22	[1024]	ReLU	2.521	7.715
Encoder 23	[1024, 512]	ReLU	1.048	5.610

Continued on next page

Model	Architecture	Activation	MSE (Train)	MSE (Test)
Encoder 24	[1024, 512, 256]	ReLU	0.751	4.676
Encoder 25	[1024, 512, 256, 128]	ReLU	0.706	4.428
Encoder 26	[1024, 512, 256, 128, 64]	ReLU	0.726	4.353
Encoder 27	[1024, 512, 256, 128, 64, 32]	ReLU	0.744	4.393
Encoder 28	[1024, 512, 256, 128, 64, 32, 16]	ReLU	0.748	4.337
Encoder 29	[1024, 512, 256, 128, 64, 32, 16, 8]	ReLU	0.733	4.333
Encoder 30	[1024, 512, 256, 128, 64, 32, 16, 8, 4]	ReLU	0.788	4.357
Encoder 31	[2048]	ReLU	2.266	7.651
Encoder 32	[2048, 1024]	ReLU	0.963	5.457
Encoder 33	[2048, 1024, 512]	ReLU	0.701	4.563
Encoder 34	[2048, 1024, 512, 256]	ReLU	0.654	4.324
Encoder 35	[2048, 1024, 512, 256, 128]	ReLU	0.632	4.349
Encoder 36	[2048, 1024, 512, 256, 128, 64]	ReLU	0.633	4.286
Encoder 37	[2048, 1024, 512, 256, 128, 64, 32]	ReLU	0.677	4.305
Encoder 38	[2048, 1024, 512, 256, 128, 64, 32, 16]	ReLU	0.693	4.237
Encoder 39	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	ReLU	0.682	4.197
Encoder 40	[2048, 1024, 512, 256, 128, 64, 32, 16, 8, 4]	ReLU	0.680	4.223
Encoder 41	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.4)	1.689	4.879
Encoder 42	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.3)	1.271	4.459
Encoder 43	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.25)	1.140	4.296
Encoder 44	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.2)	1.069	4.184
Encoder 45	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.15)	0.931	4.019
Encoder 46	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.12)	0.859	3.941
Encoder 47	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.1)	0.842	3.891
Encoder 48	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.09)	0.846	3.893

Continued on next page

Model	Architecture	Activation	MSE (Train)	MSE (Test)
Encoder 49	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.08)	0.791	3.832
Encoder 50	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.07)	0.731	3.806
Encoder 51	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.06)	0.770	3.816
Encoder 52	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.05)	0.712	3.801
Encoder 53	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.04)	0.727	3.851
Encoder 54	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.03)	0.715	3.928
Encoder 55	[2048, 1024, 512, 256, 128, 64, 32, 16, 8]	LeakyReLU(0.02)	0.715	4.226

Appendix B

Chapter 3 Supplement

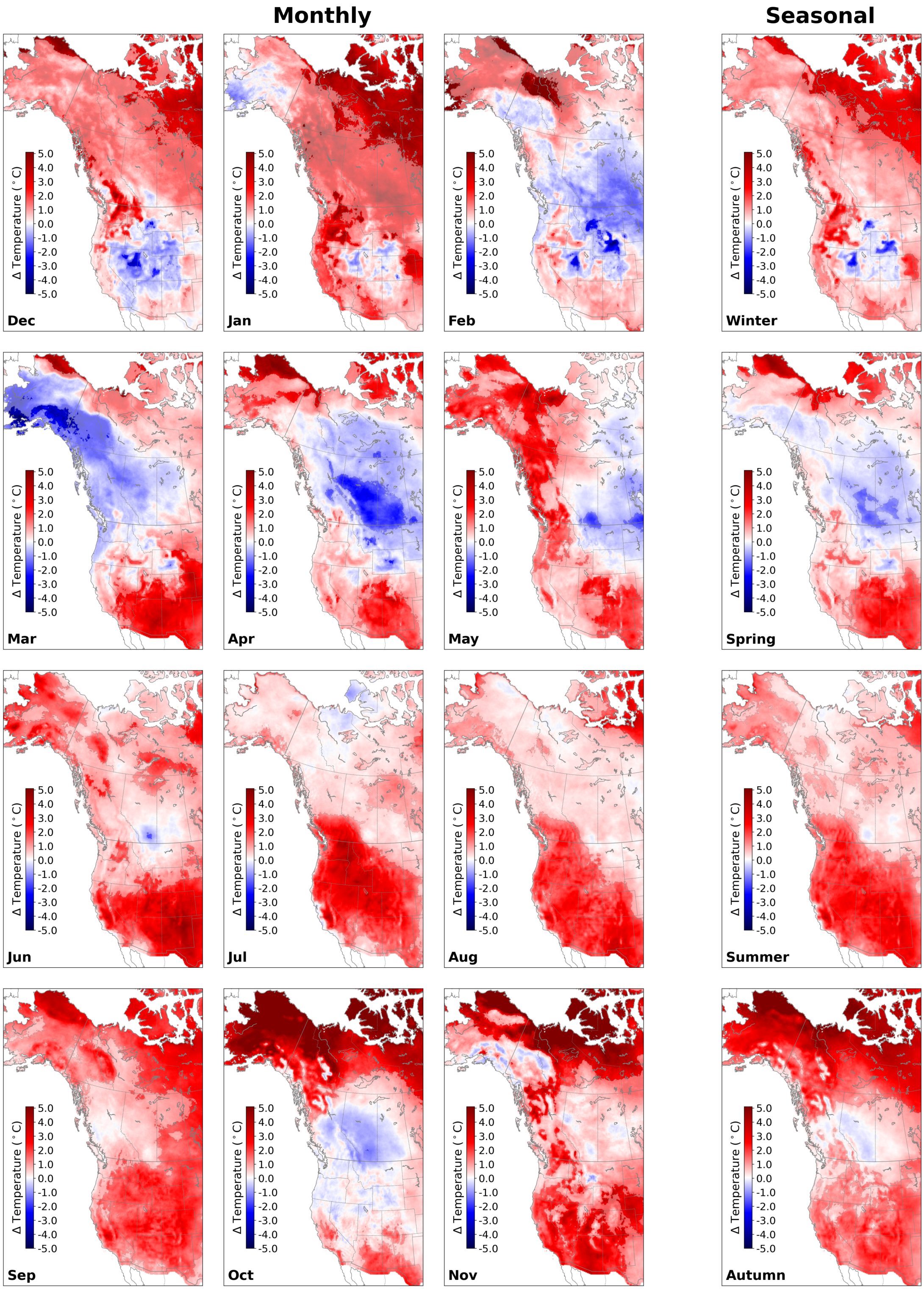


FIGURE A2.1: Maps of trends in monthly and seasonal air temperature across WNA. Non-significant changes (p-value>0.1) are masked with a 50% transparent layer of white color.

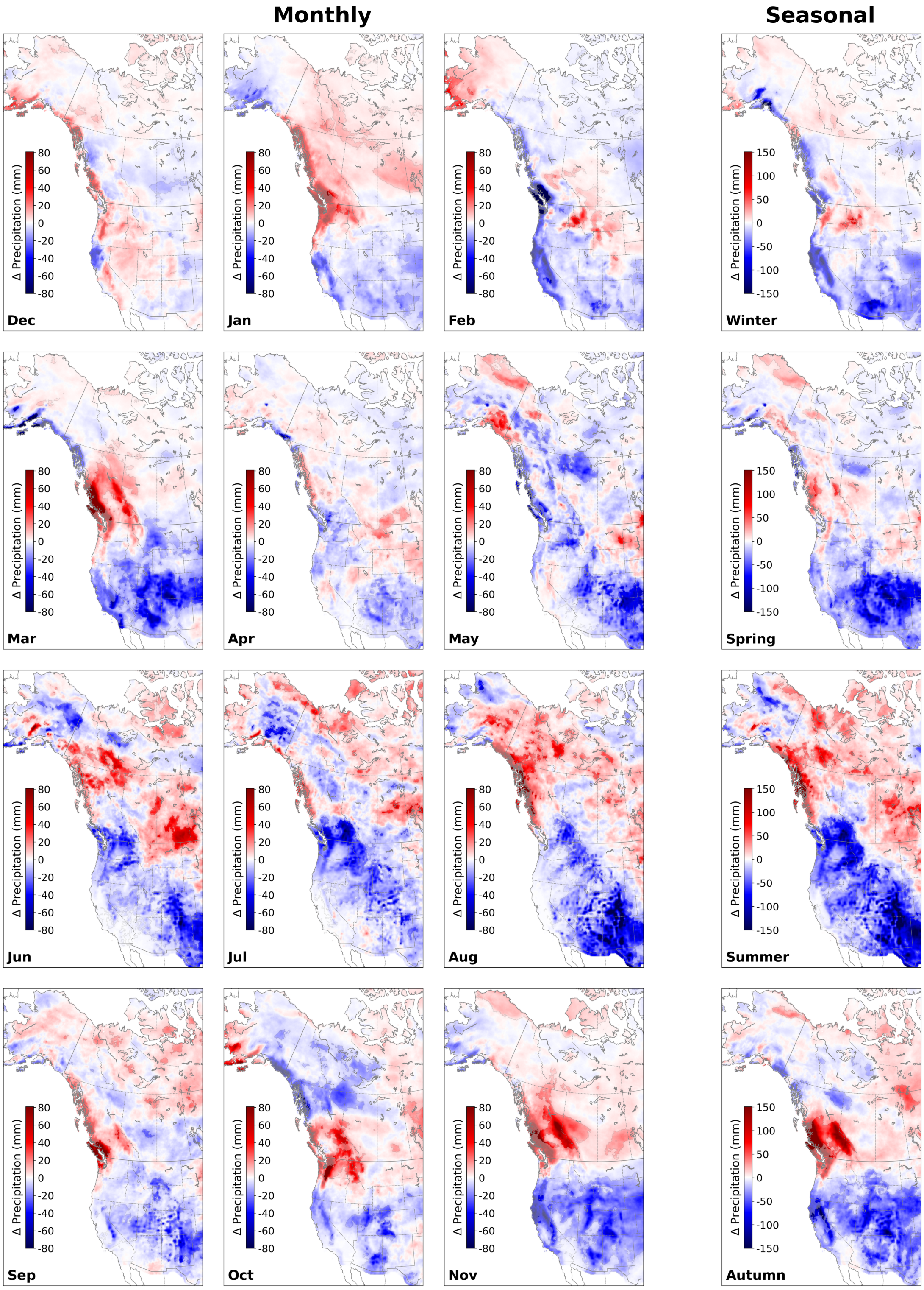


FIGURE A2.2: Maps of trends in monthly and seasonal precipitation across WNA.

