

The University of Maine

DigitalCommons@UMaine

Electronic Theses and Dissertations

Fogler Library

Summer 8-2022

On the application of dynamical measures of hydrologic response to prediction and similarity assessment in watersheds

Prashanta Bajracharya
prashanta.bajracharya@maine.edu

Follow this and additional works at: <https://digitalcommons.library.umaine.edu/etd>



Part of the [Civil Engineering Commons](#), and the [Other Civil and Environmental Engineering Commons](#)

Recommended Citation

Bajracharya, Prashanta, "On the application of dynamical measures of hydrologic response to prediction and similarity assessment in watersheds" (2022). *Electronic Theses and Dissertations*. 3667.
<https://digitalcommons.library.umaine.edu/etd/3667>

This Open-Access Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

ON THE APPLICATION OF DYNAMICAL MEASURES OF
HYDROLOGIC RESPONSE TO PREDICTION AND SIMILARITY
ASSESSMENT IN WATERSHEDS

By

Prashanta Bajracharya

B.Eng. Tribhuvan University, 2014

M.S. University of Maine, 2019

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Civil Engineering)

The Graduate School

The University of Maine

August 2022

Advisory Committee:

Shaleen Jain, Professor of Civil and Environmental Engineering, Advisor

Sean MC Smith, Associate Professor, School of Earth and Climate Sciences

Lauren Ross, Assistant Professor of Civil and Environmental Engineering

Vikrant Jain, Professor at Indian Institute of Technology, Gandhinagar

Daniel Coker, Senior Spatial Scientist at The Nature Conservancy

© Copyright 2022 Prashanta Bajracharya
All Rights Reserved

**ON THE APPLICATION OF DYNAMICAL MEASURES OF
HYDROLOGIC RESPONSE TO PREDICTION AND SIMILARITY
ASSESSMENT IN WATERSHEDS**

By Prashanta Bajracharya

Dissertation Advisor: Shaleen Jain

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Civil Engineering)
August 2022

The Prediction in Ungauged Basins (PUB) initiative set out to improve the understanding of hydrological processes with an aim of improving hydrologic models for application in ungauged basins. With a majority of basins around the world essentially ungauged, this suggests the need to shift from calibration-based models that rely on observed streamflow data to models based on process understanding. This is especially important in natural infrastructure planning projects such as investments in the conservation of wetlands across the watershed, where the lack of streamflow data hinders the quantification of their benefits (such as flood attenuation), resulting in a difficulty in prioritization. This research sought to contribute to this growing body of literature by (a) developing visual tools and metrics for assessing flow dynamics and flood attenuation benefits of wetlands in relation to their position in the watershed, (b) examining distribution-based topographic metrics in regard to their efficacy in predicting hydrologic response and providing a methodology for examining other metrics in future studies, (c) building robust functional forms for two important catchment metrics: the width function

and hypsometric curve, and (d) devising a hierarchical clustering approach to assess hydrological similarity and find analogous basins that is computationally efficient and has a potential for large-scale applications. Taken together, this study paves the way toward an analytical formulation of the geomorphological instantaneous unit hydrograph (GIUH) that can be used to assess the hydrological behavior in ungauged or data-scarce basins.

Three journal publications that resulted from the research work are listed below:

- Bajracharya, P. and Jain, S. (2020). Estimation of watershed width function: A statistical approach using LiDAR data. *Stochastic Environmental Research and Risk Assessment*, pages 1-15.
- Bajracharya, P. and Jain, S. (2021). Characterization of drainage basin hypsometry: A generalized approach. *Geomorphology*, page 107645.
- Bajracharya, P. and Jain, S. (2022). Hydrologic similarity based on width function and hypsometry: An unsupervised learning approach. *Computers & Geosciences*, page 105097.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Shaleen Jain for his support and guidance throughout this process. He is an inspiration to me, and I appreciate everything he has done for me during all these years. I would also like to thank my dissertation committee members, Dr. Sean Smith, Dr. Lauren Ross, Dr. Vikrant Jain, and Dan Coker for their valuable insights and suggestions.

I would also like to deeply thank my parents. I am who I am today because of their love and constant support. I am forever grateful to them. Finally, I would like to thank my friends. Life is hard. But having true friends makes it slightly less hard. I count myself lucky to have had some really great friends in my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	ix
LIST OF FIGURES	x
1. INTRODUCTION	1
2. VISUALIZATION TOOLS AND METRICS FOR ASSESSING FLOOD ATTENUATION BENEFITS OF WETLANDS	9
2.1 Chapter Abstract	9
2.2 Introduction	9
2.3 Study Area and Data	12
2.4 Metrics and Tools for Accessing Wetlands	15
2.4.1 Storage Capacity	15
2.4.2 Storage Potential	17
2.4.3 Inflow Dynamics	18
2.4.3.1 Stream Network	18
2.4.3.2 Altitude	19
2.4.4 Connectivity	21
2.4.5 Co-benefits	23
2.5 Concluding Remarks	26
3. ESTIMATION OF WATERSHED WIDTH FUNCTION	28

3.1	Chapter Abstract	28
3.2	Introduction	28
3.3	Background.....	31
3.3.1	LiDAR-based Terrain Data	31
3.3.1.1	Characteristics.....	31
3.3.1.2	Use Case: Upper Sheepscot River Basin, Maine, USA	32
3.3.2	Geomorphological characteristics based on stream networks	34
3.4	Estimation of Width Functions	37
3.4.1	Approach	38
3.4.1.1	Probability Distributions	38
3.4.1.2	Finite Mixture Models	40
3.4.1.3	Divergence Measures	41
3.4.2	Application	42
3.4.2.1	Comparison of Distributions.....	42
3.4.2.2	Examination of Finite Mixtures.....	43
3.4.2.3	Measure of Width Function Similarity.....	47
3.4.2.4	Computation Time	49
3.5	Discussion	50
3.6	Concluding Remarks.....	53
4.	CHARACTERIZATION OF DRAINAGE BASIN HYPSONOMETRY.....	55
4.1	Chapter Abstract	55
4.2	Introduction	55

4.3	Data and Use Cases	58
4.3.1	Data	58
4.3.2	Congo River Basin	59
4.3.3	Narmada River Basin	60
4.3.4	Yukon River Basin	62
4.4	A Review of Area-Elevation Relationships	63
4.4.1	Hypsometric Curve	63
4.4.2	Hypsometric Integral	66
4.4.3	Hypsometric Functions	68
4.5	A Generalized Hypsometric Function	73
4.5.1	Generalized Formulation	73
4.5.2	Parameter Interpretation	73
4.5.3	Inter-comparison of Functional Forms	78
4.5.4	Potential Applications	82
4.6	Summary	87
5.	HYDROLOGIC SIMILARITY BASED ON WIDTH FUNCTION AND HYPSONOMETRY	91
5.1	Chapter Abstract	91
5.2	Introduction	91
5.3	Study Area and Data	95

5.4	Methods	97
5.4.1	Hydrological Similarity	97
5.4.2	Dynamical Representation of Watershed Morphometry	98
5.4.2.1	Width Function	98
5.4.2.2	Hypsometric Curve.....	100
5.4.2.3	Geomorphological Instantaneous Unit Hydrograph	101
5.4.3	Functional Estimation	104
5.4.3.1	Width Function	104
5.4.3.2	Hypsometric Curve.....	106
5.4.4	Hierarchical Clustering	106
5.5	Results	109
5.5.1	Width Function Clusters	109
5.5.1.1	Hierarchical Clustering.....	109
5.5.1.2	Analysis of Outliers	111
5.5.1.3	Analysis of Clusters	112
5.5.2	Hypsometric Function Clusters	113
5.5.2.1	Hierarchical Clustering.....	113
5.5.3	Joint Analysis of Hierarchical Clustering of Width Functions and Hypsometric Curves	116
5.6	Discussion	117
5.7	Conclusion	121
6.	EXAMINING TOPOGRAPHY-BASED METRICS IN RELATION TO HYDROLOGICAL RESPONSE AT EVENT SCALE	123

6.1	Chapter Abstract	123
6.2	Introduction	124
6.3	Study Area and Data	126
6.4	Approach and Methods.....	128
6.4.1	Topographic Metrics.....	128
6.4.2	Hydrograph Properties	133
6.4.3	Spatial Analysis and Similarity Assessment	136
6.5	Results and Discussion	137
6.5.1	Interrelationships Among Catchment Metrics	137
6.5.2	Assessment of Catchment Metrics in Relation to Hydrograph Properties	141
6.6	Concluding Remarks.....	144
7.	CONCLUSION.....	148
	REFERENCES	153
	APPENDIX A – SUPPLEMENTAL MATERIALS FOR CHAPTER 3	180
	APPENDIX B – SUPPLEMENTAL MATERIALS FOR CHAPTER 4	182
	APPENDIX C – SUPPLEMENTAL MATERIALS FOR CHAPTER 5	188
	APPENDIX D – SUPPLEMENTAL MATERIALS FOR CHAPTER 6	189
	BIOGRAPHY OF THE AUTHOR	192

LIST OF TABLES

3.1	Four divergence metrics to assess the similarity between probability density functions.....	42
4.1	A chronological review of quantitative approaches for the estimation of hypsometry in drainage basins	70
5.1	Influence of hypsometry on hydrological response.....	102
6.1	Catchment metrics used in this study.....	129
6.2	Statistical significance of Pearson correlation between similarity based on the catchment metrics and based on hydrograph properties	143
B.1	Pearson correlation coefficients between the parameter values of Generalized Hypsometric fits and the hypsometric integral (HI) values	185
D.1	Statistical significance of Pearson correlation between similarity based on the catchment metrics and based on all hydrograph properties (including overlapping sub-basins.....	191

LIST OF FIGURES

2.1	Map of the upper Sheepscot River basin	14
2.2	Relationship between wetland area and above-water volume	16
2.3	Distribution of compound topographic index for non-wetland areas, wetland areas, and wetland inlet points	17
2.4	Width function	19
2.5	Hypsometric curve	21
2.6	Average slope along flow distance	21
2.7	Proportion of upstream wetland area along the flow path	23
2.8	Elevation profile of major streams along with locations of wetlands along the flow path	24
2.9	Variety of potential benefits provided by wetlands	25
3.1	Regional map showing the upper Sheepscot River Basin, Maine	32
3.2	The delineated channel network for upper Sheepscot River Basin based on LiDAR data	34
3.3	Channel network and width function for a subwatershed in the upper Sheepscot River Basin	36
3.4	Sample width functions for selected sub-basins	38
3.5	Comparison of unimodal Beta, Kumaraswamy, and truncated skew-normal distribution fits	43
3.6	Finite mixture of truncated skew-normal distributions	44

3.7	Finite mixture of truncated skew-normal fits for selected width functions	45
3.8	Model selection for width functions based on the truncated <i>mSN</i> distribution	46
3.9	The width function with a five-component <i>mSN</i> fit, and the BIC plots for the best model selection	47
3.10	Distribution of divergence metrics	48
3.11	Cross-comparison of four basins, based on width functions	49
3.12	Total processing time for parameter estimation	50
4.1	Regional map showing the study regions	61
4.2	The scaled hypsometric curve.....	63
4.3	Sample hypsometric curves for selected sub-basins.....	65
4.4	Map showing the distribution of hypsometric integral (HI) values in sub-basins in the three study regions.....	68
4.5	Strahler fit for three sample sub-basins	72
4.6	Synthetic plots of the Generalized Hypsometric Function with a range of values for the parameters	75
4.7	Distribution of parameter values for Strahler and Generalized Hypsometric function fits for each sub-basin	76
4.8	Hypsometric curves of northern plateaus of the Congo River basin and their parameter distribution for Generalized Hypsometric fits	78

4.9	Comparison of fits for the Strahler function, the third-order Polynomial function, the Sigmoidal function, the Double Exponential function, and the Generalized Hypsometric function for three sample sub-basins	79
4.10	Comparison of the root mean square errors (RMSE) for the Generalized Hypsometric function.....	81
4.11	Basin similarity assessment	83
4.12	Distribution of discordance index (DI) values	84
4.13	Hexagonal plot showing the joint distribution of discordance index (DI) values and the difference in hypsometric integral (HI) values	84
5.1	Map of Narmada river basin and its location.....	96
5.2	Schematic illustrating the definition of width function	99
5.3	The scaled hypsometric curve.....	101
5.4	Components of the link-based Geomorphological Instantaneous Unit Hydrograp	103
5.5	Flowchart of the study methodology	110
5.6	Determination of the optimal number of width function clusters using gap statistic	111
5.7	L_2 distances between 15 closest neighbors for each width function.....	111
5.8	Width functions in each cluster after removing the outliers	113

5.9	Dendrogram of watershed width functions using hierarchical clustering	114
5.10	Determination of the optimal number of hypsometric function clusters using gap statistic	115
5.11	Basin hypsometric curves	116
5.12	Dendrogram of basin hypsometric functions using hierarchical clustering	118
5.13	Bivariate cluster groups based on the width function and the hypsometric function	119
6.1	Surface topography of Walnut Gulch Experimental Watershed and selected sub-basins.....	127
6.2	Historical storm discharges and some key properties of hydrograph	134
6.3	Averaged hydrograph properties compared with properties of four storms plotted occurring over the entire watershed	135
6.4	Distribution of catchment metrics	138
6.5	Pairwise differences in metrics for sub-basins.....	139
6.6	Factor map of catchment metrics with the first two principal components as the axes	140
6.7	Comparison of similarity based on width function and similarity based on hydrograph properties	142
6.8	Pearson correlation between similarity based on the catchment metrics and based on two hydrograph properties	143

A.1	Distribution of residuals for selected width functions	180
A.2	Comparison of Basin 16 with all other basins with between 20% maximum hydrological distance and less than 40% overlap	181
B.1	Discordance index	184
B.2	Relationship between the three parameters of the Generalized Hypsometric functional fits and hypsometric integral (HI) values	185
B.3	Comparison of Strahler and Generalized Hypsometric function fits for a sample of nine sub-basins	186
C.1	Width functions in each cluster.....	188
D.1	Comparison of similarity based on width function and similarity based on hydrograph properties (including overlapping sub-basins)	190

CHAPTER 1

INTRODUCTION

Hydrological flow regimes are tied to the spatial pattern and distribution of a number of geophysical variables. In basins where streamflow observations are available, hydrologic models can be readily used and calibrated for the purposes of hydrologic prediction. However, the majority of catchments around the world are effectively ungauged (Hrachowitz et al., 2013). Since many of the popular hydrological models rely on parameter fitting through calibration, they are not feasible in these basins due to the absence of streamflow data (Sivapalan, 2003; Hrachowitz et al., 2013). The Prediction in Ungauged Basins (PUB) initiative took strides in improving the understanding of hydrological processes with an aim of improving hydrologic models for application in ungauged basins. As a result, there is a need to transition towards more physics-oriented methods derived that incorporate various relevant metrics of climate, topography, land use, soil, and geology (Wagener et al., 2007; Gupta et al., 2008; Hrachowitz et al., 2013). Current approaches to flow estimation in ungauged basins mainly rely on one of two techniques: (a) using regression models that use readily available or measurable attributes such as area and slope (Asquith and Roussel, 2009; Gotvald et al., 2012; Dudley, 2015); and (b) comparative hydrology that considers catchments characteristics to assess hydrologic similarity and select analogous gauged watersheds as proxies (Blöschl et al., 2013). Often, the regression equations are not based on physical processes and consequently, carry high degrees of uncertainty (Blöschl et al., 2013).