TABLE A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
05AA008	CROWSNEST RIVER AT FRANK	AB	49.597321	-114.410640	403.00	RHBN
05AA023	OLDMAN RIVER NEAR WALDRON'S CORNER	AB	49.813889	-114.183327	1446.10	RHBN
05AD003	WATERTON RIVER NEAR WATERTON PARK	AB	49.113659	-113.839400	613.00	RHBN
05AD005	BELLY RIVER NEAR MOUNTAIN VIEW	AB	49.099602	-113.697723	319.00	RHBN
05BA002	PIPESTONE RIVER NEAR LAKE LOUISE	AB	51.433109	-116.174950	306.00	RHBN
05BB001	BOW RIVER AT BANFF	AB	51.172230	-115.571770	2210.00	RHBN
05BL022	CATARACT CREEK NEAR FORESTRY ROAD	AB	50.285259	-114.588710	166.00	RHBN
05CA009	RED DEER RIVER BELOW BURNT TIMBER CREEK	AB	51.646252	-115.019250	2250.00	RHBN
05DA007	MISTAYA RIVER NEAR SASKATCHEWAN CROSSING	AB	51.884270	-116.689190	248.00	RHBN
05DA009	NORTH SASKATCHEWAN RIVER AT WHIRLPOOL POINT	AB	52.001190	-116.470930	1920.00	RHBN
05DA010	SILVERHORN CREEK NEAR THE MOUTH	AB	51.799919	-116.583950	21.00	RHBN
07AA001	MIETTE RIVER NEAR JASPER	AB	52.864120	-118.107170	629.00	RHBN
07AA002	ATHABASCA RIVER NEAR JASPER	AB	52.910191	-118.058700	3870.00	RHBN
07AG003	WOLF CREEK AT HIGHWAY NO. 16A	AB	53.598351	-116.271840	826.00	RHBN
07BB002	PEMBINA RIVER NEAR ENTWISTLE	AB	53.604191	-115.004740	4400.00	RHBN
07BJ001	SWAN RIVER NEAR KINUSO	AB	55.315540	-115.417170	1900.00	RHBN
07GG001	WASKAHIGAN RIVER NEAR THE MOUTH	AB	54.752010	-117.205700	1040.00	RHBN
15024800	STIKINE R NR WRANGELL AK	AK	56.708056	-132.130278	25692.80	HCDN
15052500	MENDENHALL R NR AUKE BAY AK	AK	58.429722	-134.572778	220.41	HCDN
15072000	FISH C NR KETCHIKAN AK	AK	55.391944	-131.193889	83.14	HCDN
15085100	OLD TOM C NR KASAAN AK	AK	55.395194	-132.408627	15.28	HCDN
15129500	SITUK R NR YAKUTAT AK	AK	59.586366	-139.494188	93.14	HCDN
15239050	MF BRADLEY R NR HOMER AK	AK	59.777746	-150.756295	23.70	HCDN
15258000	KENAI R AT COOPER LANDING AK	AK	60.492778	-149.807778	1642.06	HCDN
15266300	KENAI R AT SOLDOTNA AK	AK	60.477500	-151.079444	5205.90	HCDN
15290000	L SUSITNA R NR PALMER AK	AK	61.708889	-149.226667	160.32	HCDN
15292000	SUSITNA R AT GOLD CREEK AK	AK	62.767778	-149.691111	15954.40	HCDN
15292700	TALKEETNA R NR TALKEETNA AK	AK	62.346944	-150.016944	5195.54	HCDN
15302000	NUYAKUK R NR DILLINGHAM AK	AK	59.935556	-158.187778	3859.10	HCDN
15304000	KUSKOKWIM R AT CROOKED CREEK AK	AK	61.871111	-158.100833	2849.00	HCDN
15356000	YUKON R AT EAGLE AK	AK	64.789444	-141.197778	9065.00	HCDN
15484000	SALCHA R NR SALCHAKET AK	AK	64.472778	-146.923889	5620.30	HCDN
15493000	CHENA R NR TWO RIVERS AK	AK	64.902374	-146.359366	2419.00	HCDN
15511000	L CHENA R NR FAIRBANKS AK	AK	64.885690	-147.249648	886.71	HCDN
15514000	CHENA R AT FAIRBANKS AK	AK	64.845833	-147.701111	5128.20	HCDN
15515500	TANANA R AT NENANA AK	AK	64.565278	-149.091667	14504.00	HCDN
15565447	YUKON R AT PILOT STATION AK	AK	61.934444	-162.880556	2590.00	HCDN
15744500	KOBUK R NR KIANA AK	AK	66.973611	-160.130833	24656.80	HCDN
15896000	KUPARUK R NR DEADHORSE AK	AK	70.281667	-148.959722	8106.70	HCDN
15908000	SAGAVANIRKTOK R NR PUMP STA 3 AK	AK	69.014648	-148.820034	4821.85	HCDN
09384000	LITTLE COLORADO R ABV LYMAN LAKE NR ST. JOHNS,	AZ	34.314444	-109.361667	1934.73	HCDN
09402000	LITTLE COLORADO RIVER NEAR CAMERON, ARIZ.	AZ	35.926389	-111.566667	16835.00	HCDN
09415000	VIRGIN RIVER AT LITTLEFIELD, ARIZ.	AZ	36.891667	-113.923611	13183.10	HCDN
09444500	SAN FRANCISCO RIVER AT CLIFTON, ARIZ.	AZ	33.049444	-109.295278	7163.94	HCDN
09447800	BONITA CREEK NEAR MORENCI, AZ.	AZ	32.955618	-109.531188	781.96	HCDN
09448500	GILA R AT HEAD OF SAFFORD VALLEY NR SOLOMON ARI	AZ	32.868333	-109.510556	20450.64	HCDN
09471000	SAN PEDRO RIVER AT CHARLESTON, ARIZ.	AZ	31.625833	-110.173889	3157.21	HCDN
09480000	SANTA CRUZ RIVER NEAR LOCHIEL, ARIZ.	AZ	31.355278	-110.588889	212.90	HCDN

Continued on next page

TABLE A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
09484000	SABINO CREEK NEAR TUCSON, AZ.	AZ	32.316742	-110.810367	103.65	HCDN
09484600	PANTANO WASH NEAR VAIL, AZ.	AZ	32.035914	-110.677582	1179.59	HCDN
09492400	EAST FORK WHITE RIVER NEAR FORT APACHE, AZ.	AZ	33.822272	-109.814541	128.97	HCDN
09494000	WHITE RIVER NEAR FORT APACHE, AZ.	AZ	33.736441	-110.166767	1628.23	HCDN
09497800	CIBECUE CREEK NEAR CHYSOTILE, AZ.	AZ	33.843105	-110.557609	750.86	HCDN
09497980	CHERRY CREEK NEAR GLOBE, AZ.	AZ	33.827826	-110.856227	516.80	HCDN
09498500	SALT RIVER NEAR ROOSEVELT, ARIZ.	AZ	33.619444	-110.920833	11152.54	HCDN
09505200	WET BEAVER CREEK NEAR RIMROCK, AZ.	AZ	34.674744	-111.672094	285.68	HCDN
09505350	DRY BEAVER CREEK NEAR RIMROCK, AZ.	AZ	34.728631	-111.775708	365.66	HCDN
09505800	WEST CLEAR CREEK NEAR CAMP VERDE, AZ.	AZ	34.538636	-111.694036	614.94	HCDN
09508300	WET BOTTOM CREEK NR CHILDS, ARIZ.	AZ	34.160833	-111.692222	94.28	HCDN
09510200	SYCAMORE CREEK NEAR FORT MCDOWELL, AZ.	AZ	33.694211	-111.541802	425.31	HCDN
09512280	CAVE CREEK BLW COTTONWOOD CR NEAR CAVE CREEK, AZ.	AZ	33.887260	-111.954039	188.71	HCDN
09513780	NEW RIVER NEAR ROCK SPRINGS, AZ.	AZ	33.974202	-112.099046	177.47	HCDN
09535100	SAN SIMON WASH NEAR PISINIMO, AZ.	AZ	32.044237	-112.370971	1482.93	HCDN
07EC002	OMINECA RIVER ABOVE OSILINKA RIVER	BC	55.916859	-124.567580	5560.00	RHBN
07EE007	PARSNIP RIVER ABOVE MISINCHINKA RIVER	BC	55.081940	-122.913060	4930.00	RHBN
07EE009	CHUCHINKA CREEK NEAR THE MOUTH	BC	54.529720	-122.612220	310.00	RHBN
07FB001	PINE RIVER AT EAST PINE	BC	55.718441	-121.211560	12100.00	RHBN
07FB004	DICKEBUSCH CREEK NEAR THE MOUTH	BC	55.537781	-121.596810	82.40	RHBN
07FB008	MOBERLY RIVER NEAR FORT ST. JOHN	BC	56.092499	-121.366780	1520.00	RHBN
07FC003	BLUEBERRY RIVER BELOW AITKEN CREEK	BC	56.677639	-121.222330	1770.00	RHBN
08CD001	TUYA RIVER NEAR TELEGRAPH CREEK	BC	58.072250	-130.826450	3550.00	RHBN
08CG001	ISKUT RIVER BELOW JOHNSON RIVER	BC	56.734440	-131.668990	9500.00	RHBN
08DA005	SURPRISE CREEK NEAR THE MOUTH	BC	56.109341	-129.477400	218.00	RHBN
08DB001	NASS RIVER ABOVE SHUMAL CREEK	BC	55.262249	-129.085040	18400.00	RHBN
08EB004	KISPIOX RIVER NEAR HAZELTON	BC	55.433849	-127.716160	1880.00	RHBN
08ED001	NANIKA RIVER AT OUTLET OF KIDPRICE LAKE	BC	53.930328	-127.451690	732.00	RHBN
08EE008	GOATHORN CREEK NEAR TELKWA	BC	54.648609	-127.123330	125.00	RHBN
08EE012	SIMPSON CREEK AT THE MOUTH	BC	54.809940	-127.204160	13.20	RHBN
08EE025	TWO MILE CREEK IN DISTRICT LOT 4834	BC	55.296108	-127.620750	21.20	RHBN
08EG012	EXCHAMSIKS RIVER NEAR TERRACE	BC	54.362190	-129.312840	370.00	RHBN
08FB006	ATNARKO RIVER NEAR THE MOUTH	BC	52.360081	-126.005860	2550.00	RHBN
08FE003	KEMANO RIVER ABOVE POWERHOUSE TAILRACE	BC	53.563549	-127.952360	556.00	RHBN
08FF003	LITTLE WEDEENE RIVER BELOW BOWBYES CREEK	BC	54.136391	-128.690000	180.00	RHBN
08GA010	CAPILANO RIVER ABOVE INTAKE	BC	49.396240	-123.145843	173.00	RHBN
08GA061	MACKAY CREEK AT MONTROYAL BOULEVARD	BC	49.356159	-123.099861	3.63	RHBN
08GD004	HOMATHKO RIVER AT THE MOUTH	BC	50.986359	-124.918640	5680.00	RHBN
08HA001	CHEMAINUS RIVER NEAR WESTHOLME	BC	48.878361	-123.704640	355.00	RHBN

Continued on next page

TABLE A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
08HA003	KOKSILAH RIVER AT COWICHAN STATION	BC	48.727829	-123.670810	209.00	RHBN
08HB002	ENGLISHMAN RIVER NEAR PARKSVILLE	BC	49.316109	-124.285271	319.00	RHBN
08HB008	SPROAT RIVER NEAR ALBERNI	BC	49.289719	-124.910278	351.00	RHBN
08HB025	BROWNS RIVER NEAR COURTENAY	BC	49.692829	-125.069250	87.90	RHBN
08HB075	DOVE CREEK NEAR THE MOUTH	BC	49.736580	-125.085720	41.10	RHBN
08HF004	TSITIKA RIVER BELOW CATHERINE CREEK	BC	50.440361	-126.578510	365.00	RHBN
08JB002	STELLAKO RIVER AT GLENANNAN	BC	54.007801	-125.008920	3600.00	RHBN
08JD006	DRIFTWOOD RIVER ABOVE KASTBERG CREEK	BC	55.976139	-126.677670	403.00	RHBN
08JE001	STUART RIVER NEAR FORT ST. JAMES	BC	54.416859	-124.270620	14200.00	RHBN
08JE004	TSILCOH RIVER NEAR THE MOUTH	BC	54.610809	-124.247760	431.00	RHBN
08KA009	MCKALE RIVER NEAR 940 M CONTOUR	BC	53.444061	-120.220140	253.00	RHBN
08KB003	MCGREGOR RIVER AT LOWER CANYON	BC	54.231171	-121.668750	4780.00	RHBN
08KE016	BAKER CREEK AT QUESNEL	BC	52.973888	-122.509440	1550.00	RHBN
08LA001	CLEARWATER RIVER NEAR CLEARWATER STATION	BC	51.655720	-120.066050	10300.00	RHBN
08LB076	HARPER CREEK NEAR THE MOUTH	BC	51.354382	-119.881640	166.00	RHBN
08LD001	ADAMS RIVER NEAR SQUILAX	BC	50.937592	-119.656340	3210.00	RHBN
08LG016	PENNASK CREEK NEAR QUILCHENA	BC	49.964920	-120.134930	87.60	RHBN
08LG056	GUICHON CREEK ABOVE TUNKWA LAKE DIVERSION	BC	50.607830	-120.910860	78.20	RHBN
08MA002	CHILKO RIVER AT OUTLET OF CHILKO LAKE	BC	51.624859	-124.143360	2130.00	RHBN
08MA006	LINGFIELD CREEK NEAR THE MOUTH	BC	51.673859	-124.145310	98.80	RHBN
08MB006	BIG CREEK ABOVE GROUNDHOG CREEK	BC	51.523689	-123.115890	1010.00	RHBN
08MF062	COQUIHALLA RIVER BELOW NEEDLE CREEK	BC	49.541889	-121.119970	85.50	RHBN
08MF065	NAHATLATCH RIVER BELOW TACHEWANA CREEK	BC	49.953861	-121.862870	712.00	RHBN
08MG001	CHEHALIS RIVER NEAR HARRISON MILLS	BC	49.300030	-121.937770	383.00	RHBN
08MG005	LILLOOET RIVER NEAR PEMBERTON	BC	50.336021	-122.800250	2100.00	RHBN
08MH001	CHILLIWACK RIVER AT VEDDER CROSSING	BC	49.097382	-121.967480	1230.00	RHBN
08MH006	NORTH ALOUETTE RIVER AT 232ND STREET, MAPLE RIDGE	BC	49.242649	-122.580180	37.30	RHBN
08MH016	CHILLIWACK RIVER AT OUTLET OF CHILLIWACK LAKE	BC	49.083649	-121.458440	335.00	RHBN
08NB005	COLUMBIA RIVER AT DONALD	BC	51.483299	-117.180390	9700.00	RHBN
08NC004	CANOE RIVER BELOW KIMMEL CREEK	BC	52.731529	-119.384860	305.00	RHBN
08ND013	ILLECILLEWAET RIVER AT GREELEY	BC	51.012630	-118.085140	1150.00	RHBN
08NE006	KUSKANAX CREEK NEAR NAKUSP	BC	50.282909	-117.734480	330.00	RHBN
08NE077	BARNES CREEK NEAR NEEDLES	BC	49.905182	-118.125470	204.00	RHBN
08NE114	HIDDEN CREEK NEAR THE MOUTH	BC	49.234539	-117.239170	56.70	RHBN
08NF001	KOOTENAY RIVER AT KOOTENAY CROSSING	BC	50.887032	-116.046100	416.00	RHBN
08NH005	KASLO RIVER BELOW KEMP CREEK	BC	49.907619	-116.953490	442.00	RHBN
08NH016	DUCK CREEK NEAR WYNNDEL	BC	49.202629	-116.533910	57.00	RHBN
08NH084	ARROW CREEK NEAR ERICKSON	BC	49.159119	-116.452490	78.30	RHBN
08NH130	FRY CREEK BELOW CARNEY CREEK	BC	50.082001	-116.784860	585.00	RHBN
08NJ013	SLOCAN RIVER NEAR CRESCENT VALLEY	BC	49.460079	-117.564800	3330.00	RHBN
08NJ130	ANDERSON CREEK NEAR NELSON	BC	49.501949	-117.261280	9.07	RHBN
08NL004	ASHNOLA RIVER NEAR KEREMEOS	BC	49.207630	-119.993520	1050.00	RHBN

Continued on next page

Table A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
08NL007	SIMILKAMEEN RIVER AT PRINCETON	BC	49.459702	-120.503530	1810.00	RHBN
08NL038	SIMILKAMEEN RIVER NEAR HEDLEY	BC	49.376968	-120.152310	5580.00	RHBN
08NL070	SIMILKAMEEN RIVER ABOVE GOODFELLOW CREEK	BC	49.094101	-120.672550	408.00	RHBN
08NM174	WHITEMAN CREEK ABOVE BOULEAU CREEK	BC	50.212219	-119.538610	114.00	RHBN
08NM240	TWO FORTY CREEK NEAR PENTICTON	BC	49.650890	-119.400030	4.94	RHBN
08NM242	DENNIS CREEK NEAR 1780 METRE CONTOUR	BC	49.624168	-119.381430	3.73	RHBN
08NN015	WEST KETTLE RIVER NEAR MCCULLOCH	BC	49.702919	-119.092470	233.00	RHBN
08OA002	YAKOUN RIVER NEAR PORT CLEMENTS	BC	53.613892	-132.209720	480.00	RHBN
08OA003	PREMIER CREEK NEAR QUEEN CHARLOTTE	BC	53.258610	-132.075040	NaN	RHBN
08OB002	PALLANT CREEK NEAR QUEEN CHARLOTTE	BC	53.057301	-132.051360	76.70	RHBN
09AA006	ATLIN RIVER NEAR ATLIN	BC	59.595360	-133.814423	6860.00	RHBN
09AE003	SWIFT RIVER NEAR SWIFT RIVER	BC	59.930828	-131.768620	3390.00	RHBN
10AC005	COTTONWOOD RIVER ABOVE BASS CREEK	BC	59.117500	-129.825550	882.00	RHBN
10BE001	LIARD RIVER AT LOWER CROSSING	BC	59.412498	-126.097220	104000.00	RHBN
10BE004	TOAD RIVER ABOVE NONDA CREEK	BC	58.854969	-125.382610	2540.00	RHBN
10BE007	TROUT RIVER AT KILOMETRE 783.7 ALASKA HIGHWAY	BC	59.335972	-125.940250	1170.00	RHBN
10BE009	TEETER CREEK NEAR THE MOUTH	BC	59.454170	-126.227500	209.00	RHBN
10CB001	SIKANNI CHIEF RIVER NEAR FORT NELSON	BC	57.238220	-122.691470	2180.00	RHBN
10CD001	MUSKWA RIVER NEAR FORT NELSON	BC	58.788109	-122.661640	20300.00	RHBN
10CD004	BOUGIE CREEK AT KILOMETRE 368 ALASKA HIGHWAY	BC	58.030251	-122.720170	335.00	RHBN
10CD005	ADSETT CREEK AT KILOMETRE 386.0 ALASKA HIGHWAY	BC	58.105999	-122.715580	109.00	RHBN
12355000	FLATHEAD RIVER AT FLATHEAD, BRITISH COLUMBIA	BC	49.000556	-114.476389	1105.93	HCDN
09423350	CARUTHERS C NR IVANPAH CA	CA	35.244989	-115.298876	2.24	HCDN
10258000	TAHQUITZ CR NR PALM SPRINGS CA	CA	33.805000	-116.558333	43.77	HCDN
10258500	PALM CANYON CREEK NR PALM SPRINGS CALIF	CA	33.745000	-116.534722	241.13	HCDN
10259000	ANDREAS CREEK NEAR PALM SPRINGS, CALIF.	CA	33.760000	-116.549167	22.40	HCDN
10259200	DEEP C NR PALM DESERT CA	CA	33.631136	-116.392234	79.12	HCDN
10263500	BIG ROCK CREEK NEAR VALYERMO,CALIF.	CA	34.420833	-117.838611	59.31	HCDN
10296000	W WALKER R BL L WALKER R NR COLEVILLE, CA	CA	38.379722	-119.449167	466.20	HCDN
10296500	W WALKER R NR COLEVILLE, CA	CA	38.515278	-119.454167	647.50	HCDN
10308200	E F CARSON R BL MARKLEEVILLE C NR MARKLEEVILLEC	CA	38.713889	-119.763889	714.84	HCDN
10310000	W F CARSON R AT WOODFORDS, CA	CA	38.769444	-119.831944	169.39	HCDN
10336645	GENERAL C NR MEEKS BAY CA	CA	39.051852	-120.118521	19.60	HCDN
10336660	BLACKWOOD CREEK NR TAHOE CITY CALIF	CA	39.107500	-120.161111	29.01	HCDN
10336676	WARD C AT HWY 89 NR TAHOE PINES CA	CA	39.132129	-120.157691	24.72	HCDN
10336780	TROUT CREEK NR TAHOE VALLEY CALIF	CA	38.920000	-119.971389	95.05	HCDN
10343500	SAGEHEN CREEK NR TRUCKEE CALIF	CA	39.431667	-120.236944	27.20	HCDN
11015000	SWEETWATER R NR DESCANSO CA	CA	32.834774	-116.623075	117.76	HCDN
11025500	SANTA YSABEL CREEK NEAR RAMONA, CALIF.	CA	33.106944	-116.865278	290.08	HCDN
11042400	TEMECULA CREEK NEAR AGUANGA, CALIF.	CA	33.459167	-116.922778	339.29	HCDN
11055801	CITY C NR HIGHLAND CA.+ CANALS	CA	34.143889	-117.187778	50.76	HCDN

Continued on next page

TABLE A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
11058500	EAST TWIN CREEK NEAR ARROWHEAD SPRINGS, CALIF.	CA	34.179167	-117.264722	22.79	HCDN
11075800	SANTIAGO C A MODJESKA CA	CA	33.712778	-117.644167	33.67	HCDN
11098000	ARROYO SECO NR PASADENA CALIF	CA	34.222222	-118.176667	41.44	HCDN
11111500	SESPE CREEK NR WHEELER SPRINGS CALIF	CA	34.577778	-119.256944	128.20	HCDN
11124500	SANTA CRUZ CR NR SANTA YNEZ CA	CA	34.596667	-119.907778	191.66	HCDN
11132500	SALSIPUEDES CR NR LOMPOC CA	CA	34.588611	-120.407500	121.99	HCDN
11138500	SISQUOC RIVER NEAR SISQUOC, CALIF.	CA	34.839722	-120.167222	727.79	HCDN
11141280	LOPEZ C NR ARROYO GRANDE CA	CA	35.235556	-120.471389	54.13	HCDN
11143000	BIG SUR RIVER NR BIG SUR CALIF	CA	36.245833	-121.772222	120.44	HCDN
11148900	NACIMIENTO R BL SAPAQUE C NR BRYSON CA	CA	35.788579	-121.093805	403.45	HCDN
11149900	SAN ANTONIO RIVER NEAR LOCKWOOD CALIF	CA	35.896667	-121.087222	562.03	HCDN
11151300	SAN LORENZO C BL BITTERWATER C NR KING CITY CAL	CA	36.268056	-121.065278	603.47	HCDN
11152000	ARROYO SECO NEAR SOLEDAD, CAL.	CA	36.280556	-121.321667	631.96	HCDN
11160000	SOQUEL CR AT SOQUEL CALIF	CA	36.991389	-121.954722	104.12	HCDN
11160500	SAN LORENZO R AT BIG TREES CALIF	CA	37.044444	-122.071389	274.54	HCDN
11162500	PESCADERO CREEK NEAR PESCADERO CALIF	CA	37.260833	-122.327778	118.88	HCDN
11176400	ARROYO VALLE BL LANG CN NR LIVERMORE CALIF	CA	37.561389	-121.682778	336.70	HCDN
11180500	DRY C A UNION CITY CA	CA	37.606045	-122.023850	24.32	HCDN
11180960	CULL C AB CULL C RES NR CASTRO VALLEY CA	CA	37.717707	-122.054407	15.06	HCDN
11182500	SAN RAMON CREEK AT SAN RAMON, CALIF.	CA	37.773056	-121.993611	15.26	HCDN
11186001	COMBINED FLOW OF KERN R AND KERN R NO 3 CA	CA	35.945278	-118.476667	2191.14	HCDN
11189500	SF KERN R NR ONYX CALIF	CA	35.737500	-118.172778	1372.70	HCDN
11200800	DEER CREEK NEAR FOUNTAIN SPRINGS CALIF	CA	35.941667	-118.821944	215.75	HCDN
11202001	NF OF MF TULE R NR SPRINGVILLE CALIF (TOTAL FLO	CA	36.174722	-118.694722	101.79	HCDN
11224500	LOS GATOS CREEK AB NUNEZ CANYON NR COALINGA CAL	CA	36.214722	-120.469722	248.12	HCDN
11230500	BEAR CR NR LAKE T.A.EDISON CALIF	CA	37.338333	-118.973056	135.98	HCDN
11237500	PITMAN C BL TAMARACK CREEK CALIF	CA	37.198333	-119.213333	59.31	HCDN
11253310	CANTUA CREEK NR CANTUA CREEK CALIF	CA	36.402222	-120.432500	120.18	HCDN
11264500	MERCED R AT HAPPY ISLES BRIDGE NR YOSEMITE CALI	CA	37.731667	-119.557778	468.79	HCDN
11266500	MERCED RIVER AT POHONO BRIDGE NEAR YOSEMITE, CA	CA	37.716944	-119.665278	831.39	HCDN
11274500	ORESTIMBA CREEK NR NEWMAN CALIF	CA	37.315556	-121.124167	347.06	HCDN
11274630	DEL PUERTO C NR PATTERSON CA	CA	37.486599	-121.209103	187.44	HCDN
11284400	BIG C AB WHITES GULCH NR GROVELAND CA	CA	37.841871	-120.184910	41.72	HCDN
11294000	HIGHLAND C BL SPICER MEADOWS RES CALIF	CA	38.390000	-120.006111	117.59	HCDN
11299600	BLACK C NR COPPEROPOLIS CA	CA	37.961036	-120.615204	37.32	HCDN
11315000	COLE C NR SALT SPRINGS DAM CALIF	CA	38.519167	-120.211667	54.39	HCDN
11316800	FOREST CREEK NEAR WILSEYVILLE, CALIF.	CA	38.403333	-120.445833	53.87	HCDN
11342000	SACRAMENTO RIVER AT DELTA CALIF	CA	40.939722	-122.416111	1100.75	HCDN
11367500	MCCLOUD RIVER NR MCCLOUD CALIF	CA	41.188333	-122.064444	927.22	HCDN
11368000	MCCLOUD RIVER AB SHASTA LAKE CALIF	CA	40.958333	-122.218611	1564.36	HCDN
11372000	CLEAR CREEK NR IGO CALIF	CA	40.513333	-122.523056	590.52	HCDN