Comparative hydrology is an approach that examines a number of basins to detect patterns of hydrological behavior. While there is no universal classification framework of catchments in hydrology (Blöschl et al., 2013), the self-organizing nature of catchments has been realized (Sivapalan, 2006) and this leads to

discernible patterns that can help improve our understanding of the spatio-temporal heterogeneities of geophysical characteristics and processes. As a result, this approach has become an important tool for the process-based understanding of catchment hydrology. Numerous approaches have been developed that are based on a variety of metrics and indices such as proximity, climatic regions, and catchment properties (Horton, 1932, 1945; Strahler, 1957; Budyko et al., 1974; L’vovich, 1979; Abrahams, 1984; Bras, 1990; Burn and Boorman, 1993; Milly, 1994; Sankarasubramanian and Vogel, 2002; Tung et al., 1997; Rodríguez-Iturbe and Rinaldo, 2001; Aryal et al., 2002; McIntyre et al., 2005; Woods, 2006; Yadav et al., 2007; Wagener et al., 2007; Reichl et al., 2009; Archfield and Vogel, 2010; Oudin et al., 2010; Patil and Stieglitz, 2011, 2012; Razavi and Coulibaly, 2013; Blöschl et al., 2013; Athira et al., 2016; Brunner et al., 2018). While proximity is a commonly used, reliable metric, its applications are limited as it does not allow for the use of catchments that are not close to each other (Patil and Stieglitz, 2012). Various lumped characteristics such as the catchment area, average slope, drainage density, and Strahler ratio have been used for similarity assessment. However, simplifying complex physical phenomena into a single number leads to an inevitable loss of information (Wooldridge and Kalma, 2001; Wagener and Wheater, 2006; Tetzlaff et al., 2009; Chang et al., 2014). A better alternative is to use distribution curves that have been linked to flow processes and runoff generation, many of which have shown promising results for understanding catchment behavior (McGlynn and Seibert, 2003; Moussa, 2008; Booij et al., 2007; Ssegane et al., 2012; Hailegeorgis et al., 2015; Loritz et al., 2019).

The availability of high-resolution terrain datasets can potentially transform the analysis and modeling of land-surface hydrological processes (Liu and Zhang, 2011; Biron et al., 2013; Yang et al., 2014). With the availability of LiDAR data, various catchments metrics can be estimated with greater accuracy. Topography is a key

metric in determining hydrologic response as it drives runoff generation and routing processes (Horton, 1945), and consequently, improved estimation of topography-based metrics from LiDAR data can result in significant improvements in hydrological analysis and modeling. However, the use of these large datasets also increases the computational and storage burdens. Peters-Lidard et al. (2017) argue that the fourth paradigm of hydrology relies on the full exploitation of emerging large datasets to understand Earth system processes. As such, efficient analysis and storage of catchment characteristics are of utmost importance to advance the field of hydrology. Innovative techniques such as explainable artificial intelligence (XAI) (Maksymiuk et al., 2020; Althoff et al., 2021) provide a promising avenue for large-scale applications.

Of particular interest is the potential of the geomorphological instantaneous unit hydrograph (GIUH) (Rodriguez-Iturbe and Valdes, 1979; Kirshen and Bras, 1983; Bras, 1990; Rigon et al., 2016). Since the GIUH approach relies on geomorphological information that can be derived from catchment topography, it has considerable potential for use in ungauged basins. The GIUH formulation is based on the stream network distribution, termed width function (Bras, 1990; Rigon et al., 2016). At present, the width function is computed as an approximated frequency function with appropriate binning. Since the use of width function within the GIUH formulation involves integration, the lack of a functional formulation renders such computations to approximations using summation approaches. Furthermore, the absence of a closed-form width function representation hinders the derivation of an analytical GIUH function. Another important topographic metric is the hypsometric curve (Horton, 1932). While traditionally, the hypsometric curve has been used to study the geomorphic maturity of catchments (Strahler, 1952; Moglen and Bras, 1995; Pedrera et al., 2009; Willgoose, 2018; Guha and Jain, 2020), newer studies have shown its relation with the hydrological behavior of watersheds

(Harlin, 1980, 1984; Howard, 1990; Willgoose and Hancock, 1998; Marani et al., 2001; Luo and Harlin, 2003; Vivoni et al., 2008; Huang and Niemann, 2008a). The existing approaches for a quantitative representation of hypsometry are limited in their application due to either limited goodness of fit or their ineffectiveness in properly capturing the various elements of the hypsometric curve such as the head and the toe region (Bajracharya and Jain, 2021). An improved functional form has numerous potential applications such as in hydrologic rationalization studies and in improving hydrologic response models.

One of the primary beneficiaries of a process-based approach for accessing hydrological response in ungauged or data-scarce basins is in large-scale natural infrastructure (NI) planning projects. While engineering projects with specific application sites can rely on field measurements, watershed-scale planning projects are often limited by the lack of fine-scale observations. For example, investments in natural infrastructures such as wetlands often require prioritization due to limited funds. The lack of flow data at wetland locations can limit the amount of information available to scientists and decision-makers to make informed decisions. As an instance, efficient models that utilize the readily available terrain information to derive response characteristics can improve the quantification of flow dynamics in wetlands, resulting in a better estimation of flood attenuation potential. This can greatly help planners and decision-makers evaluate the benefits of wetlands in a sustainable development framework.

To this end, this work seeks to advance the current sciences in a few ways as discussed next.

1. The development of visualization tools and metrics for the assessment of flood attenuation benefits of wetlands. Since wetlands are often ungauged, these tools need to maximally utilize limited available data to capture the various

key wetland dynamics, including the interaction among wetlands and with the surrounding landscape.

2. The examination of relevant metrics in characterizing hydrological response. A large number of metrics and indices are available in the literature for assessing the hydrological behavior of catchments. As such, a systematic evaluation of these metrics in regards to their efficacy in predicting streamflow response can narrow down this list to a manageable number that can be further studied in detail or utilized as parameters or components in process-oriented hydrological models. As an entry point, the study was limited to topography-based distribution measures due to the importance of topographic factors in hydrology and the widespread availability of remote-sensed terrain data. However, the methodology can be extended to include other metrics in future studies.
3. The development of a functional formulation for some important distribution curves such as the width function and the hypsometric curve. These metrics have been shown to encapsulate various hydrogeomorphic information about catchments, including hydrological response characteristics. A functional form that incorporates the diversity in width function shapes and modalities can pave the way to the development of an analytical GIUH formulation. Furthermore, representing these distribution curves as functional form means that this information can be stored in the form of a small number of parameters, leading to a reduction of data and higher computational efficiency.
4. A physics-informed machine-learning approach for similarity assessment. A computationally efficient approach for finding analogous basins based on hydrological response characteristics of basins can have large-scale applications, including at national and global scales. This enables a large,

potentially global pool of donor basins with streamflow data from which information can be transferred to ungauged basins.

Additional details and a summary of each chapter are described next.

Chapter 2 provides visualization tools and metrics for evaluating the potential of wetlands for their flood attenuation benefits. These include popular metrics such as the compound topographic index (CTI) (Kirkby, 1975; Beven and Kirkby, 1979), hypsometry, and width function along with some novel tools that focus on storage capacity and connectivity. Additionally, a brief discussion on ways to access ecosystem services of wetlands is provided. As stream inflow data is not available in a majority of wetlands, this chapter highlights the need for better means of assessing flow response in ungauged basins.

Chapter 3 describes a framework for the functional estimation of width functions using a mixture of truncated skew-normal (mSN) distributions that captures a wide variety of distribution shapes. Width function is an important hydrologic metric that can be estimated solely from terrain data and represents the basin's response under idealized flow conditions of constant velocity and absence of losses. The graphic form of the width function is less tractable for further analytical applications such as in the derivation of the link-based geomorphological instantaneous unit hydrograph (GIUH), and the presented mSN formulation aims to redress this. The model selection process can be done based on the Bayesian Information Criterion (BIC). The utility of the functional form is demonstrated by identifying hydrologically similar watersheds based on divergence measures applied to the width function estimates. The benefits of the approach in data reduction and the possibility to scale the approach to very-large terrain datasets are demonstrated.

Chapter 4 provides a new, three-parameter functional form for the hypsometric curve. The elevation-area relationship within drainage basins encapsulates many biophysical processes. The understanding of erosional and hydrodynamical processes

also benefits from the quantitative characterization of elevation distribution or basin hypsometry. The head and toe regions of the hydrograph are especially important as they have shown to be related to a variety of basin characteristics (Willgoose and Hancock, 1998). This chapter provides a comprehensive review of current quantitative approaches to hypsometric analysis, ranging from simple indices to functional formulations (for example, Strahler (1952)). The provided functional form was designed to capture different head, body, and toe shapes of the hypsometric curve and detailed analyses of the goodness-of-fit over a large number of basins ($n = 419$) reveal that this function adapts well to diverse elevational profiles. Some key application areas in analyses of catchment similarity and snowmelt modeling that can benefit from the new functional characterization are discussed.

Chapter 5 then demonstrates a physics-informed machine learning approach to extract features that represent the hydrologic dynamics–width function and hypsometric curve. The functional forms developed in the previous chapters are used. The suggested approach uses an unsupervised clustering approach and divergence measures to identify dynamically-similar sub-basin groups. This paves the way toward a flexible and scalable machine learning approach, informed by the physics of surface flow generation and transport in watersheds, that can be used to assess hydrologic similarity and improve prediction at large scales, especially in data-scarce and ungauged watersheds.

In chapter 6, the width function, hypsometric curve, and CTI are tested and evaluated alongside other commonly used distribution-based topographic metrics for their efficacy in predicting hydrologic response. The other metrics include slope, long profile, height above nearest drainage (HAND), (Crave and Gascuel-Oudou, 1997; Rennó et al., 2008) and reduced dissipation per unit length index (rDUNE) (Loritz et al., 2019). The assessment is done based on representative hydrograph properties based on averaging across historical storms. The results indicate that the

width function, hypsometric curve, CTI, and long profile have high degrees of correlation with various hydrograph properties. Furthermore, a principal component analysis shows that these metrics can be grouped into three distinct clusters. These results suggest that a robust hydrological model based on geomorphology such as a geomorphological instantaneous unit hydrograph should utilize hypsometry and CTI as its components alongside the width function.

CHAPTER 2

VISUALIZATION TOOLS AND METRICS FOR ASSESSING FLOOD ATTENUATION BENEFITS OF WETLANDS

2.1 Chapter Abstract

Wetlands are valuable natural infrastructures that provide complementary hydrological benefits such as flood attenuation in tandem with co-benefits such as the regulation of groundwater and soil moisture, coastal protection, maintenance of water quality, carbon sequestration, and support for habitats. Despite their importance, there is a decline in wetland areas around the world. As such, investments are being made in the conservation and restoration of wetlands. Due to limited funding, there is a need to prioritize wetlands to maximize the benefits provided by the investment. This study developed visualization tools and metrics for evaluating the potential of freshwater wetlands for providing flood attenuation benefits and aiding decision-makers in prioritization. These include terrain-based metrics such as the compound topographic index (CTI), width function, and hypsometry along with some novel tools that focus on storage capacity and connectivity. Furthermore, a brief discussion on ways to access other co-benefits of wetlands is provided.

2.2 Introduction

Ecosystem services provide the various benefits to people and the environment obtained from natural infrastructure (Mooney et al., 1997; Assessment et al., 2005) and have gained considerable interest from governments and private sectors (Brauman et al., 2007). The Millennium Ecosystem Assessment (MA) is the largest synthesis of the study around the importance of ecosystem services (Assessment

et al., 2005), however, further research is needed to move from a conceptual to an operational framework for decision making (Carpenter et al., 2006). MA provided a synthesis of the conceptual theory and knowledge supporting the framework of ecosystem services and the next step is to make it practical and applicable for decision-makers through integrative tools (Daily et al., 2000; Falkenmark et al., 2004). Natural infrastructure refers to landscape features that provide benefits similar to built-infrastructure. These include wetlands, floodplains, forests, sand dunes, etc. In addition to providing several benefits, they can also improve the functioning of built infrastructure. (Bennett et al., 2016) Since the funds for investments are always limited, there is a need to prioritize ecosystem services in order to maximize the benefits. This often involves a wide range of competing criteria and requires the balancing of the needs of users with biophysical constraints (Murray et al., 2006). This inevitably leads to trade-offs in the production of services and the protection of ecosystems which need to be carefully analyzed (Brismar, 2002; Brauman et al., 2007).

In this research, we focus on terrestrial hydrological services (in particular, flood attenuation benefits) provided by freshwater wetlands. Wetlands are valuable natural infrastructures that provide complementary hydrological benefits such as the flood management (De Groot et al., 2002; Brauman et al., 2007; Acreman and Holden, 2013) in tandem with many other co-benefits such as regulation of groundwater and soil moisture (Hefting et al., 2004; Xiong et al., 2003), coastal protection (Temmerman et al., 2013; Gedan et al., 2011), maintenance of water quality (Verhoeven et al., 2006; Mitsch et al., 2001), carbon sequestration (Mitsch et al., 2013; Bridgman et al., 2006), and support for habitats (Gibbs, 2000; Dudgeon et al., 2006). These hydrological services are regional services as any changes upstream affect the downstream experience throughout the watershed (Brauman et al., 2007). Consequently, their effects cannot be studied in isolation which can

lead to costly and inefficient solutions that overlook the dynamic interactions between wetlands and the surrounding landscape (Thorslund et al., 2017). Often, this includes the deterioration of wetlands through landuse changes followed by a water detention structure downstream with the same total volume as the wetlands. This approach fails to take the spatial distribution and connectivity of these wetlands and results in diminished benefits. Only by considering the placement of wetlands over the watersheds can their functions and benefits be optimized (Council et al., 2001; Hansen, 2006; Diebel et al., 2008; Palmer, 2009). Furthermore, the lack of flow observations in wetlands means that watershed-level planning requires the use of metrics such as the topographic index that can be calculated from more widely available, remote-sensed datasets (Horvath et al., 2017; Bian et al., 2021).

Despite the importance of wetlands, there has been a steady decline in wetland areas around the world (Davidson, 2014; Mitsch and Gosselink, 2015; Thorslund et al., 2017). The federal regulatory bodies concerning wetlands are established under Section 404 of Clean Water Act. In Maine, the Natural Resources Protection Act (NRPA) of 1988 established a state regulatory authority over wetlands. Wetlands are regulated under the home rule provisions of the Maine Constitution and under Maine's Municipal Shoreland Zoning statute. Local governments have the authority to regulate non-forested wetlands greater than ten acres in size. The 2014 enactment of Chapter 589, "General Fund Bond Issue to Ensure Clean Water and Safe Communities" and the passage of the general fund bond promotes increased investments in natural infrastructure, including in wetlands to improve Maine's freshwater resources. This research takes steps for developing tools and methods to understand the potential for wetlands to support flood attenuation. Around 25 percent of Maine's land area is comprised of wetlands, of which over five million acres are freshwater wetlands (freshwater meadows, bogs, wooded swamps, shrub swamps, and freshwater marshes and floodplains). Current urbanization

trajectories and climate variability are likely to increase the risks of extreme hydrological events (Seneviratne et al., 2006; Bennett et al., 2016). Consequently, the use of natural infrastructure to complement built infrastructure is necessary to develop sustainable solutions. The first step is to develop meaningful ways to quantify, analyze, and integrate relevant information to assess natural infrastructure at the watershed level. Since investment budgets for natural infrastructure conservation and restoration are limited, means for prioritizing them to maximize benefits are necessary. More research is needed to improve the metrics for understanding ecosystem benefits (Bennett et al., 2016; Thorslund et al., 2017). In particular, there is a need to improve the understanding of the interrelation between wetlands and their surroundings and the means to evaluate this interaction (Thorslund et al., 2017). To this end, this study seeks to develop metrics and visualization tools to prioritize wetlands and support decision-making through the morphometric analysis of geospatial data at watershed scales that are linked to various aspects of flood attenuation potential. Due to the availability of terrain data over the world, we focus mostly on developing metrics that can be computed using catchment topography.

2.3 Study Area and Data

The Sheepscot River is a 106 km-long river in the State of Maine, USA that drains into the Atlantic ocean. It starts in the hills of Montville Maine and is joined by the West Branch upstream of North Whitefield. The river continues through the Head Tide Dam and mixes with the incoming tide in an estuary in the town of Wiscasset. The path of the river is influenced by the underlying geology of the Norubega shear zone (Osberg et al., 1985) and the basin landscape is heavily shaped by glaciation (Belknap et al., 1986, 2002). These glaciations occurred during the Late Pleistocene and were followed by the retreating of ice sheets, creating wide,

shallow valleys (Smith and Hunter, 1989). Deglaciation occurred simultaneously as sea-level inundation due to the isostatic depression of the lithosphere (Belknap et al., 1986; Hooke and Winski, 2014). A majority of the basin is forested, with agricultural and low-density residential areas occupying most of the remaining land (Brady, 2007). A number of dams have been built along the river to support various purposes including supporting the timber industry (Halsted, 2002), mining, textile, and grain production. Some dams are believed to obstruct the upstream passage of fish species (Halsted, 2002). In addition to artificial damming, beaver dams continue to influence fluvial morphology (Halsted, 2002; Rosell et al., 2005). The Sheepscot River provides spawning grounds for the remnant population of Atlantic salmon and also supports a number of other species, including brook trouts, sticklebacks, perch, shiners, white-tailed deer, otters, and minks (McLean et al., 2007). The lower Sheepscot River supports lobster fishery and the bait-worm industry (McLean et al., 2007). The Maine State legislature designates a region of the Sheepscot River as an outstanding river segment and provides special protection due to its socioeconomic importance (State of Maine, 1983). To focus on freshwater hydrology, we have constrained the study site to the upper Sheepscot River Watershed (Figure 2.1), a 386 km² basin with the US Geological Survey stream gauging station number 0103800 in North Whitefield set as the watershed outlet.

The geospatial data was obtained from the publicly available Maine GeoLibrary Data Catalog maintained by the Maine Office of GIS (MEGIS) (State of Maine, 2018). MEGIS is an organization within the State of Maine’s Office of Information Technology. The geospatial data includes several socio-economic and ecological layers, such as wildlife and fish habitat, land use and land cover, schools, lakes, and wetlands. The Maine Land Cover Dataset (MELCD) is based on the National Land Cover Database (NLCD), which was refined to the State of Maine requirements using SPOT 5 panchromatic imagery from 2004. Furthermore, this dataset uses a

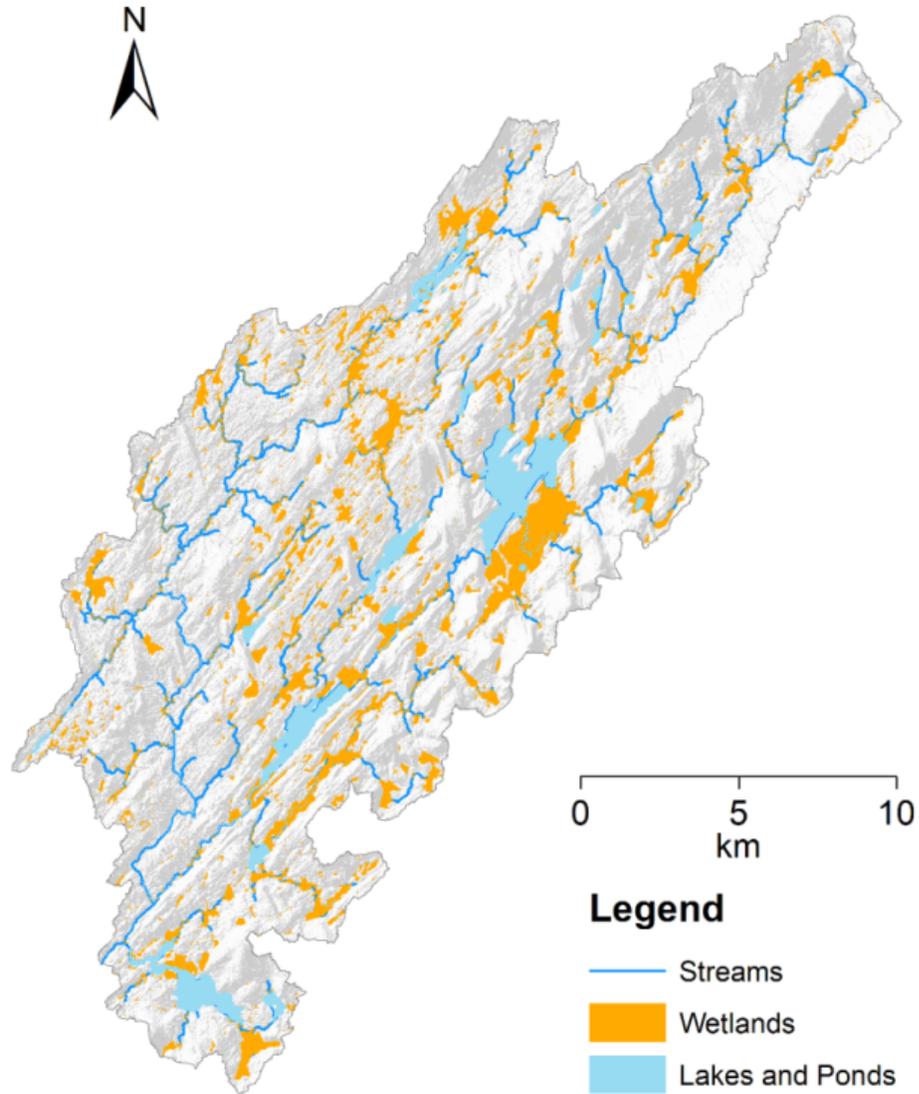


Figure 2.1: Map of the upper Sheepscot River basin. The wetlands, lakes, ponds, and streams are displayed over the shaded relief.

refined classification system to fit the Maine-specific classes. The spatial resolution of this data is 5m. Additionally, MEGIS also provides the high-resolution Light Detection And Ranging- (LIDAR) based Digital Elevation Model (DEM). This data was collected between 2006 and 2012, and has a horizontal and vertical resolution of 2m and 15 cm, respectively. The resolution of the dataset used is one-third arc-second (approximately 10 meters).

The high-resolution LiDAR DEM was first modified to drain through obstructions such as roads (Poppenga et al., 2010). The modified DEM was used to delineate streams using Geographic Information Systems (GIS) techniques. A stream accumulation threshold of 1 km² was used to delineate the streams to closely match the USGS National Hydrography Dataset (NHD). Additionally, a threshold of 1000 m² was used to delineate streams for the visualization of flow paths.

2.4 Metrics and Tools for Accessing Wetlands

2.4.1 Storage Capacity

Wetlands can absorb and store flood waters in depressions, thereby reducing flood surges downstream. As such, the storage capacity of wetlands is a principal indicator of how much flood attenuation benefit they can provide. Since the storage capacity data for most wetlands around the world is not readily available, a common method is to use either wetland area as a proxy for volume, or to use an empirical relationship between volume and area to determine the storage capacity (Gleason et al., 2007; Liu and Wang, 2008; Huang et al., 2011; Shook and Pomeroy, 2011; Shaw et al., 2012; Kessler and Gupta, 2015; Wu and Lane, 2016; Jones et al., 2018). These empirical relationships carry a high degree of uncertainty (Gleason et al., 2007). However, for the purposes of wetland prioritization, actual volumes are not as important as relative volumes. We test the efficacy of using wetland area as a proxy for volumetric capacity. The wetland volumes were computed using the existing LiDAR-based data using GIS techniques similar to the work by Jones et al. (2018). For each wetland polygon, the lowest elevation of the wetland boundary was taken as the highest potential water level in the wetland. The height difference between this elevation and the LiDAR elevation at each cell within the wetland was used to compute the maximum potential volume above water level. These volumes indicate the above-water volumes since LiDAR only captures the terrain above the

water level. Figure 2.2 shows a strong log-linear relationship between wetland area and volume, which is consistent with prevalent literature (Gleason et al., 2007; Kessler and Gupta, 2015; Wu and Lane, 2016). In general, the relationship is of the exponential form:

$$V = kA^m \tag{2.1}$$

where A and V represent the wetland area and volume respectively, and k and m are constants. (Gleason et al., 2007) presents some values for these constants based on physiographic region. In our study area, we found $k = 4.84$ and $m = 0.81134$. Furthermore, the log-linear relationship had a high statistical significance with an R^2 value of 0.88. This indicated that wetland area can generally be used as a reliable relative indicator for its storage capacity.

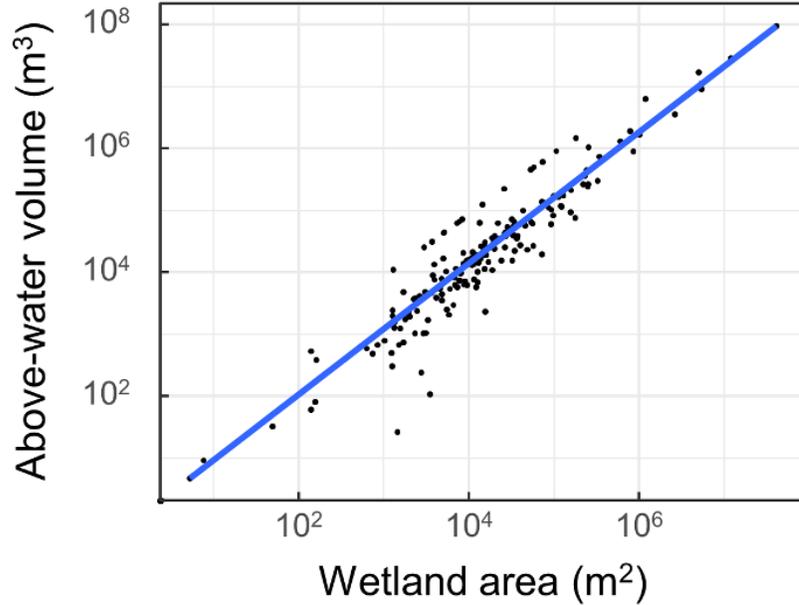


Figure 2.2: Relationship between wetland area and above-water volume. The above-water volume is the volumetric capacity of wetland over the water level, and was computed using LiDAR data. Both axes have been log-scaled. A linear regression line is shown in blue.