Continued on next page

Table A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
11379500	ELDER CREEK NEAR PASKENTA CALIF	CA	40.024722	-122.508611	239.32	HCDN
11381500	MILL C NR LOS MOLINOS CALIF	CA	40.054722	-122.023056	339.29	HCDN
11383500	DEER CREEK NEAR VINA CALIF	CA	40.014167	-121.947222	538.72	HCDN
11402000	SPANISH CREEK ABOVE BLACKHAWK CREEK AT KEDDIE,	CA	40.003056	-120.953333	476.56	HCDN
11413000	NORTH YUBA RIVER BELOW GOODYEARS BAR, CALIF.	CA	39.525000	-120.936944	647.50	HCDN
11425500	SACRAMENTO RIVER AT VERONA, CALIF.	CA	38.774444	-121.597222	3240.09	HCDN
11427000	NF AMERICAN R AT NORTH FORK DAM CALIF	CA	38.936111	-121.022778	885.78	HCDN
11427700	DUNCAN CREEK NR FRENCH MEADOWS CALIF	CA	39.135833	-120.477500	25.74	HCDN
11441500	S.F. SILVER CREEK NEAR ICE HOUSE, CALIF	CA	38.818889	-120.364167	71.22	HCDN
11449500	KELSEY CREEK NEAR KELSEYVILLE, CALIF.	CA	38.927500	-122.842500	94.79	HCDN
11451100	NF CACHE C AT HOUGH SPRING NEAR CLEARLAKE OAKS,	CA	39.165556	-122.618889	155.92	HCDN
11456000	NAPA RIVER NEAR ST. HELENA CALIF	CA	38.497778	-122.426944	210.83	HCDN
11461000	RUSSIAN RIVER NEAR UKIAH, CALIF.	CA	39.195556	-123.193889	259.00	HCDN
11465200	DRY CREEK NR GEYSERVILLE CALIF	CA	38.698611	-122.956944	419.58	HCDN
11468000	NAVARRO RIVER NEAR NAVARRO, CALIF.	CA	39.172222	-123.668333	784.77	HCDN
11468500	NOYO RIVER NR FORT BRAGG CALIF	CA	39.428333	-123.736667	274.54	HCDN
11469000	MATTOLE RIVER NR PETROLIA CALIF	CA	40.311667	-124.263333	621.60	HCDN
11473900	MIDDLE FORK EEL R NR DOS RIOS CALIF	CA	39.706389	-123.324167	1929.55	HCDN
11475560	ELDER CREEK NEAR BRANSCOMB CALIF	CA	39.729722	-123.642778	16.84	HCDN
11475800	SF EEL RIVER AT LEGGETT CALIF	CA	39.874722	-123.719444	642.32	HCDN
11476500	SF EEL RIVER NR MIRANDA CALIF	CA	40.181944	-123.775000	1390.83	HCDN
11476600	BULL CREEK NEAR WEOTT, CALIF.	CA	40.351389	-124.002778	72.78	HCDN
11477000	EEL RIVER AT SCOTIA CALIF	CA	40.491667	-124.098611	8062.67	HCDN
11478500	VAN DUZEN RIVER NR BRIDGEVILLE CALIF	CA	40.480556	-123.889722	574.98	HCDN
11480390	MAD R AB RUTH RES NR FOREST GLEN CA	CA	40.284313	-123.335303	242.60	HCDN
11481200	LITTLE R NR TRINIDAD CALIF	CA	41.011111	-124.080556	104.90	HCDN
11481500	REDWOOD C NR BLUE LAKE CALIF	CA	40.906111	-123.814167	175.34	HCDN
11482500	REDWOOD CREEK AT ORICK CALIF	CA	41.299444	-124.050000	717.43	HCDN
11519500	SCOTT RIVER NEAR FORT JONES, CALIF.	CA	41.640833	-123.013889	1691.27	HCDN
11521500	INDIAN CREEK NEAR HAPPY CAMP, CALIF.	CA	41.835278	-123.381944	310.80	HCDN
11522500	SALMON RIVER AT SOMES BAR CALIF	CA	41.377778	-123.476389	1945.09	HCDN
11523200	TRINITY R AB COFFEE C NR TRINITY CTR CA	CA	41.111389	-122.704444	385.91	HCDN
11525500	TRINITY RIVER AT LEWISTON CALIF	CA	40.719444	-122.802500	1862.21	HCDN
11528700	S F TRINITY RIVER BL HYAMPOM, CALIF.	CA	40.650000	-123.493056	1978.76	HCDN
11530000	TRINITY R AT HOOPA CALIF	CA	41.050000	-123.670833	7389.27	HCDN
11532500	SMITH RIVER NEAR CRESCENT CITY, CALIF.	CA	41.789444	-124.053889	1577.31	HCDN
06614800	MICHIGAN RIVER NEAR CAMERON PASS, CO.	CO	40.496094	-105.865012	4.03	HCDN
06620000	NORTH PLATTE RIVER NEAR NORTHGATE, CO	CO	40.937500	-106.337778	3706.29	HCDN
06746095	JOE WRIGHT CREEK ABOVE JOE WRIGHT RESERVOIR, CO.	CO	40.539982	-105.882790	8.90	HCDN
07083000	HALFMOON CREEK NEAR MALTA, CO.	CO	39.172222	-106.388611	61.12	HCDN

Continued on next page

Table A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
07096000	ARKANSAS RIVER AT CANON CITY, CO.	CO	38.433889	-105.256667	8073.03	HCDN
07105945	ROCK CREEK ABOVE FORT CARSON RESERVATION, CO.	CO	38.707493	-104.846640	17.48	HCDN
09034900	BOBTAIL CREEK NEAR JONES PASS, CO.	CO	39.760264	-105.906401	15.65	HCDN
09035800	DARLING CREEK NEAR LEAL, CO.	CO	39.800542	-106.026407	22.94	HCDN
09035900	SOUTH FORK OF WILLIAMS FORK NEAR LEAL, CO.	CO	39.795820	-106.030573	72.84	HCDN
09047700	KEYSTONE GULCH NEAR DILLON, CO.	CO	39.594431	-105.972516	23.57	HCDN
09065500	GORE CREEK AT UPPER STATION, NEAR MINTURN, CO.	CO	39.625819	-106.278082	37.78	HCDN
09066000	BLACK GORE CREEK NEAR MINTURN, CO.	CO	39.596375	-106.265026	32.41	HCDN
09066200	BOOTH CREEK NEAR MINTURN, CO.	CO	39.648319	-106.323083	16.10	HCDN
09066300	MIDDLE CREEK NEAR MINTURN, CO.	CO	39.645819	-106.382251	15.52	HCDN
09081600	CRYSTAL RIVER AB AVALANCHE C, NEAR REDSTONE, CO	CO	39.232222	-107.226667	432.53	HCDN
09085000	ROARING FORK RIVER AT GLENWOOD SPRINGS, CO.	CO	39.543611	-107.328889	3758.09	HCDN
09107000	TAYLOR RIVER AT TAYLOR PARK, CO.	CO	38.860271	-106.566697	331.62	HCDN
09110000	TAYLOR RIVER AT ALMONT, CO.	CO	38.664444	-106.844722	1235.43	HCDN
09112500	EAST RIVER AT ALMONT CO.	CO	38.664444	-106.847500	748.51	HCDN
09119000	TOMICHI CREEK AT GUNNISON, CO.	CO	38.521667	-106.940278	2747.99	HCDN
09124500	LAKE FORK AT GATEVIEW, CO.	CO	38.298889	-107.229444	865.06	HCDN
09132500	NORTH FORK GUNNISON RIVER NEAR SOMERSET, CO.	CO	38.925833	-107.433611	1362.34	HCDN
09146200	UNCOMPAHGRE RIVER NEAR RIDGWAY, CO.	CO	38.183889	-107.745278	385.91	HCDN
09147500	UNCOMPAHGRE RIVER AT COLONA, CO.	CO	38.331389	-107.778889	1160.32	HCDN
09165000	DOLORES RIVER BELOW RICO, CO.	CO	37.638889	-108.059722	271.95	HCDN
09239500	YAMPA RIVER AT STEAMBOAT SPRINGS, CO.	CO	40.483611	-106.831667	1564.36	HCDN
09251000	YAMPA RIVER NEAR MAYBELL, CO.	CO	40.502778	-108.029167	8831.90	HCDN
09304500	WHITE RIVER NEAR MEEKER, CO.	CO	40.033611	-107.861667	1955.45	HCDN
09306242	CORRAL GULCH NEAR RANGELY, CO.	CO	39.920250	-108.472872	81.99	HCDN
09352900	VALLECITO CREEK NEAR BAYFIELD, CO.	CO	37.477500	-107.543056	186.74	HCDN
09361500	ANIMAS RIVER AT DURANGO, CO.	CO	37.279167	-107.879722	1792.28	HCDN
08NH032	BOUNDARY CREEK NEAR PORTHILL	ID	48.997219	-116.568060	242.00	RHBN
12306500	MOYIE RIVER AT EASTPORT, IDAHO	ID	48.999444	-116.178611	1476.30	HCDN
12321500	BOUNDARY CREEK NR PORTHILL ID	ID	48.997222	-116.568056	251.23	HCDN
12411000	COEUR D'ALENE R AB SHOSHONE CK NR PRICHARD, IDA	ID	47.708333	-115.976389	867.65	HCDN
12413000	COEUR D'ALENE RIVER AT ENAVILLE IDAHO	ID	47.572222	-116.252778	2318.05	HCDN
12413500	COEUR D'ALENE RIVER NR CATALDO, IDAHO	ID	47.563889	-116.306944	3159.80	HCDN
12414500	ST. JOE RIVER AT CALDER, ID	ID	47.274722	-116.188056	2667.70	HCDN
12414900	ST. MARIES RIVER NEAR SANTA IDAHO	ID	47.176389	-116.491667	712.25	HCDN
13037500	SNAKE RIVER NR HEISE,ID	ID	43.612500	-111.659167	14897.68	HCDN
13073000	PORTNEUF RIVER AT TOPAZ ID	ID	42.625000	-112.088889	1476.30	HCDN
13075000	MARSH CREEK NR MCCAMMON ID	ID	42.630000	-112.224722	914.27	HCDN
13082500	GOOSE CREEK AB TRAPPER CREEK NR OAKLEY ID	ID	42.125000	-113.938889	1639.47	HCDN
13083000	TRAPPER CREEK NR OAKLEY ID	ID	42.165833	-113.983611	133.18	HCDN

Continued on next page

TABLE A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
13120000	N FK BIG LOST RIVER AT WILD HORSE NR CHILLY ID	ID	43.933056	-114.113056	295.26	HCDN
13120500	BIG LOST RIVER AT HOWELL RANCH NR CHILLY ID	ID	43.998333	-114.020000	1165.50	HCDN
13139510	COMBINATION BIG WOOD R/SLOUGH AT HAILEY ID	ID	43.518056	-114.319444	1657.60	HCDN
13147900	LITTLE WOOD RIVER AB HIGH FIVE CREEK NR CAREY I	ID	43.491667	-114.058333	642.32	HCDN
13168500	BRUNEAU RIVER NR HOT SPRING ID	ID	42.771111	-115.719444	6811.70	HCDN
13185000	BOISE RIVER NR TWIN SPRINGS ID	ID	43.659167	-115.726111	2149.70	HCDN
13186000	SF BOISE RIVER NR FEATHERVILLE ID	ID	43.494444	-115.305556	1644.65	HCDN
13200000	MORES CREEK AB ROBIE CREEK NR ARROWROCK DAM ID	ID	43.648056	-115.988889	1033.41	HCDN
13235000	SF PAYETTE RIVER AT LOWMAN ID	ID	44.085278	-115.621111	1181.04	HCDN
13240000	LAKE FORK PAYETTE RIVER AB JUMBO CR NR MCCALL I	ID	44.913611	-115.996389	126.65	HCDN
13269000	SNAKE RIVER AT WEISER ID	ID	44.245556	-116.980000	23828.00	HCDN
13296500	SALMON RIVER BL YANKEE FORK NR CLAYTON ID	ID	44.268333	-114.731944	2077.18	HCDN
13302500	SALMON RIVER AT SALMON ID	ID	45.183333	-113.894444	9738.40	HCDN
13305000	LEMHI RIVER NR LEMHI ID	ID	44.940000	-113.637778	2318.05	HCDN
13310700	SF SALMON RIVER NR KRASSEL RANGER STATION ID	ID	44.986944	-115.725000	853.13	HCDN
13313000	JOHNSON CREEK AT YELLOW PINE ID	ID	44.962222	-115.499444	551.67	HCDN
13316500	LITTLE SALMON RIVER AT RIGGINS ID	ID	45.413056	-116.324722	1491.84	HCDN
13317000	SALMON RIVER AT WHITE BIRD ID	ID	45.750278	-116.323056	9194.50	HCDN
13336500	SELWAY RIVER NR LOWELL ID	ID	46.086667	-115.512778	4946.90	HCDN
13337000	LOCHSA RIVER NR LOWELL ID	ID	46.150556	-115.586389	3056.20	HCDN
13338500	S FK CLEARWATER RIVER AT STITES ID	ID	46.086667	-115.975556	2978.50	HCDN
13340000	CLEARWATER RIVER AT OROFINO ID	ID	46.478611	-116.256389	14452.20	HCDN
13340600	N FK CLEARWATER RIVER NR CANYON RANGER STA ID	ID	46.840556	-115.619722	3522.40	HCDN
13342500	CLEARWATER RIVER AT SPALDING, ID	ID	46.448611	-116.826389	24786.30	HCDN
13345000	PALOUSE RIVER NR POTLATCH, ID.	ID	46.915278	-116.950000	821.03	HCDN
05014500	SWIFTCURRENT CREEK AT MANY GLACIER MT.	MT	48.799167	-113.655833	81.33	HCDN
06019500	RUBY RIVER ABOVE RESERVOIR NEAR ALDER MT	MT	45.175278	-112.147778	1393.42	HCDN
06037500	MADISON RIVER NEAR WEST YELLOWSTONE, MT.	MT	44.656944	-111.067500	1087.80	HCDN
06043500	GALLATIN RIVER NEAR GALLATIN GATEWAY MT	MT	45.497500	-111.420000	2136.75	HCDN
06061500	PRICKLY PEAR CREEK NEAR CLANCY, MT.	MT	46.519167	-111.945833	497.28	HCDN
06099500	MARIAS RIVER NEAR SHELBY, MT.	MT	48.427222	-111.888889	8396.78	HCDN
06191000	GARDNER RIVER NEAR MAMMOTH YNP	MT	44.992500	-110.690556	523.18	HCDN
06191500	YELLOWSTONE RIVER AT CORWIN SPRINGS, MT.	MT	45.111944	-110.793611	6793.57	HCDN
06192500	YELLOWSTONE RIVER NEAR LIVINGSTON, MT.	MT	45.597222	-110.565278	9197.09	HCDN
06207500	CLARKS FORK YELLOWSTONE RIVER NEAR BELFRY MT	MT	45.010278	-109.064722	2988.86	HCDN
06214500	YELLOWSTONE RIVER AT BILLINGS MT	MT	45.796667	-108.470000	4649.05	HCDN
06289000	LITTLE BIGHORN RIVER AT STATE LINE NEAR WYOLA M	MT	45.006944	-107.614444	499.87	HCDN
12302055	FISHER RIVER NEAR LIBBY, MT.	MT	48.355556	-115.313889	2170.42	HCDN
12304500	YAAK RIVER NEAR TROY, MT.	MT	48.561944	-115.969167	1983.94	HCDN
12330000	BOULDER CREEK AT MAXVILLE, MT.	MT	46.472222	-113.233056	184.67	HCDN
12332000	MIDDLE FORK ROCK CREEK NEAR PHILIPSBURG, MT.	MT	46.195000	-113.500000	318.57	HCDN
12354500	CLARK FORK AT ST. REGIS, MT.	MT	47.301944	-115.086389	1836.31	HCDN