2.4.2 Storage Potential

The compound topographic index (CTI), also known as the topographic wetness index (TWI), is defined as the ratio of upstream drainage area to the local slope (Equation 2.2).

$$CTI = \ln \left(\frac{\alpha}{\tan \beta} \right) \quad (2.2)$$

where, at each point (grid cell), α is the contributing area, and β is the slope. This index represents the tendency of water to accumulate based on upstream contribution and ground slope. The hydrological model named TOPMODEL utilizes this index based on the assumption that regions in the watershed with

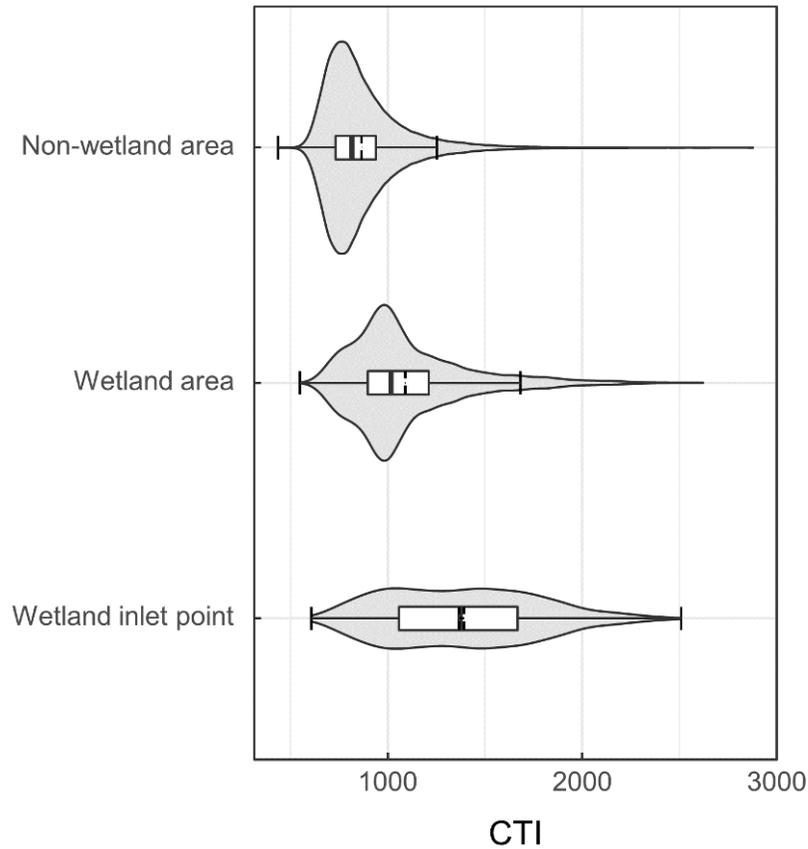


Figure 2.3: Distribution of compound topographic index (CTI) for non-wetland areas, wetland areas, and wetland inlet points. The box plot overlaid on the violin plot demonstrates the distribution. Additionally, the dashed line shows the mean values.

similar CTI values have similar hydrological behavior (Beven and Kirkby, 1979; Beven, 1997, 2011). Furthermore, high CTI areas indicate regions with low slopes and high overland runoff and result in a greater potential to support wetlands. Consequently, CTI has been used as a predictor of potential water accumulation to delineate wetlands Rampi et al. (2014); Horvath et al. (2017); Bian et al. (2021). We suggest that the CTI distribution of wetlands can be used as a proxy for the storage potential of wetlands and can be used for prioritization. However, since each downstream grid point would include the accumulated drainage area of upstream grid points, we instead suggest using one critical point in the wetland as a representative CTI to measure storage potential. We propose the inlet point of the wetland as this critical point as it represents the total potential runoff entering the wetland. The distribution of CTI for wetland and non-wetland areas along with the distribution for the critical points is shown in Figure 2.3. The CTI values for wetlands regions are higher than for non-wetland areas, as is to be expected. However, the CTI values for the critical points are even higher, indicating that this CTI value can be used to compare and prioritize wetlands.

2.4.3 Inflow Dynamics

2.4.3.1 Stream Network

Stream network topology plays a vital role in shaping the hydrologic response of basins. One way to represent the stream network is using the width function. The width function is the distribution of stream links across the watershed (Gupta et al., 1986; Bras, 1990). Under the idealized conditions of contact flow velocity, the width function represents the probability distribution of flow travel times and reflects the watershed's streamflow response (Lashermes and Foufoula-Georgiou, 2007; Moussa, 2008). It is an important hydrological metric and has been linked with the shape of the hydrograph (Kirkby, 1976; Gupta and Waymire, 1983; Troutman and Karlinger,

1984, 1989). The geomorphological instantaneous unit hydrograph approach is a unit hydrograph approach based on the geomorphological structure of the river network and uses the width function as one of its primary components. Consequently, the width function is a powerful tool for predicting hydrologic response in ungauged basins (Moussa, 2008). This can be useful in understanding hydrologic dynamics in wetlands where stream gauge data are not readily available. Comparison of width functions for drainage areas upstream of wetlands and classifying similar wetlands can aid in the prioritization of wetlands. An example of the width function for one of the wetlands in the upper Sheepscot River basin is shown in Figure 2.4 . Here, the x-axis has been scaled to a unit distance. This normalized width function allows for basins to be readily compared for similarity assessment, and a basis for this has been demonstrated by Bajracharya and Jain (2020).

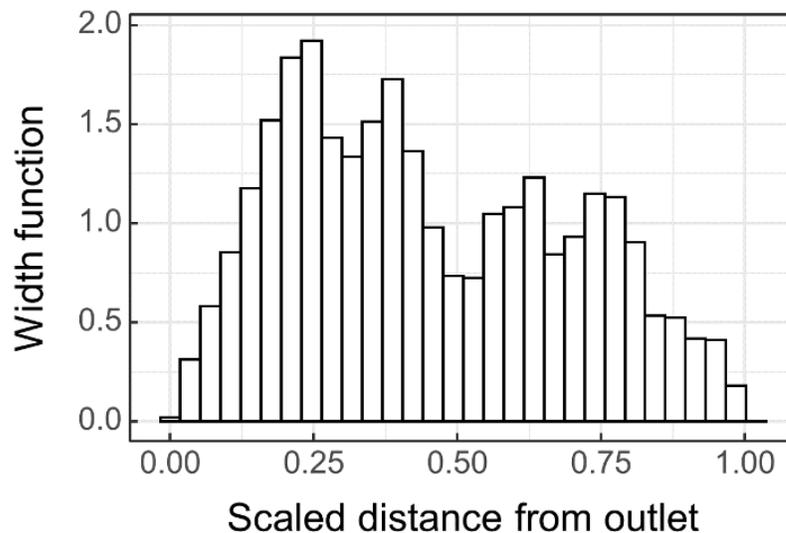


Figure 2.4: Width function. The x-axis of the width function is the flow distance from the outlet. This axis has been scaled to a unit distance.

2.4.3.2 Altitude

Gupta et al. (1986) suggested that altitude is the missing dimension to linking networks with basin hydrology. A hypsometric curve is an area-altitude distribution

curve that shows the proportion of land area that exists at various elevations (Horton, 1932; Langbein, 1947). The curve can be plotted in absolute units or relative units, with the latter allowing for the comparison of basins of different sizes and altitudes (Strahler, 1952; Bajracharya and Jain, 2021). An example of a hypsometric curve plotted with relative units for a wetland in the upper Sheepscot River basin is shown in Figure 2.5. The hypsometric curve is most commonly used for studying the geomorphic maturity of catchments (Strahler, 1952; Moglen and Bras, 1995; Pedrera et al., 2009), with younger basins exhibiting convex-up shapes while mature basins exhibiting concave-up shapes. Due to the influence of topography on drainage patterns and flows, newer studies have linked the hypsometric curve with various hydrological and hydraulic attributes such as the watershed time-to-peak, headward drainage development, mid-basin slope, rate of change in upper and lower basin slope (Harlin, 1984), average channel gradient (Howard, 1990), landscape runoff and erosion processes, planar form of the drainage network (Willgoose and Hancock, 1998), groundwater interactions, fluctuations of the water table (Marani et al., 2001), and surface and subsurface runoff components of basins (Vivoni et al., 2008). Consequently, hypsometric curves of contributing drainage areas of wetlands can be used as another metric to classify and prioritize wetlands.

Alongside elevation, slope plays a vital role in routing stream flow across the watershed. We suggest slope along the flow path as another metric for accessing the hydrological property of basins. By comparing the average slope and distribution along the flow distance (Figure 2.6), basins with similar routing properties can potentially be classified.

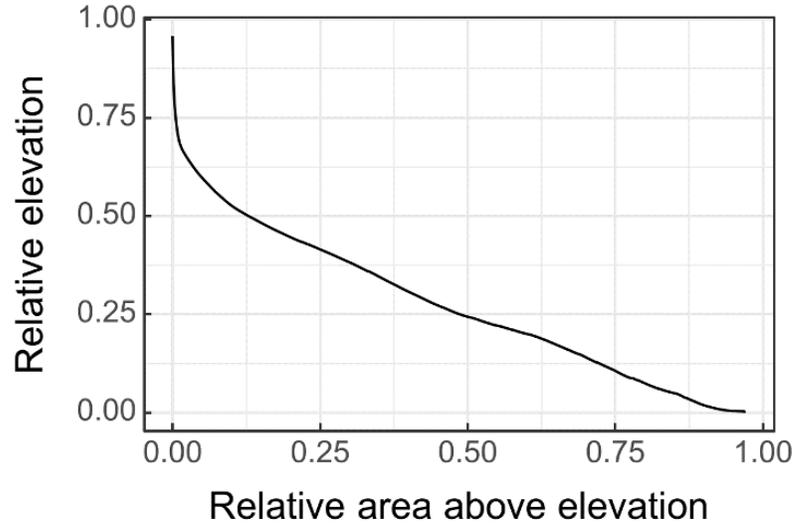


Figure 2.5: Hypsometric curve. Here, both x- and y-axes are scaled to unit values.

2.4.4 Connectivity

While engineering decisions often focus on individual wetland projects, large-scale planning, conservation, and restoration efforts require consideration of the functioning of hydrologically connected system of multiple wetlands within the associated hydrological catchment (Thorslund et al., 2017). Only a broader view of

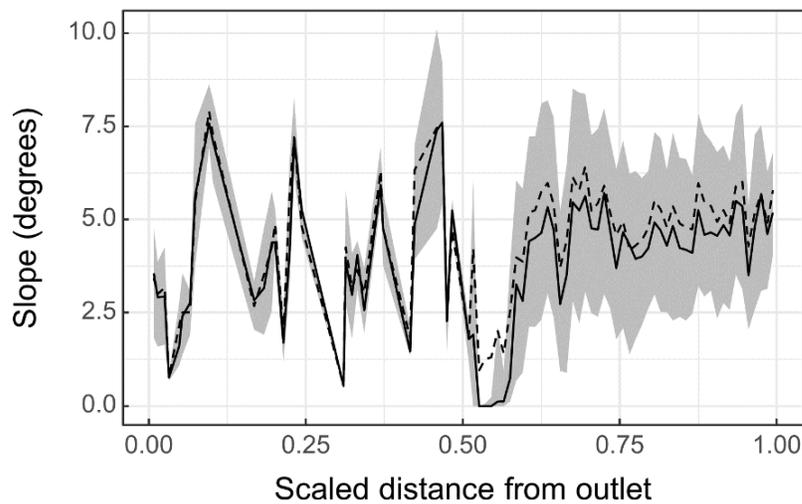


Figure 2.6: Average slope along flow distance. The x-axis has been scaled to a unit distance. The median slope is shown using a solid black line, while the mean is shown using a dashed black line. The grey envelope presents the slope distribution within one standard deviation.

the landscape can provide a fuller context for decision-makers to assess the potential cumulative effects of conservation and restoration efforts on the large-scale wetland diversity and functioning (Preston and Bedford, 1988), including flood mitigation as well as ecosystem services (Cohen et al., 2016). Thorslund et al. (2017) coined the term, "wetlandscape" to denote multiple, hydrologically connected wetlands and their aggregated effects of interacting with other connected wetlands as well as their surrounding landscape. A fragmented approach that fails to account for these dynamics can lead to inefficient and costly solutions (Thorslund et al., 2017). The level of interaction between wetlands is determined by the degree to which they are joined by various transport mechanisms (Alexander et al., 2015). To a large degree, the connectivity of wetlands and their contribution to landscape-scale functioning may depend on the flow path (Thorslund et al., 2017). As a starting point, we suggest exploring the distribution of wetlands along the flow path as an indicator of their connectedness. Figure 2.7 shows the proportion of other wetlands moving upstream from a selected wetland along the flow path. On a relative scale, this plot helps visualize how wetlands are distributed upstream, and on an absolute scale, it can demonstrate the total storage benefits across the given wetlandscape. Comparing the distribution of connected wetlands can be another tool for prioritizing which wetlands are in need of more investment.

Figure 2.8 provides a better, albeit more elaborate, picture of the wetlandscape. The plot shows the elevation and location of wetlands in relation to the major streams inside the entire watershed. This helps visualize the connectivity between different wetlands as well as how they are distributed across the basin. Furthermore, important properties of wetlands such as their area (as a proxy for storage) and location within or outside flood zones are also shown. Wetlands within floodplains and flood-prone areas are more valuable due to their potential to reduce flood damages downstream (Highfield and Brody, 2006). Finally, the upstream drainage

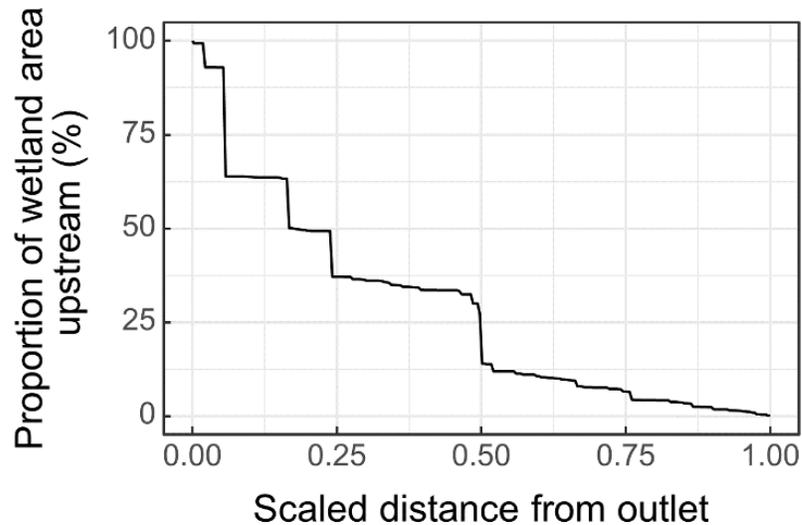


Figure 2.7: Proportion of upstream wetland area along the flow path. The upstream wetland area is shown in percentage, while the flow path is scaled to a unit distance.

area along the stream provides information about potential inflows into wetlands.

This plot can aid in the initial screening of wetlands across the watershed to narrow down a list of potential candidates in need of more attention. These selected wetlands can then be further evaluated using the other tools outlined earlier.

2.4.5 Co-benefits

Finally, we briefly discuss some ways to access wetland co-benefits. Wetlands provide various ecosystem services including water purification, sediment retention, habitat for species, and recreational opportunities. Often, they also hold cultural significance to the local people. The very first step in accessing these benefits is the identification and inventory. Here, we present the approach developed by the Maine State Planning Office for characterizing watershed-scale wetland benefits using geographic information system (GIS) analysis (Office, 2001). The office in conjunction with the Maine Department of Environmental Protection cooperated with other state and federal agencies to develop a set of criteria for identifying wetland functions. Figure 2.9 maps these benefits for the wetlands in the upper Sheepscot River watershed. The flood attenuation benefits are based on location

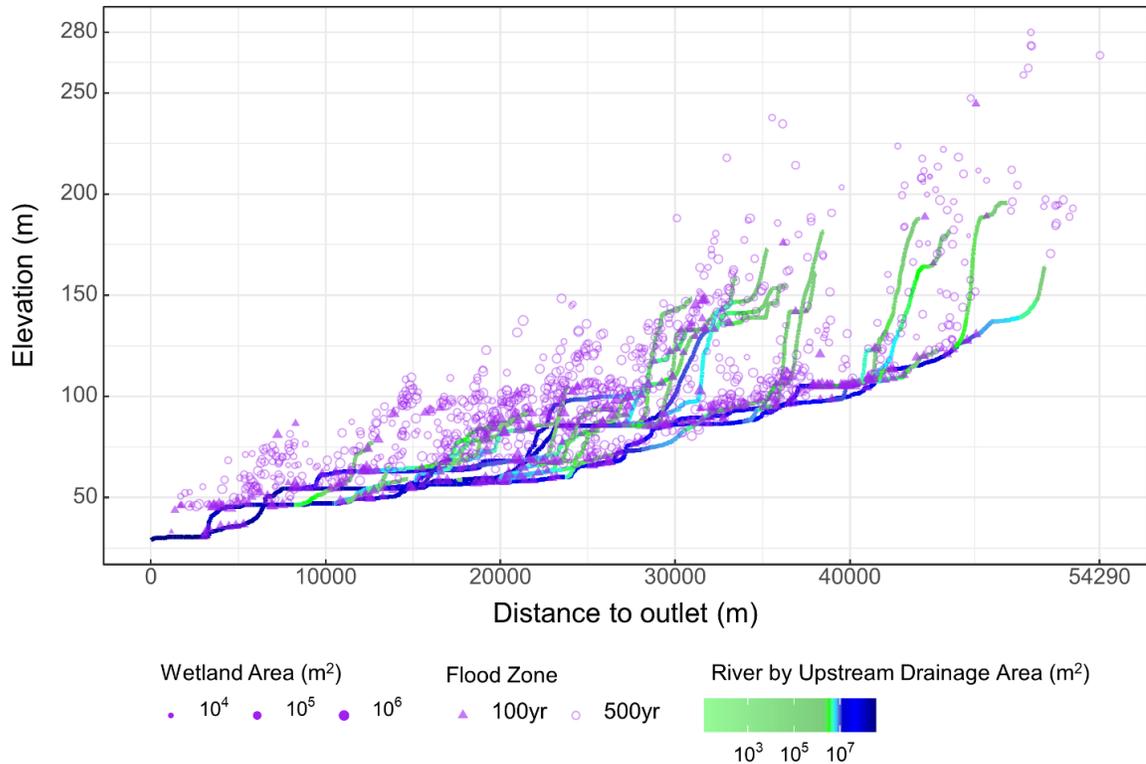


Figure 2.8: Elevation profile of major streams along with locations of wetlands along the flow path. The streams are colored based on the total upstream contributing area. The sizes of the points indicate the wetland areas while their shapes indicate whether they fall within 100-year or 500-year flood zones. These flood zones are based on Federal Emergency Management Agency’s (FEMA) National Flood Hazard Layer (NFHL).

within flood zones or in proximity to water bodies that have a ground slope of less than 3%. Habitat benefits are characterized by wetlands with open water or vegetation classes that are adjacent to habitats supporting rare, threatened, and endangered habitats. Further criteria can be set for specific species and habitats based on their importance in the region. Wetlands with less than 3% ground slope with emergent vegetation and in close proximity to water bodies are classified as beneficial for sediment retention. Wetlands with close proximities to boat ramps and schools were seen as likely candidates for use as educational resources or as wetlands with built-in constituency and were classified as having cultural and

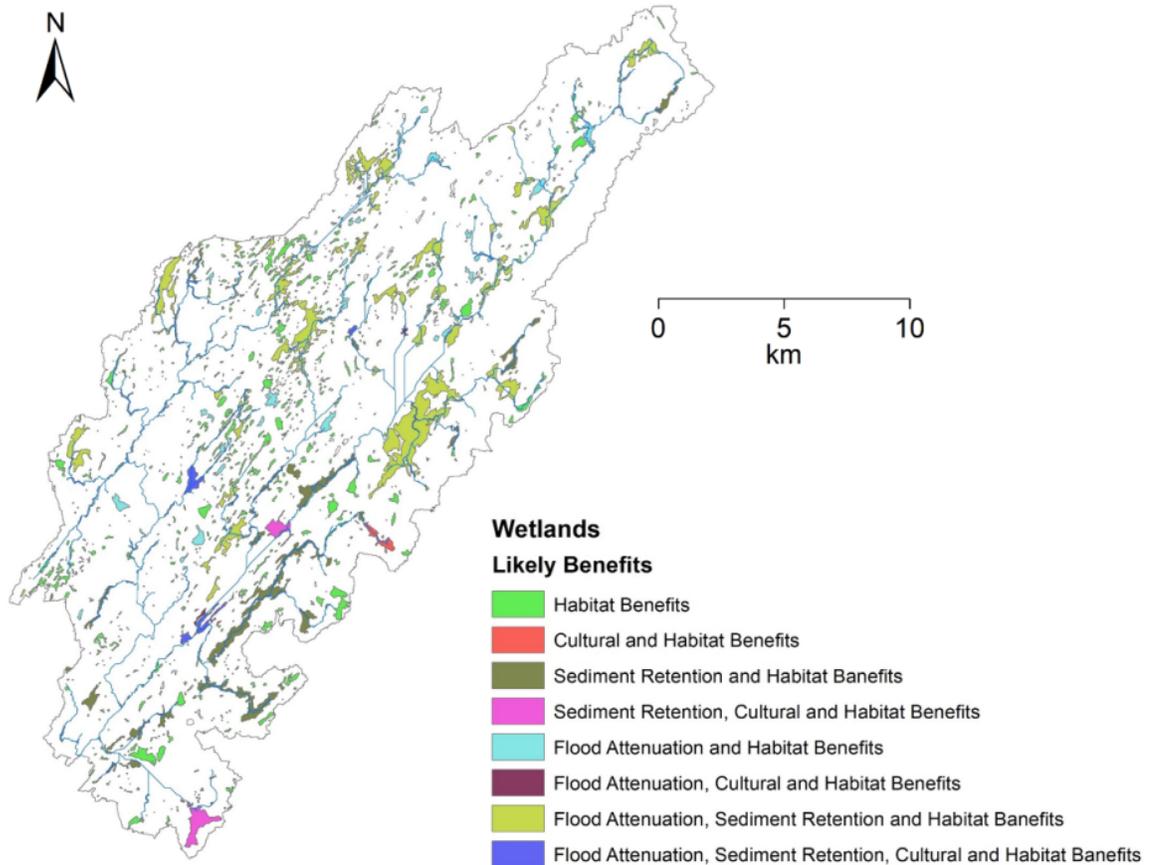


Figure 2.9: Variety of potential benefits provided by wetlands. These benefits are determined through spatial analysis based on criteria set by the Maine State Planning Office for wetland characterization in the Casco Bay Watershed.

educational benefits. Identifying and mapping these benefits provides a qualitative basis for the prioritization of wetlands based on their co-benefits.

Economic valuation of these co-benefits can be powerful tools to quantify their value and help decision-makers prioritize investments. However, since most of the benefits of wetlands do not have market prices, non-market valuation techniques are often required (Woodward and Wui, 2001). Some of the methods include the net factor income method, replacement cost method, travel cost method, contingent valuation method, and hedonic pricing method. Woodward and Wui (2001) provides an overview and analysis of these methods. Newer tools such as the Natural Capital Project's Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST) have

been developed to specifically address the difficulty in the valuation of ecosystem services and have been used in the context of wetlands (Cragg et al., 2011; Flight et al., 2012; Watson et al., 2016; Gittman et al., 2019). While this is still an emerging field in need of more research attention, these studies provide guidance for the valuation of wetland benefits.

2.5 Concluding Remarks

While wetlands provide a range of valuable ecosystem benefits, total wetland areas are still in decline around the world (Davidson, 2014; Mitsch and Gosselink, 2015; Thorslund et al., 2017). This has led to increased investments in wetland conservation and restoration. However, due to limited funds, there is a need to prioritize wetlands to maximize potential benefits. This requires an understanding of the wetland flow dynamics, interconnectedness between wetlands, and interplay between wetlands and the surrounding landscape. In this study, we demonstrated several metrics and visualization tools for evaluating the wetland flow dynamics and potential flood attenuation benefits. These include popular metrics such as the compound topographic index (CTI), width function, and hypsometry along with some novel tools that focus on storage capacity and connectivity. We also provided a brief discussion on ways to access other co-benefits of wetlands. Streamflow observations into wetlands are not widely available. This leads to the difficulty in analyzing the upstream hydrologic characteristics. The metrics developed in this study go beyond what is currently available by suggesting process-based metrics and tools such as width function and hypsometry. Wetland connectivity has been taken keenly into consideration in developing some of these metrics. Taken together, they provide a comprehensive set of additional aids to planners and policy makers in deciding which wetlands should be prioritized for investments in conservation and restoration projects. Finding the best locations for investing in natural

infrastructure involves the analysis of many alternatives to select the optimal choice in a multi-criteria decision making (MCDM) framework (Linkov and Moberg, 2011). Several MCDM techniques such as TOPSIS (the Technique for Order Preference by Similarity to Ideal Solution) and AHP (Analytical Hierarchy Process) have been adapted to prioritize and rank wetlands (Liu et al., 2006; Peñacoba-Antona et al., 2021). To test the accuracy of decision-making models, the results should be tested for stability through sensitivity analysis (Rolander et al., 2003).

While these tools provide an improved understanding of important wetland properties and phenomena, the lack of inflow stream data limits the amount of information that is available to decision-makers, since most wetlands around the world are not gauged. In the following chapters, we improve ways of accessing flow dynamics in the absence of streamflow observations.

CHAPTER 3

ESTIMATION OF WATERSHED WIDTH FUNCTION: A STATISTICAL APPROACH BASED ON LIDAR DATA

3.1 Chapter Abstract

Terrain variability and channel network characteristics critically influence the hydrologic response of watersheds. The width function represents this response under idealized flow conditions of constant velocity and absence of losses, and can be estimated solely using the terrain data. However, the width function in its graphical form is less tractable for further analytical applications such as in the derivation of the link-based geomorphological instantaneous unit hydrograph (GIUH). In this study, we systematically redress these issues in the following manner: (a) develop a framework for the functional estimation of width functions using a mixture of truncated skew-normal distributions that captures a wide variety of distribution shapes, (b) provide a basis for model selection based on the Bayesian Information Criterion (BIC), (c) demonstrate the utility of a functional estimation approach by identifying hydrologically similar watersheds based on divergence measures applied to the width function estimates, and (d) illustrate the utility of efficient statistical estimation of geomorphic functions and metrics, which affords data reduction and can be scaled to very-large terrain datasets.

3.2 Introduction

The recent proliferation of high-resolution terrain data has the potential to transform the analysis and modeling of land-surface hydrological processes (Liu and Zhang, 2011; Biron et al., 2013; Yang et al., 2014). For instance, higher spatial resolution afforded by Light Detection and Ranging (LiDAR) data (often at ≤ 2

meter resolution, as compared to the routinely available 30-meter Digital Elevation Model data) is poised to enable advancements in (a) delineation of channel networks (Sahoo and Jain, 2018) and wetlands (Wu and Lane, 2017), (b) understanding of runoff generation processes (Degetto et al., 2015), and (c) analytical modeling using the geomorphological unit hydrographs (Rodriguez-Iturbe and Valdes, 1979; Gupta et al., 1980; Cheng, 1982; Gupta and Mesa, 1988; Troutman and Karlinger, 1988).

The availability of high-resolution terrain elevation data is of particular interest in improving geomorphological instantaneous unit hydrograph (GIUH) theory-based derivations of hydrologic response (Rodriguez-Iturbe and Valdes, 1979; Kirshen and Bras, 1983; Bras, 1990; Rigon et al., 2016). For example, Hallema and Moussa (2014) devised a hydrologic modeling approach based on the representative elementary hillslope formulation to model distributed runoff processes. In a recent study, Sahoo and Jain (2018) derived a comprehensive set of metrics representing morphometric ratios to model GIUH-based hydrograph parameters; statistical tests to assess differences in the metrics derived based on two spatial resolutions (30 and 90 meters) revealed consistency. As such, the study noted that compared to topography-based models, the GIUH approach appeared less sensitive to spatial resolution. The generalization of their results to finer spatial scales remains an interesting and open question. Furthermore, the GIUH approach is particularly suited to modeling the hydrologic response in ungauged basins (Kumar et al., 2002; Ellouze-Gargouri and Bargaoui, 2012; Sahoo and Saritha, 2015; Swain et al., 2015), wherein terrain information can be used profitably, yet no flow data is available.

Peters-Lidard et al. (2017) argue that the fourth paradigm of hydrology focuses on advancing hydrologic science by full exploitation of emerging large datasets and improvement of data support for mechanisms of Earth system processes. With the availability of LiDAR data, the metrics and the hydrograph parameters can be estimated with greater accuracy, however, at the expense of the computational

burden stemming from the voluminous high-resolution data. Therefore, one approach to the maximal use of LiDAR data would involve efficient analysis and storage of watershed characteristics that can be readily used for hydrograph and peak flow estimation, as well as assessment of uncertainty.

Of particular interest is the width function, defined as the distribution of the number of stream links at a given distance from the outlet of the watershed (Bras, 1990; Rigon et al., 2016). In this context, stream links represent the channel network topology, as determined from the analysis of LiDAR data. The width function shows a diversity of functional forms within the longitudinal distance from the outlet of the longest stream segment. As such, the function does not lend itself to estimation; as a result, the watershed width function would need to be computed as an approximated frequency function with appropriate binning. Not only is such an approach cumbersome and subjective, but the use of width function within the flow estimation approaches also involves integration, and once again the lack of a function estimate renders such computations to approximations using summation approaches. Furthermore, the absence of a closed-form width function representation precludes the analytical derivation of the watershed GIUH function.

In this study, we redress these issues by (a) developing a mixed skew-normal (mSN) distribution approach to estimating width functions, with three to twelve parameters, (b) selection of model based on the Bayesian Information Criterion (BIC), (c) identifying hydrologically similar watersheds based on divergence measures applied to the mSN characterization of width functions, and (d) illustrating the utility of efficient statistical estimation of geological functions and metrics that can be scaled to very-large terrain data for Earth science problem-solving.

In what follows, first the background information regarding LiDAR, GIUH approaches, and the study region is presented. The functional estimation approach

involving skew-normal distribution is presented next. This is followed by the application and testing of the approach for the study region. Finally, the computational statistics related to the mSN approach and the prospects for efficient GIUH-based estimation at multiple scales are discussed.