Continued on next page

TABLE A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
12355500	N F FLATHEAD RIVER NEAR COLUMBIA FALLS, MT.	MT	48.495556	-114.126667	4009.32	HCDN
12358500	MIDDLE FORK FLATHEAD RIVER NEAR WEST GLACIER MT	MT	48.495278	-114.009167	2921.52	HCDN
12370000	SWAN RIVER NEAR BIGFORK, MT.	MT	48.024444	-113.978889	1737.89	HCDN
12374250	Mill Cr ab Bassoo Cr nr Niarada MT	MT	47.829664	-114.697630	50.79	HCDN
12375900	South Crow Creek near Ronan MT	MT	47.491600	-114.026774	19.72	HCDN
12377150	Mission Cr ab Reservoir nr ST Ignatius MT	MT	47.322987	-113.979547	32.18	HCDN
12381400	South Fork Jocko River near Arlee MT	MT	47.195486	-113.850653	150.95	HCDN
12390700	PROSPECT CREEK AT THOMPSON FALLS, MT.	MT	47.586111	-115.354167	471.38	HCDN
07203000	VERMEJO RIVER NEAR DAWSON, NM	NM	36.680556	-104.785556	779.59	HCDN
07207500	PONIL CREEK NEAR CIMARRON, N. MEX.	NM	36.573611	-104.946111	442.89	HCDN
07208500	RAYADO CREEK AT SAUBLE RANCH NEAR CIMARRON, N.M	NM	36.372222	-104.969444	168.35	HCDN
07211500	CANADIAN R NR TAYLOR SPRINGS, NM	NM	36.296944	-104.493333	7381.50	HCDN
07216500	MORA RIVER NEAR GOLONDRINAS N. MEX.	NM	35.890833	-105.163056	691.53	HCDN
07218000	COYOTE CREEK NEAR GOLONDRINAS, N. MEX.	NM	35.916667	-105.163611	556.85	HCDN
07226500	UTE CREEK NEAR LOGAN, NM	NM	35.438382	-103.525794	5242.75	HCDN
08267500	RIO HONDO NEAR VALDEZ, N. MEX.	NM	36.541667	-105.555833	93.76	HCDN
08269000	RIO PUEBLO DE TAOS NEAR TAOS, N. MEX.	NM	36.439444	-105.503056	172.49	HCDN
08271000	RIO LUCERO NEAR ARROYO SECO, N. MEX.	NM	36.508333	-105.530278	42.99	HCDN
08275500	RIO GRANDE DEL RANCHO NEAR TALPA, N. MEX.	NM	36.297778	-105.581944	214.97	HCDN
08276500	RIO GRANDE BELOW TAOS JUNCTION BRIDGE NR TAOS,N	NM	36.320000	-105.753889	25200.70	HCDN
08279000	EMBUDO CREEK AT DIXON, NM	NM	36.210833	-105.913056	789.95	HCDN
08289000	RIO OJO CALIENTE AT LA MADERA, NM	NM	36.349722	-106.043611	1085.21	HCDN
08291000	SANTA CRUZ RIVER NEAR CUNDIYO, N. MEX.	NM	35.964722	-105.903889	222.74	HCDN
08324000	JEMEZ RIVER NR JEMEZ,NM	NM	35.661667	-106.742778	1217.30	HCDN
08377900	RIO MORA NEAR TERRERO, NM	NM	35.777222	-105.657500	137.79	HCDN
08378500	PECOS R NR PECOS, NM	NM	35.708333	-105.681944	489.51	HCDN
08380500	GALLINAS CREEK NEAR MONTEZUMA, N. MEX.	NM	35.651944	-105.318333	217.56	HCDN
08387000	RIO RUIDOSO AT HOLLYWOOD, N. MEX.	NM	33.326667	-105.627222	310.80	HCDN
08405500	BLACK RIVER ABOVE MALAGA, N. MEX.	NM	32.228889	-104.150556	888.37	HCDN
08408500	DELAWARE RIVER NR RED BLUFF, N M	NM	32.023056	-104.054167	1784.51	HCDN
09364500	ANIMAS RIVER AT FARMINGTON, NM	NM	36.721389	-108.201389	3522.40	HCDN
09386900	RIO NUTRIA NEAR RAMAH, NM	NM	35.282529	-108.553411	184.94	HCDN
09430500	GILA RIVER NEAR GILA, NM	NM	33.061111	-108.536667	4827.76	HCDN
09430600	MOGOLLON CREEK NEAR CLIFF, NM	NM	33.166667	-108.649167	178.71	HCDN
09431500	GILA RIVER NEAR REDROCK, NM	NM	32.726944	-108.675000	7327.11	HCDN
10EB001	SOUTH NAHANNI RIVER ABOVE VIRGINIA FALLS	NT	61.636108	-125.797030	14500.00	RHBN
10ED002	LIARD RIVER NEAR THE MOUTH	NT	61.742722	-121.227970	275000.00	RHBN
10ED003	BIRCH RIVER AT HIGHWAY NO. 7	NT	61.333580	-122.094250	542.00	RHBN
10FA002	TROUT RIVER AT HIGHWAY NO. 1	NT	61.139809	-119.843060	9270.00	RHBN
10FB005	JEAN-MARIE RIVER AT HIGHWAY NO. 1	NT	61.445530	-121.238190	1310.00	RHBN
10GA001	ROOT RIVER NEAR THE MOUTH	NT	62.479721	-123.433060	9820.00	RHBN
10GB006	WILLOWLAKE RIVER ABOVE METAHDALI CREEK	NT	62.650059	-122.899250	20200.00	RHBN

Continued on next page

Table A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
10LA002	ARCTIC RED RIVER NEAR THE MOUTH	NT	66.787437	-133.089650	18750.00	RHBN
10LC007	CARIBOU CREEK ABOVE HIGHWAY NO. 8 (DEMPSTER HIG...	NT	68.087280	-133.492320	590.00	RHBN
10MC002	PEEL RIVER ABOVE FORT MCPHERSON	NT	67.258888	-134.888810	70600.00	RHBN
10ND002	TRAIL VALLEY CREEK NEAR INUVIK	NT	68.736191	-133.493100	68.30	RHBN
10244950	STEPTOE C NR ELY, NV	NV	39.201389	-114.687500	28.75	HCDN
10249300	S TWIN R NR ROUND MOUNTAIN, NV	NV	38.887500	-117.244444	51.80	HCDN
10301500	WALKER R NR WABUSKA, NV	NV	39.152778	-119.097222	6734.00	HCDN
10310500	CLEAR CREEK NEAR CARSON CITY, NV	NV	39.113242	-119.798237	39.49	HCDN
10312000	CARSON R NR FORT CHURCHILL, NV	NV	39.291667	-119.311111	3372.18	HCDN
10316500	LAMOILLE C NR LAMOILLE, NV	NV	40.690833	-115.475556	64.75	HCDN
10322500	HUMBOLDT R AT PALISADE, NV	NV	40.606944	-116.201389	12975.90	HCDN
10329500	MARTIN C NR PARADISE VALLEY, NV	NV	41.533333	-117.427778	445.48	HCDN
10352500	MC DERMITT C NR MC DERMITT, NV	NV	41.966667	-117.833611	582.75	HCDN
13161500	BRUNEAU RIVER AT ROWLAND NV	NV	41.933333	-115.673611	989.38	HCDN
10396000	DONNER UND BLITZEN RIVER NR FRENCHGLEN,OREG.	OR	42.791111	-118.866667	518.00	HCDN
11501000	SPRAGUE RIVER NEAR CHILOQUIN,OREG.	OR	42.584722	-121.848611	4092.20	HCDN
13331500	MINAM RIVER AT MINAM,OREG.	OR	45.620000	-117.725556	621.60	HCDN
14020000	UMATILLA RIVER AB MEACHAM CR NR GIBBON,OREG.	OR	45.719722	-118.322222	339.29	HCDN
14034500	WILLOW CREEK AT HEPPNER, OREG.	OR	45.350556	-119.548889	250.71	HCDN
14044000	M FK JOHN DAY R AT RITTER, OREG.	OR	44.888889	-119.140278	1333.85	HCDN
14046500	JOHN DAY RIVER AT SERVICE CREEK, OREG.	OR	44.793889	-120.005556	13183.10	HCDN
14048000	JOHN DAY R AT MCDONALD FERRY,OREG.	OR	45.587778	-120.408333	19632.20	HCDN
14092750	SHITIKE CR, AT PETERS PASTURE, NR WARM SPRINGS, OR	OR	44.750395	-121.633397	57.48	HCDN
14096850	BEAVER CREEK, BLW QUARTZ CR, NR SIMNASHO, OR.	OR	44.958727	-121.394229	374.62	HCDN
14105700	COLUMBIA RIVER AT THE DALLES,OREG.	OR	45.607500	-121.172222	18130.00	HCDN
14137000	SANDY RIVER NEAR MARMOT, OREG.	OR	45.391667	-122.127778	678.58	HCDN
14138800	BLAZED ALDER CREEK NEAR RHODODENDRON, OREG.	OR	45.452618	-121.891468	21.29	HCDN
14138870	FIR CREEK NEAR BRIGHTWOOD, OR	OR	45.480119	-122.025638	14.02	HCDN
14138900	NORTH FORK BULL RUN RIVER NEAR MULTNOMAH FALLS, OR	OR	45.494286	-122.035917	21.68	HCDN
14139800	SOUTH FORK BULL RUN RIVER NEAR BULL RUN, OR	OR	45.444564	-122.109529	40.70	HCDN
14141500	LITTLE SANDY RIVER NEAR BULL RUN, OREG.	OR	45.415398	-122.171475	59.87	HCDN
14154500	ROW RIVER ABOVE PITCHER CREEK NEAR, DORENA, ORE	OR	43.736111	-122.872222	546.49	HCDN
14158500	MCKENZIE RIVER AT OUTLET OF CLEAR LAKE, OR	OR	44.360955	-121.995617	237.08	HCDN
14158790	SMITH R AB SMITH R RES NR BELKNAP SPRGS,OREG.	OR	44.334567	-122.047007	40.57	HCDN
14166500	LONG TOM RIVER NEAR NOTI, OREG.	OR	44.049845	-123.426210	226.52	HCDN
14178000	NO SANTIAM R BL BOULDER CR NR DETROIT,OREG.	OR	44.706944	-122.100000	559.44	HCDN
14179000	BREITENBUSH RIVER ABV FRENCH CR, NR DETROIT	OR	44.752778	-122.127778	279.72	HCDN
14182500	LITTLE NORTH SANTIAM RIVER NEAR MEHAMA, OR	OR	44.791511	-122.578974	286.85	HCDN
14185000	SOUTH SANTIAM RIVER BELOW CASCADIA, OREG.	OR	44.393056	-122.509722	450.66	HCDN
14185900	QUARTZVILLE CREEK NEAR CASCADIA, OREG.	OR	44.540278	-122.434722	256.93	HCDN

Continued on next page

Table A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
14187000	WILEY CREEK NEAR FOSTER, OR	OR	44.372348	-122.623138	134.75	HCDN
14190500	LUCKIAMUTE RIVER NEAR SUVER, OREG.	OR	44.783333	-123.233333	621.60	HCDN
14191000	WILLAMETTE RIVER AT SALEM,OREG.	OR	44.944444	-123.041667	18855.20	HCDN
14301000	NEHALEM RIVER NEAR FOSS, OREG.	OR	45.704167	-123.754167	1727.53	HCDN
14301500	WILSON RIVER NEAR TILLAMOOK, OREG.	OR	45.484722	-123.688889	416.99	HCDN
14303200	TUCCA CREEK NEAR BLAINE, OR.	OR	45.324276	-123.546501	8.06	HCDN
14305500	SILETZ RIVER AT SILETZ, OR	OR	44.715117	-123.887335	526.33	HCDN
14306340	EAST FORK LOBSTER CREEK NEAR ALSEA, OR.	OR	44.247899	-123.636494	14.73	HCDN
14306500	ALSEA RIVER NEAR TIDEWATER, OREG.	OR	44.386111	-123.830556	865.06	HCDN
14307620	SIUSLAW R NR MAPLETON, OREG.	OR	44.062500	-123.881944	1522.92	HCDN
14308000	SOUTH UMPQUA RIVER AT TILLER, OREG.	OR	42.930556	-122.947222	1162.91	HCDN
14308990	COW CREEK ABV GALESVILLE RES, NR AZALEA, OR.	OR	42.823177	-123.125892	167.76	HCDN
14309500	WEST FORK COW CREEK NEAR GLENDALE, OR	OR	42.804004	-123.610907	224.92	HCDN
14316700	STEAMBOAT CREEK NEAR GLIDE,OREG.	OR	43.350000	-122.727778	587.93	HCDN
14318000	LITTLE RIVER AT PEEL, OREG.	OR	43.252778	-123.025000	458.43	HCDN
14321000	UMPQUA RIVER NEAR ELKTON, OREG.	OR	43.586111	-123.554167	9538.97	HCDN
14325000	SOUTH FORK COQUILLE RIVER AT POWERS, OREG.	OR	42.891667	-124.069444	437.71	HCDN
14338000	ELK CREEK NEAR TRAIL, OREG.	OR	42.675000	-122.743889	334.11	HCDN
14359000	ROGUE RIVER AT RAYGOLD NR CENTRAL PT,OREG.	OR	42.437500	-122.986111	5317.27	HCDN
14362000	APPLEGATE RIVER NEAR COPPER, OREG.	OR	42.058333	-123.113889	582.75	HCDN
14362250	STAR GULCH NEAR RUCH, OR.	OR	42.154014	-123.075323	41.42	HCDN
14377100	ILLINOIS RIVER NEAR KERBY, OREG	OR	42.231944	-123.662500	984.20	HCDN
14400000	CHETCO RIVER NR BROOKINGS, OREG.	OR	42.123611	-124.186111	701.89	HCDN
09180500	COLORADO RIVER NEAR CISCO UTAH	UT	38.810556	-109.292778	10619.00	HCDN
09183500	MILL CREEK AT SHELEY TUNNEL, NEAR MOAB, UT.	UT	38.483040	-109.404004	74.30	HCDN
09299500	WHITEROCKS RIVER NEAR WHITEROCKS, UTAH	UT	40.586944	-109.926944	292.67	HCDN
09310500	FISH CREEK ABOVE RESERVOIR NEAR SCOFIELD, UTAH	UT	39.774444	-111.190278	155.66	HCDN
09312600	WHITE RIVER BL TABBYUNE C NEAR SOLDIER SUMMIT, UT	UT	39.875793	-111.037388	195.29	HCDN
09315000	GREEN RIVER AT GREEN RIVER, UTAH	UT	38.986111	-110.150556	12561.50	HCDN
09330500	MUDDY CREEK NEAR EMERY, UTAH	UT	38.981944	-111.248611	271.95	HCDN
09378170	SOUTH CREEK ABOVE RESERVOIR NEAR MONTICELLO, UT	UT	37.846661	-109.369563	21.90	HCDN
09378630	RECAPTURE CREEK NEAR BLANDING, UT	UT	37.755550	-109.476511	10.42	HCDN
09379500	SAN JUAN RIVER NEAR BLUFF, UTAH	UT	37.146944	-109.864167	7770.00	HCDN
09404450	EAST FORK VIRGIN RIVER NR GLENDALE, UTAH	UT	37.338611	-112.603611	192.18	HCDN
09408400	SANTA CLARA RIVER NR PINE VALLEY UTAH	UT	37.383333	-113.482500	48.43	HCDN
10023000	BIG CREEK NEAR RANDOLPH, UT	UT	41.609942	-111.254088	131.46	HCDN
10128500	WEBER RIVER NEAR OAKLEY, UTAH	UT	40.736111	-111.245833	419.58	HCDN
10131000	CHALK CREEK AT COALVILLE UTAH	UT	40.920556	-111.400833	647.50	HCDN
10166430	WEST CANYON CREEK NEAR CEDAR FORT, UT	UT	40.405226	-112.100496	70.17	HCDN
10172200	RED BUTTE CREEK AT FT. DOUGLAS NR. SLC, UTAH	UT	40.780000	-111.805278	18.78	HCDN
10172700	VERNON CREEK NEAR VERNON, UTAH	UT	39.979444	-112.379444	64.75	HCDN

Continued on next page

Table A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
10172800	SOUTH WILLOW CREEK NEAR GRANTSVILLE, UT	UT	40.496331	-112.574403	11.07	HCDN
10172870	TROUT CR NR CALLAO UTAH	UT	39.744167	-113.889167	21.21	HCDN
10173450	MAMMOTH CREEK ABV WEST HATCH DITCH, NEAR HATCH, UT	UT	37.622757	-112.516877	268.53	HCDN
10174500	SEVIER RIVER AT HATCH UTAH	UT	37.651111	-112.429444	880.60	HCDN
10205030	SALINA CREEK NEAR EMERY UTAH	UT	38.911944	-111.529722	134.16	HCDN
10234500	BEAVER RIV NR BEAVER UTAH	UT	38.280556	-112.573611	235.69	HCDN
10242000	COAL CREEK NEAR CEDAR CITY, UT	UT	37.672199	-113.034670	208.69	HCDN
12013500	WILLAPA RIVER NEAR WILLAPA, WASH.	WA	46.650000	-123.652778	336.70	HCDN
12020000	CHEHALIS RIVER NEAR DOTY, WASH.	WA	46.617500	-123.276389	292.67	HCDN
12025000	NEWAUKUM RIVER NEAR CHEHALIS, WA	WA	46.620104	-122.945126	404.90	HCDN
12025700	SKOOKUMCHUCK RIVER NEAR VAIL, WA	WA	46.772603	-122.594009	103.30	HCDN
12027500	CHEHALIS RIVER NEAR GRAND MOUND, WASH.	WA	46.776111	-123.034444	2318.05	HCDN
12035000	SATSOP RIVER NEAR SATSOP, WA	WA	47.001944	-123.493611	774.41	HCDN
12039500	QUINAULT RIVER AT QUINAULT LAKE, WASH.	WA	47.457778	-123.888056	683.76	HCDN
12040500	QUEETS RIVER NEAR CLEARWATER, WASH.	WA	47.538889	-124.313889	1152.55	HCDN
12041200	HOH RIVER AT US HIGHWAY 101 NEAR FORKS, WA	WA	47.806747	-124.251033	655.80	HCDN
12043000	CALAWAH RIVER NEAR FORKS, WA	WA	47.960077	-124.392987	337.18	HCDN
12048000	DUNGENESS RIVER NEAR SEQUIM, WASH.	WA	48.014444	-123.131389	404.04	HCDN
12054000	DUCKABUSH RIVER NEAR BRINNON, WASH.	WA	47.684167	-123.010278	172.24	HCDN
12056500	NF SKOKOMISH R BLW STRCSE RPDS NR HDSPRT, WASH.	WA	47.514444	-123.328611	148.15	HCDN
12073500	HUGE CREEK NEAR WAUNA, WA	WA	47.389262	-122.699024	15.67	HCDN
12082500	NISQUALLY RIVER NEAR NATIONAL, WA	WA	46.752608	-122.083719	350.00	HCDN
12092000	PUYALLUP RIVER NEAR ELECTRON, WA	WA	46.903716	-122.035109	240.88	HCDN
12095000	SOUTH PRAIRIE CREEK AT SOUTH PRAIRIE, WA	WA	47.139546	-122.092613	205.84	HCDN
12114500	CEDAR RIVER BELOW BEAR CREEK NEAR CEDAR FALLS, WA	WA	47.342054	-121.548986	66.56	HCDN
12115000	CEDAR RIVER NEAR CEDAR FALLS, WA	WA	47.370107	-121.625101	102.80	HCDN
12115500	REX RIVER NEAR CEDAR FALLS, WA	WA	47.350662	-121.663158	34.72	HCDN
12117000	TAYLOR CREEK NEAR SELLECK, WA	WA	47.386491	-121.846222	44.85	HCDN
12134500	SKYKOMISH RIVER NEAR GOLD BAR, WASH.	WA	47.837500	-121.665556	1385.65	HCDN
12141300	MIDDLE FORK SNOQUALMIE RIVER NEAR TANNER, WA	WA	47.485940	-121.647881	401.51	HCDN
12143600	SF SNOQUALMIE RIVER AT EDGEWICK, WA	WA	47.452604	-121.717885	165.01	HCDN
12144000	SF SNOQUALMIE RIVER AT NORTH BEND, WA	WA	47.492880	-121.790111	210.11	HCDN
12144500	SNOQUALMIE RIVER NEAR SNOQUALMIE, WASH.	WA	47.545278	-121.841111	971.25	HCDN
12145500	RAGING RIVER NEAR FALL CITY, WA	WA	47.539824	-121.909005	78.98	HCDN
12147500	NORTH FORK TOLT RIVER NEAR CARNATION, WA	WA	47.712324	-121.788727	102.89	HCDN
12147600	SOUTH FORK TOLT RIVER NEAR INDEX, WA	WA	47.706771	-121.600107	14.13	HCDN
12167000	NF STILLAGUAMISH RIVER NEAR ARLINGTON, WA	WA	48.261492	-122.047641	683.75	HCDN
12175500	THUNDER CREEK NEAR NEWHALEM, WA	WA	48.672629	-121.072896	273.82	HCDN
12178100	NEWHALEM CREEK NEAR NEWHALEM, WA	WA	48.655959	-121.238456	69.68	HCDN

Continued on next page

Table A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
12186000	SAUK R ABV WHITECHUCK R NR DARRINGTON, WASH.	WA	48.168889	-121.469444	393.68	HCDN
12189500	SAUK RIVER NEAR SAUK, WASH.	WA	48.424722	-121.567222	1849.26	HCDN
12401500	KETTLE RIVER NEAR FERRY, WASH.	WA	48.981389	-118.765278	5749.80	HCDN
12404500	KETTLE RIVER NEAR LAURIER, WASH.	WA	48.984444	-118.215278	9842.00	HCDN
12409000	COLVILLE RIVER AT KETTLE FALLS, WASH.	WA	48.594444	-118.061389	2608.13	HCDN
12422500	SPOKANE RIVER AT SPOKANE, WASH.	WA	47.659444	-117.448056	11111.10	HCDN
12424000	HANGMAN CREEK AT SPOKANE, WASH.	WA	47.652778	-117.448611	1784.51	HCDN
12431000	LITTLE SPOKANE RIVER AT DARTFORD, WASH.	WA	47.784722	-117.403333	1722.35	HCDN
12442500	SIMILKAMEEN RIVER NEAR NIGHTHAWK, WASH.	WA	48.984722	-119.617222	9194.50	HCDN
12445000	OKANOGAN RIVER NEAR TONASKET, WASH.	WA	48.632500	-119.460556	18803.40	HCDN
12447200	OKANOGAN RIVER AT MALOTT, WASH.	WA	48.280556	-119.703056	20927.20	HCDN
12447390	ANDREWS CREEK NEAR MAZAMA, WASH.	WA	48.823056	-120.144722	57.24	HCDN
12449950	METHOW RIVER NR PATEROS, WASH.	WA	48.077500	-119.983889	4589.48	HCDN
12451000	STEHEKIN RIVER AT STEHEKIN, WASH.	WA	48.329722	-120.690556	831.39	HCDN
12452800	ENTIAT RIVER NEAR ARDENVOIR, WASH.	WA	47.818611	-120.421944	525.77	HCDN
12457000	WENATCHEE RIVER AT PLAIN, WASH.	WA	47.763056	-120.665000	1530.69	HCDN
12459000	WENATCHEE RIVER AT PESHASTIN, WASH.	WA	47.583333	-120.612778	2590.00	HCDN
12462500	WENATCHEE RIVER AT MONITOR, WASH.	WA	47.499444	-120.423333	3369.59	HCDN
12465000	CRAB CREEK AT IRBY, WASH.	WA	47.360556	-118.848889	2698.78	HCDN
12488500	AMERICAN RIVER NEAR NILE, WASH.	WA	46.977500	-121.168056	204.35	HCDN
13351000	PALOUSE RIVER AT HOOPER, WASH.	WA	46.758611	-118.147778	6475.00	HCDN
14113000	KLICKITAT RIVER NEAR PITT, WASH.	WA	45.756667	-121.208889	3359.23	HCDN
14216500	MUDDY CREEK BELOW CLEAR CREEK NEAR COUGAR, WA	WA	46.075669	-121.998698	349.52	HCDN
14222500	EAST FORK LEWIS RIVER NEAR HEISSON, WASH.	WA	45.836944	-122.465000	323.75	HCDN
14236200	TILTON RIVER AB BEAR CANYON CREEK NEAR CINEBAR, WA	WA	46.595384	-122.459556	360.99	HCDN
06186500	YELLOWSTONE RIVER AT YELLOWSTONE LK OUTLET, YNP	WY	44.567500	-110.380000	2605.54	HCDN
06188000	LAMAR RIVER NR TOWER FALLS RANGER STATION, YNP	WY	44.927778	-110.393056	1709.40	HCDN
06218500	WIND RIVER NEAR DUBOIS, WYO.	WY	43.578611	-109.759167	600.88	HCDN
06221400	DINWOODY CREEK ABOVE LAKES, NEAR BURRIS, WYO.	WY	43.345556	-109.409444	228.44	HCDN
06224000	BULL LAKE C AB BULL LAKE WYO	WY	43.176944	-109.202222	484.33	HCDN
06278300	SHELL CREEK ABOVE SHELL CREEK RESERVOIR, WYO.	WY	44.508056	-107.403056	59.83	HCDN
06280300	SOUTH FORK SHOSHONE RIVER NEAR VALLEY, WYO.	WY	44.208333	-109.554167	769.23	HCDN
06298000	TONGUE RIVER NR DAYTON WYO	WY	44.849444	-107.303889	528.36	HCDN
06309200	MIDDLE FORK POWDER RIVER NEAR BARNUM, WY	WY	43.577740	-107.138400	117.54	HCDN
06311000	NORTH FORK POWDER RIVER NEAR HAZELTON, WY	WY	44.027778	-107.080278	63.46	HCDN
06622700	NORTH BRUSH CREEK NEAR SARATOGA, WY	WY	41.370278	-106.520000	96.87	HCDN
06623800	ENCAMPMENT RIV AB HOG PARK CR NR ENCAMPMENT WYO	WY	41.023611	-106.824167	188.29	HCDN
06630000	N PLATTE R AB SEMINOE RES NR SINCLAIR WYO	WY	41.872222	-107.056944	10813.25	HCDN
06632400	ROCK CR AB KING CANYON CANAL, NR ARLINGTON, WYO	WY	41.585278	-106.222222	162.91	HCDN