3.3 Background

3.3.1 LiDAR-based Terrain Data

3.3.1.1 Characteristics

LiDAR is an active remote sensing method that uses light to measure the distance to the ground through the principle of radar (Dong and Chen, 2017; NOAA, 2020). The distance information is then converted to elevation to generate precise, three-dimensional information about the surface characteristics of the Earth. The major components of a LiDAR system comprise of a laser source and laser detector, a timing device to record the precise time for the laser to leave and return to the scanner, a Global Positioning System (GPS) to record precise location, an Inertial Measurement Unit (IMU) to measure the altitude, an on-board computer, and data storage devices. Based on functionality, LiDAR is classified as topographic or bathymetric. Topographic LiDAR typically uses a near-infrared laser to map land surfaces. Bathymetric LiDAR uses the laser to map land and water surface, while an additional water-penetrating green light can measure the bed elevations of water bodies.

Despite its many advantages, LiDAR is a technology that collects very large datasets that require careful analysis and interpretation. For this reason, the processing and analysis can be time- and computationally intensive. Moreover, LiDAR systems are prone to errors related to atmospheric and ground conditions (Xhardé et al., 2006; Karp and Stotts, 2012). LiDAR pulses may be affected by heavy rains or low-hanging clouds because of the effects of refraction. However, the

data collected can still be used for analysis. LiDAR technology does not work well in areas or situations with high sun angles or significant reflections since the laser pulses depend on the principle of reflection. When it is used on water surfaces or where the surface is not uniform, it may not return accurate data since high water depth will affect the reflection of the pulses. In some instances where the canopies over the forests are dense, the LiDAR pulses may not be able to penetrate the canopies, thus returning incomplete data. LiDAR pulses may not penetrate thick vegetation when collecting data, which can lead to inaccuracy.

3.3.1.2 Use Case: Upper Sheepscot River Basin, Maine, USA

The case study was conducted in the upper Sheepscot River Watershed Area in Maine, USA (Figure 3.1). This is a 386 km² area that drains along the Sheepscot River, with the basin outlet coinciding with the US Geological Survey stream

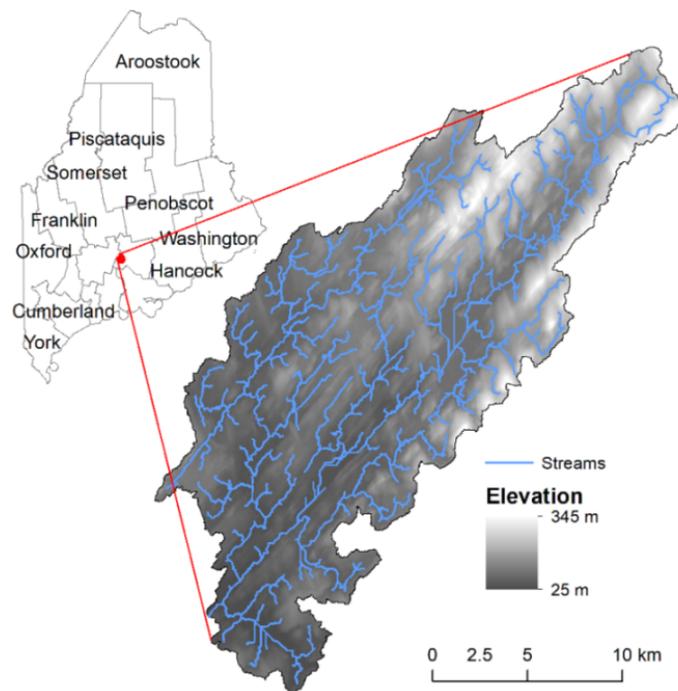


Figure 3.1: Regional map showing the upper Sheepscot River Basin, Maine. The drainage area is 386 km². The elevation ranges from 29 meters to 341 meters above the NGVD. The outlet of the basin considered here is coincident with the US Geological Survey stream gauging station number 01038000, located in North Whitefield, Maine.

gauging station number 0103800, located in North Whitefield, Maine. The mainstem of the Sheepscot River spans 93 km. The headwaters of the river are in the hills of West Montville, widening into the Sheepscot Pond in Palmero, and then dropping over the Head Tide dam to mix with the incoming tide. This area is vitally important to the remnant population of Atlantic salmon and other native species living in the river and surrounding watershed. The landscape was primarily shaped by glaciations during the Late Pleistocene, followed by the retreating of ice sheets 13,000 to 11,000 years ago, leaving behind wide, shallow valleys (Smith and Hunter, 1989), and subsequent inundation with seawater due to isostatic depression (Belknap et al., 2002; Hooke and Winski, 2014). The lower Sheepscot Watershed area was not included in the study. The Head Tide dam is located 14.5 km downstream of the study area.

The LiDAR-based Digital Elevation Model (DEM) was obtained from Maine Office of GIS (State of Maine, 2018). The horizontal and vertical resolution of the dataset are 2 m and 0.15 m respectively. This data was collected from various LiDAR collectors in Maine between 2006 to 2012.

The DEM data was used to delineate the channel network using standard Geographical Information Systems (GIS) procedures. A total of 53 outlets were selected within this watershed, and their sub-basins were delineated. Starting from the highest order streams, outlet points were placed at each location where large order streams branched out. This process was then repeated on the branches. The outlet points are shown in Figure 3.2 (a). Figure 3.2 (b) shows a strong linear relationship between the basin areas and the maximum hydrological distances, defined here as the flow path distance to the outlets, of corresponding basins.

3.3.2 Geomorphological characteristics based on stream networks

The width function is an important basin descriptor that integrates the channel network structure to provide hydrograph information, including the shape and the peak (Kirkby, 1976; Gupta and Waymire, 1983; Troutman and Karlinger, 1984, 1989). In this study, the width function, $N(x)$ is defined as the areal extent between x and $x + dx$, where x represents the flow path distance from the outlet (Veneziano

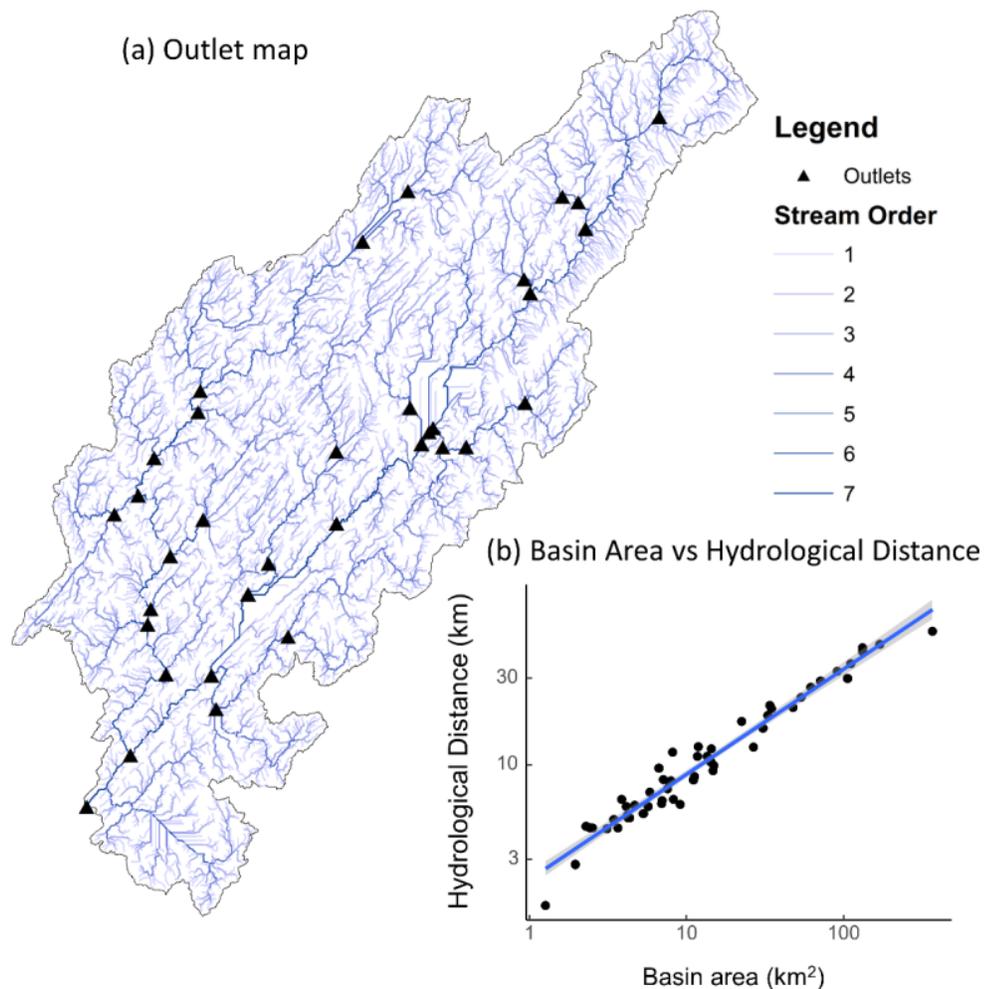


Figure 3.2: (a) The delineated channel network for upper Sheepscot River Basin based on LiDAR data. The area threshold used for this analysis was 1000 m^2 . A total of 53 outlet points were manually selected to delineate the streams at confluences of relatively large order streams to create sub-basins for the study. (b) Plot showing relationship between basin areas and maximum hydrological distances of corresponding basins.

et al., 2000). Under idealized conditions of the constant flow velocity of water without any losses, the width function represents the unit hydrograph for a basin, and reflects the topological features of the basin's stream response (Moussa, 2008). The flow distance from the outlet has been termed the hydrological distance, which has two components, the distance along a hillslope, and the distance along a channel network. In this study, we make no distinction between the two, and as such, assume that the flow remains constant. Consequently, the width function resembles an area function. We selected an area threshold of 1000 m^2 for delineating streams. Hence,

$$Z = \int_0^{x_{max}} N(x)dx \quad (3.1)$$

where x_{max} is the longest hydrological distance of the catchment ($0 \leq x \leq x_{max}$), and Z is the total catchment area.

Figure 3.3 demonstrates the width function for an example drainage network. The x -axis of the width function denotes the hydrological distance, while the y -axis denotes the frequency or density of the areal extent of the stream. It should be noted that the width function is unit-independent, i.e. the choice of units for measuring the flow path distance conveniently does not affect the shape of the width function. The distance can be scaled from 0 to 1 without affecting the width function shape. This allows for efficient integration of the width function within analytical approaches.

One analytical approach for modeling and transformation of rainfall into surface runoff is the GIUH. This approach uses the geomorphological structure of the river basin to determine the hydrologic response as a convolution of travel time distribution. The need for solely the spatial network information allows this model to assist in the prediction of surface runoff at the outlet point of an ungauged basin.

Gupta and Waymire (1983) and Mesa and Mifflin (1986) derived a generic formulation for GIUH as a function of width function.

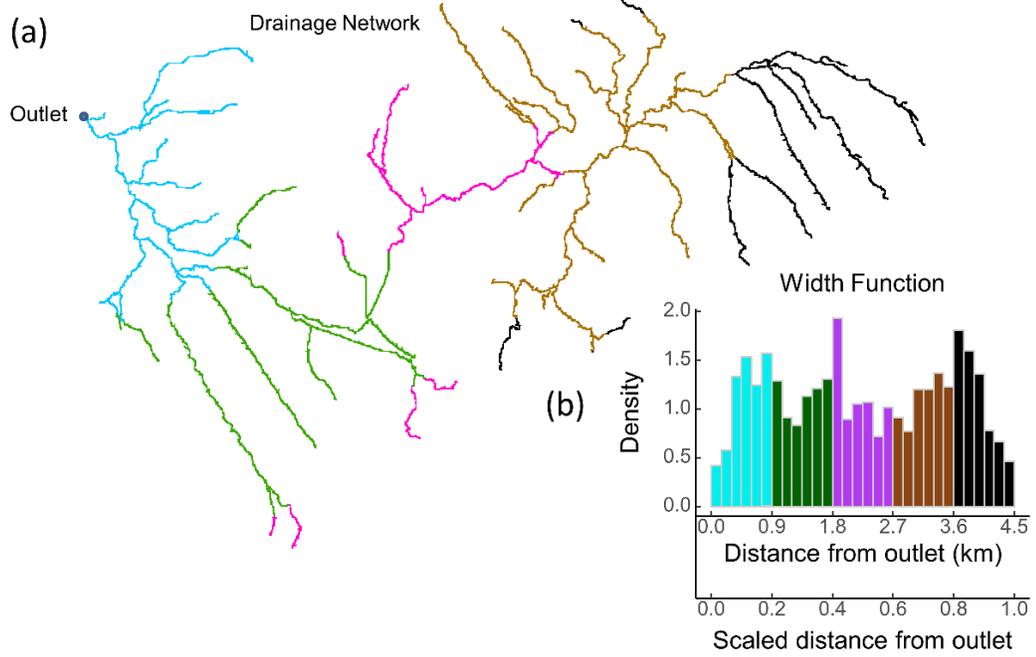


Figure 3.3: Channel network and width function for a subwatershed in the upper Sheepscot River Basin. (a) Drainage network, with color gradations across flow path distance from the outlet denoting hydrological distances, and (b) width function with corresponding color gradation. The upper x-axis presents the hydrological distance, while the lower x-axis presents the corresponding scaled hydrological distance.

$$h(t) = \int_0^{x_{max}} g(x, t) N(x) Z^{-1} dx \quad (3.2)$$

where $g(x, t)$ is the hydraulic response function of a single channel at a distance x from the outlet of the basin. The term $N(x)Z^{-1}$ represents a non-dimensionalized width function with the hydrological distance, x scaled between 0 and 1. From this point, we use this scaled hydrological distance for width function representation. As such, we define the normalized width function, $N^*(x^*)$ as

$$N^* = \frac{N}{Z}; \int_0^1 N^*(x^*) dx^* = 1 \quad (3.3)$$

and x^* as

$$x^* = \frac{x}{x_{max}}; 0 \leq x \leq x_{max} \quad (3.4)$$

Furthermore, under idealized conditions of no groundwater exchanges and losses, $g(x, t)$ becomes a constant, g_0 and the GIUH of the basin becomes proportional to the normalized width function.

$$h(t) = g_0 \int_0^1 N^*(x^*) dx^* \quad (3.5)$$

While the normalized width function was used in this study, it should be noted that the non-scaled width function can be recovered using Z and x_{max} . This prevents the loss of the basin size information.

3.4 Estimation of Width Functions

The form of width function can vary widely depending on the shape of the watershed, its terrain, and the channel network characteristics. Width functions can have a single peak or multiple peaks. The shapes can be symmetric, positively skewed, or negatively skewed. Thus, careful consideration of various functional forms, ones that can accommodate the variations in branching structure observed in channel networks, is needed.

In this study, we investigate the various types of width functions of the 53 sub-basins in the Upper Sheepscot watershed. The width functions of five selected sub-basins in the study area are shown in Figure 3.4. These five basins illustrate diversity in the shape of width functions. Basins 3 and 8 have unimodally distributed width functions (Figure 3.4 (a, b)). While the former is close to symmetric, the latter is highly skewed. Similarly, Basins 20 and 23 have bimodally distributed width functions, with the former being symmetric, and the latter being skewed (Figure 3.4 (c, d)). The width function for Basin 24 is multimodally distributed (Figure 3.4 (e)). The functional form that properly represents the width function should thus capture the skewness and the modality. In this section, we

investigate select probability distributions and identify one that can adapt to various distributional shapes seen in our study cases.

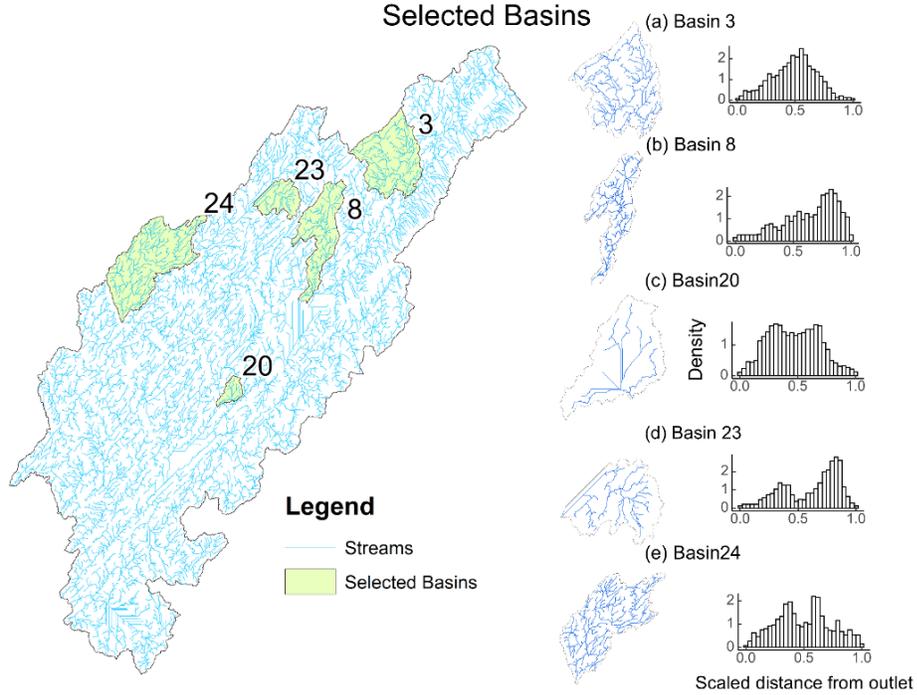


Figure 3.4: Sample width functions for selected sub-basins. These width functions exemplify the diversity in shapes, and the accompanying modality and asymmetry.

3.4.1 Approach

3.4.1.1 Probability Distributions

We first examined a select set of bounded probability distributions for the functional representation of width functions. Since the scaled distances from the outlet were bounded ($x \in [0, 1]$), Beta and Kumaraswamy distributions (Jones, 2009) were first tested.

A continuous random variable X has a Beta distribution, $\beta(x; a, b)$ if its probability density function (*pdf*) has the following form:

$$\beta(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)}, x \in [0, 1], a \in (0, \infty), b \in (0, \infty) \quad (3.6)$$

where a and b are the parameters, and $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Similarly, a continuous random variable X has a Kumaraswamy distribution, $K(x; a, b)$, if its *pdf* has the form (Jones, 2009):

$$K(x) = abx^{a-1}(1 - x^a)^{b-1}, x \in [0, 1], a \in (0, \infty), b \in (0, \infty) \quad (3.7)$$

where a and b are the parameters.

Lastly, a continuous random variable X has a univariate skew-normal (SN) distribution if it has the *pdf* of the following form (Azzalini, 1985):

$$f(x) = 2\phi(x)\Phi(\alpha x) \quad (3.8)$$

where α is a shape parameter, $\phi(x)$ denotes the standard normal density function of x , and $\Phi(x)$ denotes the cumulative distribution function (*cdf*) of the standard normal (Equation S1, S2). Further, this can be linearly transformed as $Y = \xi + \omega X$ to form the 3-parameter skew-normal distribution, $SN(\xi, \omega^2, \alpha)$, where ξ is the location parameter, and ω is the scale parameter. Consequently, the distribution of the three-parameter SN distribution is of the form:

$$f(x) = \frac{2}{\omega} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left(\alpha \frac{x - \xi}{\omega}\right), x \in (-\infty, \infty) \quad (3.9)$$

The SN distribution is an extension of the normal (Gaussian) probability distribution, allowing for the presence of skewness. An SN distribution becomes normal when the shape parameter is equal to zero. As the shape parameter tends to infinity, the distribution converged to the folded normal distribution.

Since the domain of SN distribution is $(-\infty, \infty)$, we truncate the distribution to restrict the domain to $[0, 1]$. Suppose X is a continuous random variable with *pdf*, $f(x)$ and *cdf*, $F(x)$, and with a domain of $(-\infty, \infty)$. This distribution can be truncated to a domain of $[a, b]$, given $a < b$ by multiplying $f(x)$ by a correcting factor, which guarantees the validity of the normalization condition (Thomopoulos, 2017):

$$g(x; a, b) = \begin{cases} \frac{f(x)}{F(b)-F(a)} & , x \in [a, b] \\ 0 & , x \in (-\infty, a) \cup (b, \infty) \end{cases} \quad (3.10)$$

In the context of normalized width function (Figure 3.3 and Equations 3.3–3.4) modeled as a skew-normal distribution (Equation 3.9), $a = 0$ and $b = 1$. For a continuous random variable, $X \sim SN(\xi, \omega^2, \alpha)$, with a *pdf* of $f(x)$, a *cdf* of $F(x)$, and a domain of $(-\infty, \infty)$, if $g(x; 0, 1)$ denotes the *pdf* of the corresponding distribution truncated to a domain of $[0, 1]$, the functional form of $g(x; 0, 1)$ is:

$$g(x; 0, 1) = \frac{f(x)}{c(0, 1)} \quad (3.11)$$

where, $c(0, 1) = F(1) - F(0) = \int_0^1 f(x)dx$.

3.4.1.2 Finite Mixture Models

As noted in previous sections, width functions are often multimodal. Most statistical formulations deal with unimodal data. Although multimodal data can be modeled with some single distributions, the functional forms lack the diversity of forms seen in empirical estimates of width functions in watersheds (for example, as shown in Figure 3.4). Multimodal distributions can result from sub-populations. If they cannot be identified, a mixture distribution can be used to analyze the dataset (McLachlan and Peel, 2004). The *pdf* of a finite mixture distribution of n -dimensional random vector, X takes the form (McLachlan et al., 2019):

$$f(x) = \sum_{i=1}^n w_i g_i(x) \quad (3.12)$$

where, the mixing proportions, w_i are non-negative and sum to one, and where the $g_i(x)$ are the component densities.

Now for a finite mixture of n truncated skew-normal distributions, the *pdf* thus becomes:

$$f(x) = \sum_{i=1}^n w_i g(x; \xi_i, \omega_i^2, \alpha_i) \quad (3.13)$$

The skew-normal parameters, ξ_i , ω_i , and α_i , and the mixing proportions, w_i can be estimated through maximum likelihood estimation (MLE) (Casella and Berger, 2002; Rossi, 2018).

Finite mixture model selection entails a choice of the number of components. In this regard, two considerations are deemed relevant: (a) a penalized form of the log-likelihood function, and (b) a likelihood ratio test (McLachlan et al., 2019). We demonstrate the model selection process using BIC. BIC is defined as

$$BIC = -2 \ln(L) + k \ln(n) \quad (3.14)$$

where L is the likelihood, k is the number of parameters, and n is the sample size. The model with the lowest BIC is preferred. Keribin (2000) showed that BIC is reliable in choosing the true number of components in a mixture model. BIC has been applied extensively in hydrologic analysis and modeling applications (Tsai, 2010; Haddad and Rahman, 2011).

3.4.1.3 Divergence Measures

In this section, we examine one potential application of developing a functional form for representing width functions. Making hydrologic predictions is important in various fields of engineering and water resources. There is an increasing need for the ability to predict the future hydrologic response in any given area. But the lack of data everywhere places a limit on this. For instance, without the presence of stream gauge data, it is difficult to predict the flood response to a storm event. This has led to the study of hydrologic similarity being a viable option to predict the hydrologic response in an ungauged basin by using the information available from a hydrologically similar gauged basin on the basis of physical proximity or similar catchment properties (Burn and Boorman, 1993; Tung et al., 1997; McIntyre et al., 2005; Wagener et al., 2007; Reichl et al., 2009; Archfield and Vogel, 2010; Oudin et al., 2010; Patil and Stieglitz, 2011, 2012; Razavi and Coulibaly, 2013; Athira

et al., 2016; Brunner et al., 2018). We propose that GIUH, or more specifically for this study, the width function can serve as a measure of hydrologic similarity, given that it represents the hydrologic response of a catchment under idealized conditions.

Divergence measure or contrast function is a function that assesses the "distance" (Mahalanobis, 1936; Bhattacharyya, 1943; Kailath, 1967) between two probability distributions. For two probability distributions, p and q with common support, the divergence function is denoted as $D(p||q)$. In order to compare width functions as a proxy for determining similarity between drainage basin characteristics, we investigated various divergence metrics (Tsybakov, 2008), which have been summarized in Table 3.1. Due to the complexity of comparing the moments of a finite mixture of probability distributions, we opted for numerical analysis to compute the divergence measures, with discretized $\Delta x = 0.01$.

3.4.2 Application

3.4.2.1 Comparison of Distributions

We first visually compare the goodness of fit of the three distributions (Figure 3.5). Beta and Kumaraswamy distributions provide similar fits. These distributions have the advantage of not having to truncate the domain to $[0, 1]$. However, the truncated SN distribution better captures the peak of the width function, which is an important value in hydrologic analysis. The concave shape of the tail is also

Table 3.1: Four divergence metrics to assess the similarity between probability density functions. $N_1 = N_1(x)$ and $N_2 = N_2(x)$ indicate the two probability distributions being compared, i.e. the width functions of sub-basins. $D(N_1||N_1)$ indicates the value of the metric when the two distributions, i.e. width functions, are identical.

Metric	Definition, $D(N_1 N_2)$	$D(N_1 N_1)$
L_1 Distance (L_1)	$\int N_1 - N_2 dx$	0
L_2 Distance (L_2)	$\sqrt{\int (N_1 - N_2)^2 dx}$	0
Hellinger Distance (HD)	$\sqrt{\int (\sqrt{N_1} - \sqrt{N_2})^2 dx}$	0
Hellinger Affinity (HA)	$\int \sqrt{N_1 N_2} dx$	1

better captured by the truncated SN distribution. For Basin 8, the Beta and Kumaraswamy distribution curves fail to capture the shape of the right tail of the histogram. Skew-normal provides a better fit. For these reasons, despite the need to truncate the SN distribution, the better fit suggests that SN distribution should be preferred to model the width function.

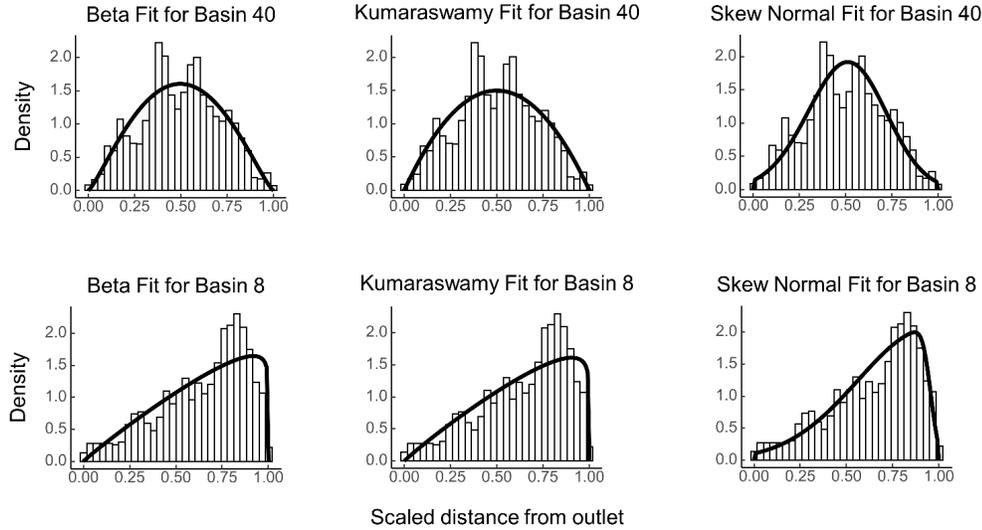


Figure 3.5: Comparison of unimodal Beta, Kumaraswamy, and truncated skew-normal distribution fits. $\beta(2.25, 2.27)$, $K(2.01, 2.06)$, and $SN(0.96, 0.15, -11.00)$ are the fitted distributions for Basin 40. For Basin 8, the fitted distributions are $\beta(1.96, 1.09)$, $K(1.95, 1.15)$, and $SN(0.63, 0.06, -0.91)$.