Continued on next page

Table A2.1: Information of selected stations from HCDN and RHBN

STATION ID	STATION NAME	PROV	LATITUDE	LONGITUDE	AREA (SQKM)	NETWORK
06635000	MEDICINE BOW R AB SEMINOE RESERVOIR, NR HANNA,	WY	42.009722	-106.512500	6055.42	HCDN
09188500	GREEN RIVER AT WARREN BRIDGE, NEAR DANIEL, WYOM	WY	43.018889	-110.117500	1212.12	HCDN
09196500	PINE CREEK ABOVE FREMONT LAKE, WYO.	WY	43.030556	-109.769444	196.32	HCDN
09210500	FONTENELLE CR NR HERSCHLER RANCH, NR FONTENELLE	WY	42.096111	-110.415833	393.68	HCDN
09223000	HAMS FORK BELOW POLE CREEK, NEAR FRONTIER, WY	WY	42.110556	-110.708889	331.52	HCDN
10032000	SMITHS FORK NEAR BORDER, WY	WY	42.293333	-110.868056	427.35	HCDN
13010065	SNAKE RIVER AB JACKSON LAKE AT FLAGG RANCH WY	WY	44.098889	-110.667500	1222.29	HCDN
13011500	PACIFIC CREEK AT MORAN, WY	WY	43.851111	-110.516389	437.71	HCDN
13011900	BUFFALO FORK ABOVE LAVA CREEK NEAR MORAN,WY	WY	43.837222	-110.439167	836.57	HCDN
13018300	CACHE CREEK NEAR JACKSON WY	WY	43.452222	-110.703333	27.45	HCDN
13023000	GREYS RIVER AB RESERVOIR, NR ALPINE, WY	WY	43.143056	-110.976111	1160.32	HCDN
08AA009	GILTANA CREEK NEAR THE MOUTH	YT	61.195000	-136.982190	190.00	RHBN
08AC001	TAKHANNE RIVER AT KM 167 HAINES HIGHWAY	YT	60.113441	-136.927640	375.00	RHBN
09AA012	WHEATON RIVER NEAR CARCROSS	YT	60.127781	-134.883610	864.00	RHBN
09AC001	TAKHINI RIVER NEAR WHITEHORSE	YT	60.851109	-135.741100	7050.00	RHBN
09AH004	NORDENSKIOLD RIVER BELOW ROWLINSON CREEK	YT	62.051170	-136.280270	6410.00	RHBN
09BC001	PELLY RIVER AT PELLY CROSSING	YT	62.829720	-136.580570	48900.00	RHBN
09CB001	WHITE RIVER AT KILOMETRE 1881.6 ALASKA HIGHWAY	YT	61.988171	-140.558670	6230.00	RHBN
09EA004	NORTH KLONDIKE RIVER NEAR THE MOUTH	YT	64.001968	-138.596280	1090.00	RHBN
09EB003	INDIAN RIVER ABOVE THE MOUTH	YT	63.770000	-139.629580	2210.00	RHBN
09FC001	OLD CROW RIVER NEAR THE MOUTH	YT	67.634438	-139.696400	13900.00	RHBN
10AB001	FRANCES RIVER NEAR WATSON LAKE	YT	60.473888	-129.118900	12800.00	RHBN

Bibliography

Adam, J. C., Hamlet, A. F., and Lettenmaier, D. P. (2009). Implication of global climate change for snowmelt hydrology in the twenty-first century. *Hydrological Processes* 23, 962–972.

Aksamit, N. O. and Whitfield, P. H. (2019). Examining the pluvial to nival river regime spectrum using nonlinear methods: Minimum delay embedding dimension. *Journal of Hydrology* 572, 851–868.

Ali, G., Tetzlaff, D., Soulsby, C., Mcdonnell, J. J., and Capell, R. (2012). A comparison of similarity indices for catchment classification using a cross-regional dataset. *Advances in Water Resources* 40, 11–22.

Aziz, O. I. and Burn, D. H. (2006). Trend and variability in the hydrological regime of the Mackenzie River Basin. *Journal of Hydrology* 319, 2006.

Barhney, J., Menounos, B., Wei, X., and Curtis, P. J. (2017). Determining annual crysphere storage contributions to streamflow using historical hydrometrics records. *Hydrological Processes* 31, 1590–1601.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* 438, 303–309.

Battin, J., Wiley, M. W., Ruckelshaus, M. H., Palmer, R. N., Korb, E., Bartz, K. K., and Imaki, H. (2007). Projected impacts of climate change on salmon habitat restoration. *Preceedings of the National Academy of Sciences* 104(16), 6720–6725.

Bawden, A. J., Burn, D. H., and Prowse, T. D. (2015). Recent changes in patterns of western Canadian river flow and association with climatic drivers. *Hydrology Research* 46(4), 551–565.

Beechie, T., Buhle, E., Ruckelshaus, M., Fullerton, A., and Holsinger, L. (2006). Hydrologic regime and the conservation of salmon life history diversity. *Biological Convervation* 30, 560–572.

Beven, K. (2020). Deep Learning, hydrological processes and the uniqueness of place. *Hydrological Processes* 34, 3608–3613.

Bintanja, R. and Andry, O. (2017). Towards a rain-dominated Arctic. *Nature Climate Change* 7(4), 263–267.

Bormann, H. (2010). Towards a hydrologically motivated soil texture classification. *Geoderma* 157(3-4), 142–153.

Brabets, T. P. and Walvoord, M. A. (2009). Trends in streamflow in the Yukon River Basin from 1944 to 2005 and the influence of the Pacific Decadal Oscillation. *Journal of Hydrology* 371(1-4), 108–119.

Burn, D. H., Aziz, O. I. A., and Pietroniro, A. (2004). A comparison of trends in hydrological variables for two watersheds in the Mackenzie River Basin. *Canadian Water Resources Journal* 29(4), 283–298.

Burn, D. H. and Hag Elnur, M. A. (2002). Detection of hydrologic trend and variability. *Journal of Hydrology* 2002, 107–122.

Burn, D. H., Sharif, M., and Zhang, K. (2010). Detection of trends in hydrological extremes for Canadian watersheds. *Hydrological Processes* 24, 1781–1790.

Burn, D. H. and Whitfield, P. H. (2017). Changes in cold region flood regimes inferred from long-record reference gauging stations. *Water Resources Research* 53, 2643–2658.

Buttle, J. (2006). Mapping first-order controls on streamflow from drainage basins: the T3 template. *Hydrological Processes* 20(15), 3415–3422.

Carey, S. K., Tetzlaff, D., Seibert, J., Soulsby, C., Buttle, J., Laudon, H., . . ., and Pomeroy, J. W. (2010). Inter-comparison of hydro-climatic regimes across northern catchments: synchronicity, resistance and resilience. *Hydrological Processes* 24, 3591–3602.

Cayan, D. R., Kammerdiener, S. A., Dettinger, M. D., Caprio, J. M., and Peterson, D. H. (2001). Changes in the Onset of Spring in the Western United States. *Bulletin of the American Meteorological Society* 82(3), 399–416.

Chang, H., Jung, L. W., Steele, M., and Gannett, M. (2012). Spatial patterns of March and September streamflow trends in Pacific Northwest streams, 1958-2008. *Geographical Analysis* 44, 177–201.

Cherry, J. E., Knapp, C., Trainor, S., Ray, A. J., Tedesche, M., and Walker, S. (2017). Planning for climate change impacts on hydropower in the Far North. *Hydrology and Earth System Sciences* 21, 133–151.

Chollet, F. et al. (2015). *Keras*. `https://keras.io`.

Chollet, F. (2017). *Deep Learning with Python*. 1st. 3 Lewis Street Greenwich, CT, USA: Manning Publications Co. ISBN: 1617294438.

Cook, B. I., Ault, T. R., and Smerdon, J. E. (2015). Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Science Advances* 1(1), e1400082.

Coopersmith, E. J., Minsker, B. S., and Sivapalan, M. (2014). Patterns of regional hydroclimate shifts: An analysis of changing hydrological regimes. *Water Resources Research* 50, 1960–1983.

Coulibaly, P., Samuel, J., Pietroniro, A., and Harvey, D. (2013). Evaluation of Canadian National Hydrometric Network density based on WMO 2008 standards. *Canadian Water Resources Journal* 38(2), 259–167.

DeBeer, C. M., Wheater, H. S., Carey, S. K., and Chun, K. P. (2016). Recent climatic, cryospheric, and hydrological changes over the interior of western

Canada: a review and systhesis. *Hydrology and Earth System Sciences* 20, 1573–1598.

Dery, S. J., Hernandez-Henrıquez, M. A., Burford, J. E., and Wood, E. F. (2009a). Observational evidence of an intensifying hydrological cycle in northern Canada. *Geophysical Research Letters* 36, L13402.

Dery, S. J., Stah, K., Moore, R. D., Whitfield, P. H., Menounos, B., and Burford, J. E. (2009b). Detection of runoff timing changes in pluvial, nival, and glacial rivers of western Canada. *Water Resources Research* 45, W04426.

Dery, S. J. and Wood, E. F. (2005). Decreasing river discharge in northern Canada. *Geophysical Research Letters* 32, L10401.

Dethier, E. N., Sartain, S. L., Renshaw, C. E., and Magilligan, F. J. (2020). Spatially coherent regional changes in seasonal extreme streamflow events in the United States and Canada since 1950. *Science Advances* 6, eaba5939.

Di Prinzio, M., Castellarin, A., and Toth, E. (2011). Data-driven catchment classification: application to the PUB problem. *Hydrology and Earth System Sciences* 15, 1921–1935.

Dudley, R. W., Hodgkins, G. A., Mchale, M. R., Kolian, M. J., and Renard, B. (2017). Trends in snowmelt-related streamflow timing in the conterminous United States. *Journal of Hydrology* 547, 208–221.

Durocher, M., Requena, A. I., Burn, D. H., and Pellerin, J. (2019). Analysis of trends in annual streamflow to Arctic Ocean. *Hydrological Processes* 33, 1143–1151.

Gobena, A. K. and Gan, T. Y. (2006). Low-frequency variablility in southwestern Canadian stream flow: Link with large-scale climate anomalies. *International Journal of Climatology* 26, 1843–1869.

Goldberger, J., Roweis, S., Hinton, G., and Salakhutdinov, R. (2005). Neighbourhood Components Analysis. *Advances in Neural Information Processing Systems* 17, 513–520.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Nets. *Proceedings of the 27th International Conference on Neural Information Processing Systems* 2, 2672–2680.

Groisman, P. Y., Knight, R. W., and Karl, T. R. (2000). Heavy Precipitation and High Streamflow in the Contiguous United States: Trends in the Twentieth Century. *Bulletin of the American Meteorological Society* 82(2), 219–246.

Haines, A. T., Finlayson, B. L., and McMahon, T. A. (1988). A global classification of river regimes. *Applied Geography* 8, 255–272.

Hall, M. J. and Minns, A. W. (1999). The classification of hydrologically homogeneous regions. *Hydrological Sciences Journal* 44(5), 693–704.

Hamed, K. H. and Rao, R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology* 204, 182–196.

Hernandez-Henriquez, M. A., Sharma, A. R., and Dery, S. J. (2017). Variability and trends in runoff in the rivers of British Columbia's Coast and Insular Mountains. *Hydrological Processes* 31(18), 3269–3282.

Hersbach, B., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., ..., and Thepaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146(730), 1999–2049.

Hicke, J., Lucatello, S., Mortsch, L., Dawson, J., Domínguez Aguilar, M., Enquist, C., Gilmore, E., Gutzler, D., Harper, S., Holsman, K., Jewett, E., Kohler, T., and Miller, K. (2022). North America. In: *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Ed. by H. O. Pörtner, D. Roberts, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, and B. Rama. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Chap. 14.

Hinton, G. E. and Roweis, S. T. (2002). Stochastic Neighbor Embedding. In: *Advances in Neural Information Processing Systems.* Cambridge, MA, USA: MIT Press, 833–840.

Hodgkins, G. A. (2009). Streamflow changes in Alaska between the cool phase (1947-1976) and the warm phase (1977-2006) of the Pacific Decadal Oscillation: The influence of glaciers. *Water Resources Research* 45, W06502.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., ..., and Kaab, A. (2021). Accelerated global glacier mass loss in the early twenty-first century. *Nature* 592, 726–731.

Hussain, M. and Mahmud, I. (2019). pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *The Journal of Open Source Software* 4(39), 1556.

Jones, J. B. and Rinehart, A. J. (2010). The long-term reponse of stream flow to climatic warming in headwater streams of interior Alaska. *Canadian Journal of Forest Research* 40, 1210–1218.

Kattenborn, T., Leitloff, J., Schiefer, F., and Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 173, 24–49.

Kendall, M. (1975). *Rank Correlation Measures.* London: Charles Griffin.

Kennard, M. J., Pusey, B. J., Olden, J. D., Mackay, S. J., Stein, J. L., and Marsh, N. (2010). Classificaiton of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology* 55, 171–193.

Kingma, D. and Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations.*

Kingma, D. and Welling, M. (2014). Auto-encoding variational bayes. *Proceedings of the 2nd International Conference on Learning Representations.*

Knowles, N., Dettinger, M. D., and Cayan, D. R. (2006). Trends in snowfall vs Rainfall in the western United States. *Journal of Climate* 19, 4545–4559.

Kormos, P. R., Luce, C. H., Wenger, S. J., and Berghuijs, W. R. (2016). Trends and sensitivities of low streamflow extremes to discharge timing and magnitude in Pacific Northwest mountain streams. *Water Resources Research* 52, 4990–5007.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. (2018). Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences* 22(11), 6005–6022.

La Frenierre, J. and Mark, B. G. (2014). A review of methods for estimating the contribution of glaciers meltwater to total watershed discharge. *Progress in Physical Geography* 38(2), 173–200.

Laudon, H., Spence, C., Buttle, J., Carey, S. K., McDonnell, J. J., McNamara, J. P., Soulsby, C., and Tetzlaff, D. (2017). Save northern high-latitude catchments. *Nature Geosciences* 10(5), 324–325.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature* 521, 436–444.

Leppi, J. C., DeLuca, T. H., Harrar, S. W., and Running, S. W. (2012). Impacts of climate change on Auguest stream discharge in the Central-Rocky Mountains. *Climatic Change* 112, 997–1014.

Ley, R., Casper, M. C., Hellebrand, H., and Merz, R. (2011). Catchment classification by runoff behaviours with self-organizing maps (SOM). *Hydrology and Earth System Sciences* 15, 2947–2962.

Li, S., Wu, L., Yang, Y., Geng, T., Cai, W., Gan, B., ..., and Ma, X. (2020). The Pacific Decadal Oscillation less predictable under greenhouse warming. *Nature Climate Change* 10(1), 30–34.

Lins, H. F. and Slack, J. R. (1999). Streamflow trends in the United States. *Geophysical Research Letters* 26(2), 227–230.

Liu, H., Yang, J., Ye, M., James, S. C., Tang, Z., Dong, J., and Xing, T. (2021). Using t-distributed Stochastic Neighbor Embedding (t-SNE) for cluster analysis and spatial zone delineation of groundwater geochemistry data. *Journal of Hydrology* 597, 126146.

Luce, C. H. and Holden, Z. A. (2009). Declining annual streamflow distributions in the Pacific Northwest United States, 1948-2006. *Geophysical Research Letters* 36, L16401.

Maas, A. L., Hannun, A. Y., and Ng, A. Y. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proceedings of the 30th International Conference on Machine Learning* 28.

Maaten, L. van der (2009). Learning a Parametric Embedding by Preserving Local Structure. *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics* 5, 384–391.

Maaten, L. van der and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 2579–2605.

MacDonald, G. M. (2010). Water, climate change, and sustainability in the southwest. *Proceedings of the National Academy of Sciences* 107(50), 21256–21262.

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica* 13(3), 245–259.

Mantua, N. J. and Hare, S. R. (2002). The Pacific decadal oscillation. *Journal of Oceanography* 58(1), 35–44.

Mazher, A. (2020). Visualization Framework for High-Dimensional Spatio-Temporal Hydrological Gridded Datasets using Machine-Learning Techniques. *Water* 12, 590.

McClelland, J. W., Dery, S. J., Peterson, B. J., Holmes, R. M., and Wood, E. F. (2006). A pan-arctic evaluation of changes in river discharge during the latter half of the 20th century. *Geophysical Research Letters* 33, L06715.

McDonnell, J. J. and Woods, R. (2004). On the need for catchment classification. *Journal of Hydrology* 299, 03–Feb.

Merz, R. and Bloschl, G. (2005). Flood frequency regionalisation - spatial proximity vs catchment attributes. *Journal of Hydrology* 302(1-4), 283–306.

Mishra, A. K. and Coulibaly, P. (2009). Developmentsin hydrological network design: a review. *Reviews of Geophysics* 47, RG2001.

Moore, R. D., Fleming, S. W., Menounos, B., Wheate, R., Fountain, A., Stahl, K., ..., and Jakob, M. (2009). Glaicer change in western North America: influences on hydrology, geomorphic hazards and water quality. *Hydrological Processes* 23, 42–61.

Moore, R. D., Pelto, B., Menounos, B., and Hutchinson, D. (2020). Detecting the Effects of Sustained Glacier Wastage on Streamflow in Variably Glacierized Catchments. *Frontiers in Earth Science* 8, 136.

Mote, P. W., Hamlet, A. F., Clark, M. P., and Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America. *Bulletin of American Meteorological Society* 86, 39–49.

Musselman, K. N., Addor, N., Vano, J. A., and Molotch, N. P. (2021). Winter melt trends portend widespread declines in snow water resources. *Nature Climate Change* 11(5), 418–424.

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., ..., and Rasmussen, R. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. *Nature Climate Change* 8(9), 808–812.

Nash, J. E. and Sutcliffe, J. V. (1970). River flow regime forcasting through conceptual models. Part I - A discussion of princeiples. *Journal of Hydrology* 10, 282–290.

Olden, J. D. and Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications* 19, 101–121.

Oudin, L., Kay, A., Andreassian, V., and Perrin, C. (2010). Are seemingly physically similar catchments truly hydrologically similar. *Water Resources Research* 46, W11558.

Padarian, J., Minasny, B., and McBratney, A. B. (2019). Using deep learning for digital soil mapping. *Soil* 5(1), 79–89.

Patil, S. and Stieglitz, M. (2012). Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrology and Earth System Sciences* 16, 551–562.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . ., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., and Stromberg, J. C. (1997). The natural flow regime. *BioScience* 47(11), 769–784.

Razavi, T. and Coulibaly, P. (2013). Classification of Ontario watersheds based on physical attributes and streamflow series. *Journal of Hydrology* 493, 81–94.

Regonda, S. K. and Rajagopalan, B. (2005). Seasonal cycle shifts in hydroclimatology over the western United States. *Journal of Climate* 18, 372–384.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature* 566(7743), 195–204.

Ren, H., Cromwell, E., Kravitz, B., and Chen, X. (2019). Using Deep Learning to Fill Spatio-Temporal Data Gaps in Hydrological Monitoring Networks. *Hydrology and Earth System Sciences*, 20–Jan.

Renard, B., Lang, M., Bois, P., Dupeyrat, A., Mestre, O., Niel, H., Sauquet, E., Prudhomme, C., Parey, S., Paquet, E., Neppel, L., and Gailhard, J. (2008). Regional methods for trend detection: Accessing field significance and regional consistency. *Water Resources Research* 44, W08419.

Rocca, J. (2019). *Understanding Variational Autoencoders.* `https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73`. [Online; accessed March 21, 2021].

Rood, S. B., Pan, J., Gill, K. M., Franks, C. G., Samuelson, G. M., and Shepherd, A. (2008). Declining summer flows of Rocky Mountain rivers: Changing seasonal hydrology and probable imapcts on floodplain forests. *Journal of Hydrology* 349, 397–410.

Rood, S. B., Samuelson, G. M., Weber, J. K., and Wywrot, K. A. (2005). Twentieth-century decline in streamflows from the hydrographic apex of North America. *Journal of Hydrology* 306, 215–233.

Sabo, J. L., Sinha, T., Bowling, L. C., Schoups, G. H. W., Wallender, W. W., Campana, M. E., Cherkauer, K. A., Fuller, P. L., Graf, W. L., Hopmans, J. W., Kominoski, J. S., Taylor, C., Trimble, S. W., Webb, R. H., and Wohl, E. E.

(2010). Reclaiming freshwater sustainability in the Cadillac Desert. *Preceedings of the National Academy of Sciences* 107(50), 21263–21270.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G. (2011). Catchment classification: emperical anlaysis of hydrological similarity based on catchment function in the eastern USA. *Hydrological and Earth System Sciences* 15, 2895–2911.

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., and Carrillo, G. (2014). Characterizing hydrological change through catchment classification. *Hydrology and Earth System Sciences* 18, 273–285.

Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association* 63(324), 1379–1389.

Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resources Research* 54(11), 8558–8593.

Singh, V. P. (1997). The use of entropy in hydrology and water resources. *Hydrological Processes* 11, 587–626.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., . . ., and Zehe, E. (2003). IAHS decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal* 48(6), 857–880.

Smith, S. L., Romanovsky, V. E., Lewkowicz, A. G., Burn, C. R., Allard, M., Clow, G. D., Yoshikawa, K., and Throop, J. (2010). Thermal State of Permafrost in North America: A Contribution to the International Polar Year. *Permafrost and Periglacial Processes* 21(2), 117–135.

Snelder, T. H., Biggs, B. J. F., and Woods, R. A. (2005). Improved eco-hydrological classification of rivers. *River Research and Applications* 21, 609–628.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, L., and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15, 1929–1958.

St. Jacques, J. M., Huang, Y. A., Zhao, Y., Lapp, S. L., and Sauchyn, D. J. (2014). Detection and attribution of variablity and trends in steramflow records from the Canadian Prairie Provinces. *Canadian Water Resources Journalaa* 39(3), 270–284.

St. Jacques, J. M. and Sauchyn, D. J. (2009). Increasing winter baseflow and mean annual streamflow from possible permafrost thawing in the Northwest Territories, Canada. *Geophysical Research Letters* 36, L01401.

St. Jacques, J. M., Sauchyn, D. J., and Zhao, Y. (2010). Northern Rocky Mountain streamflow records: Global warming trends, human impacts or natural variability. *Geophysical Research Letters* 37, L06407.

Stewart, I. T., Cayan, D. R., and Dettinger, M. D. (2005). Changes toward earlier streamflow timing across western North America. *Journal of Climate* 18, 1136–1154.

Sungmin, O. and Orth, R. (2021). Global soil moisture data derived through machine learning trained with in-situ measurements. *Scientific Data* 8(1), 1–14.

Toth, E. (2013). Catchment classification based on characterization of streamflow and precipitation time series. *Hydrology and Earth System Sciences* 17, 1149–1159.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography Compass* 1/4, 901–931.

Walvoord, M. A. and Striegl, R. G. (2007). Increased groundwater to stream discharge from permafrost thawing in the Yukon River basin: Potential impacts on lateral export of carbon and nitrogen. *Geophysical Research Letters* 34, L12402.

Wang, J., Yuan, Q., Shen, H., Liu, T., Li, T., Yue, L., ..., and Zhang, L. (2020). Estimating snow depth by combining satellite data and ground-based observations over Alaska: A deep learning approach. *Journal of Hydrology* 585, 124828.

Whitfield, P. H. and Cannon, A. J. (2000). Recent variations in climate and hydrology in Canada. *Canadian Water Resources Journal* 25(1), 19–65.

Whitfield, P. H., Moore, R. D., Fleming, S. W., and Zawadzki, A. (2010). Pacific decadal oscillation and the hydroclimatology of western Canada - Review and Prospects. *Canadian Water Resources Journal* 35(1), 1–28.

Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T., Bolles, K., Baek, S. H., Badger, A. M., and Livneh, B. (2020). Large contribution from anthropogenic warming to an emerging North American megadrought. *Science* 368, 314–318.

Williams, J. E., Isaak, D. J., Imhof, J., Hendrickson, D. A., and McMillan, J. R. (2015). Cold-water fishes and climate change in North America. *Reference Module in Earth Systems and Environmental Sciences.*

Woo, M. and Thorne, R. (2003). Commnet on 'Detection of hydrologic trends and variability' by Burn, D. H. and Hag Elnur, M. A., 2002. Journal of Hydrology 255, 107-122. *Journal of Hydrology* 277, 150–160.

Yue, S., Pilon, P., Phinney, B., and Cavadias, G. (2002). The influence of autcorrelation on the ability to detect trend in hydrological series. *Hydrological Processes* 16, 1807–1829.

Yue, S. and Wang, C. (2002). Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test. *Water Resources Research* 38(6), 41–47.

Yue, S. and Wang, C. (2004). The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. *Water Resources Management* 18, 201–218.

Zemp, M., Huss, M., Thibert, E., Eckert, N., McNabb, R., Huber, J., ..., and Cogley, J. G. (2019). Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* 568, 382–386.

Zhang, X., Harvey, K. D., Hogg, W. D., and Yuzyk, T. R. (2001). Trends in Canadian streamflow. *Water Resources Research* 37(4), 987–998.