3.4.2.2 Examination of Finite Mixtures

Next, the effects of finite mixtures are examined. While increasing the number of mixture components provides a better fit, each increase comes at the cost of four added parameters. Each individual SN component captures a separate local peak of the width function as seen in Figure 3.6. Increasing the number of components more than that is necessary can lead to the capture of extremely small peaks. On the other hand, a lower number of SN components can lead to the distribution being unable to capture the shape of the width function overall. For the 53 selected

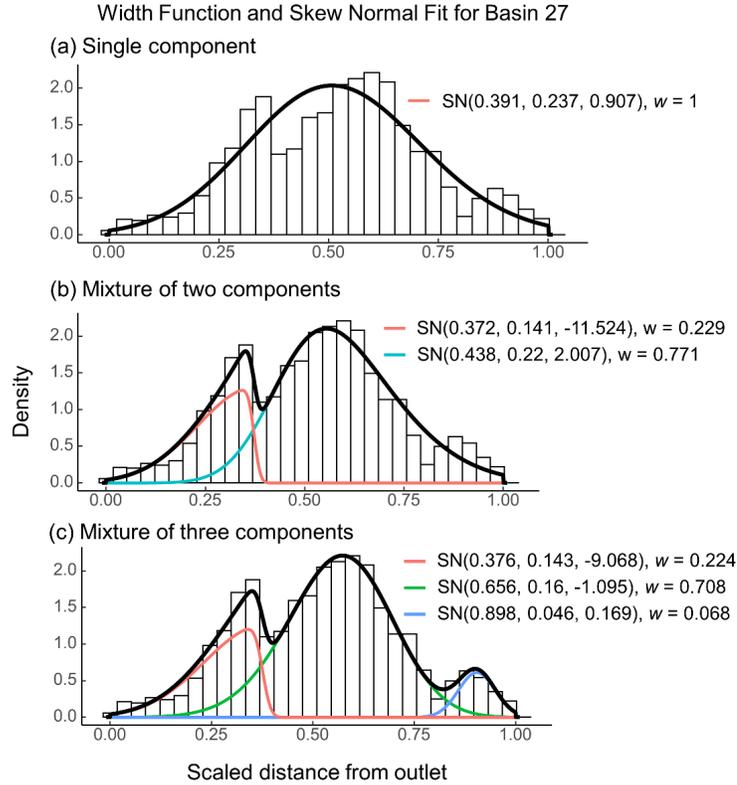


Figure 3.6: Finite mixture of truncated skew-normal distributions. Linear mixtures of (a) one, (b) two, and (c) three SN components for fitting the given width functions are demonstrated.

sub-basins, we find that a maximum of three components performs adequately for 46 cases.

Single SN component suffices for Basin 3 (Figure 3.7 (a)). However, mixtures of two components are required for Basins 8, 20, and 23 (Figure 3.7 (b), (c), and (d)). Moreover, for Basin 24, a mixture of three components is required (Figure 3.7 (e)). For Basin 3, the mixture model fit does not offer an appreciable improvement over a single-component SN fit. Similarly, for Basin 8, 23, and 20, the increase in the number of components above two groups does not contribute to a substantial change in the overall distribution.

Similar results are seen in cumulative distribution plots in Figure 3.8 (b), (c), and (d). As expected, increasing the number of components increases the goodness

of fit. But visual inspection of the cumulative distributions seems to indicate that the improvements level off at the two-component mixtures. Examining the change in BIC with an increase in the number of components provide mostly similar results (Figure 3.8 (a)). For Basin 20 and 23, the lowest BIC occurs for two-component mixtures, indicating that this is the most appropriate model to use. It can be noted that for Basin 23, the BIC values for two- and three-component mixtures are similar, but since a lower number of parameters is preferred, the former is the best model. For Basin 8, the BIC values indicate that the three-component model is the best. Additionally, Figure A.1 shows the distribution of residuals as a means of comparing the quality of fit. For choosing the most optimal number of components, we want the average of the residuals to be close to zero and the variance to be low. Figure A.1 indicates that increasing the number of components leads to an increase

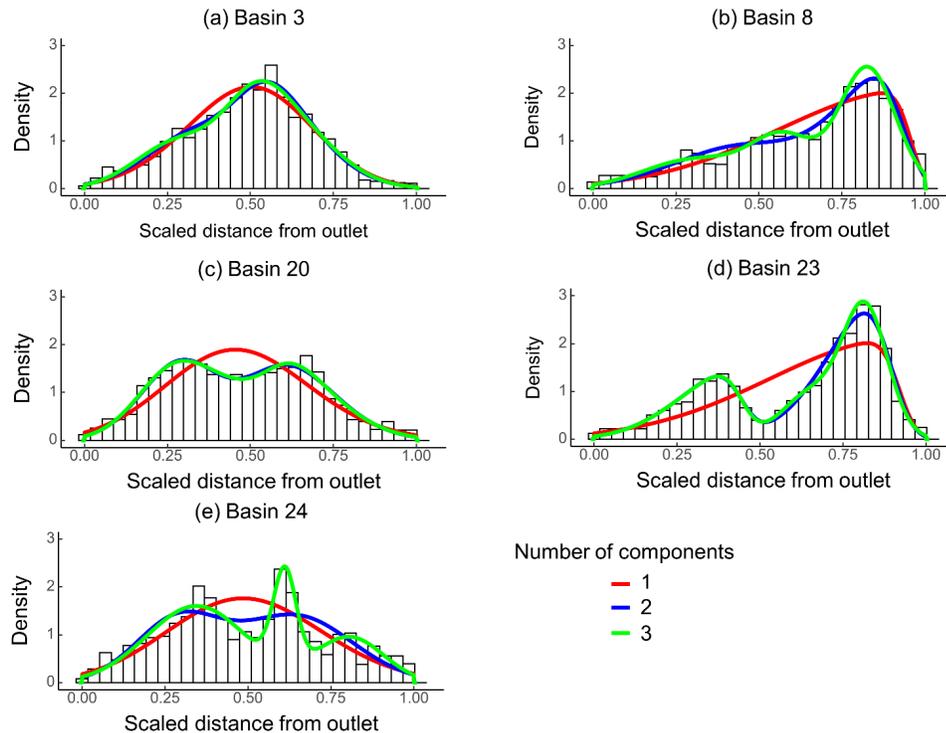


Figure 3.7: Finite mixture of truncated skew-normal fits for selected width functions. A mixture of one to three SN components was tested for unimodal shapes ((a) Basin 3 and (b) Basin 8), bimodal shapes ((c) Basin 20 and (d) Basin 23), and trimodal shapes ((e) Basin 24).

in the quality of fit. But the improvement from additional components is seen to tail off at some point.

While the lowest number of components is preferred to limit the number of parameters, an advantage of using mixtures is that they can be adjusted to address the degree of complexity required. We note that width functions computed based on DEMs with different resolutions may differ in shape, as finer-scale features may

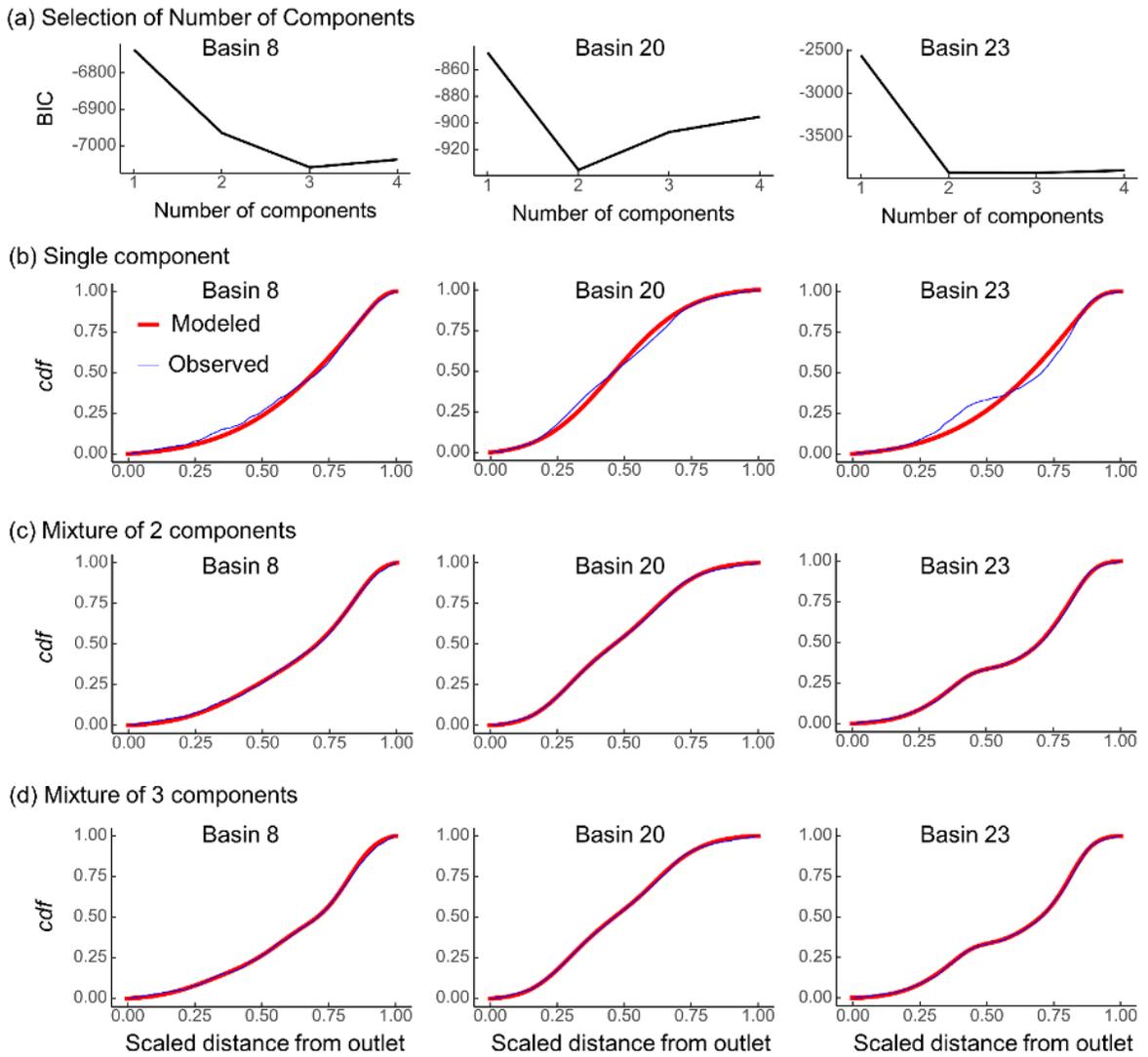


Figure 3.8: Model selection for width functions based on the truncated mSN distribution. (a) Comparison of BIC for different number of components, and (b, c, d) Empirical $cdfs$ overlaid with $cdfs$ based on fitted mSN distributions with one, two, and three components.

appear; in all those instances, the choice of mixtures would afford appropriate fits, likely with differing numbers of components. Figure 3.9 shows an example where a mixture of five components provides a substantially better fit based on the BIC values.

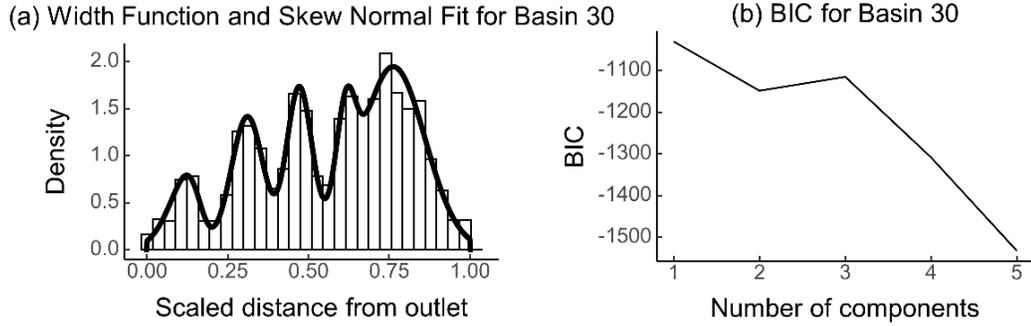


Figure 3.9: (a) The width function with a five-component mSN fit, and (b) the BIC plots for the best model selection for Basin 30.

3.4.2.3 Measure of Width Function Similarity

In this section, we investigate the applicability of width functions as a measure of hydrologic similarity. We begin by comparing the four divergence metrics shown in Table 3.1. All four metrics are comparable in terms of rank, but the distribution of metric values varies (Figure 3.10). HA has a left-skewed distribution while the rest are symmetric to varying degrees. L_2 has the most even and symmetric distribution between a convenient range of 0 to 1. As such, we chose L_2 as the divergence metric for determining hydrologic similarity in this study. Moreover, the wide variance in the metrics indicates heterogeneity in the width function shapes.

Each basin was compared with other basins whose maximum hydrological distances were within 20% and which had less than a 40% overlap. We note that the hydrological distance is strongly correlated with drainage area (Figure 3.2 (b)), and as such, restricts the comparisons to watersheds with similar drainage areas.

Furthermore, the choice of the allowance of variance in basin sizes (here, 20%) may be revised to a lower value, especially in cases where a large number of candidate

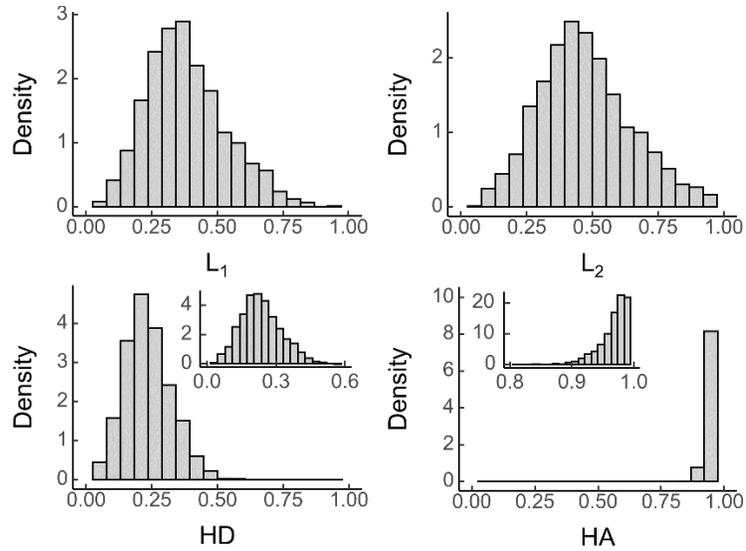


Figure 3.10: Distribution of (a) L_1 distances, (b) L_2 distances, (c) Hellinger distances (HD), and (d) Hellinger affinities (HA) (see Table 4.1 for details) for the 53 sub-basins. A variety in the range and skewness of these indices can be observed.

basins are available. The divergence metric provides reliable estimates of the similarity in width functions as compared to visual inspection (Figure 3.11). Basins 16 and 38 have the most similar width function distributions visually and according to the metric. Interestingly, the metric does not place strong importance on the modality of the distribution. Basins 16 and 21 have a better metric score than Basins 38 and 43, despite the former having different modalities and the latter having the same. Figure A.2 provides the comparison between Basin 16 and all other basins that meet the aforementioned criteria, and illustrates an example of using this method to find a set of similar basins that can be used to derive hydrologic information for a given ungauged basin. This would include specifying the range of drainage areas or maximum hydrological distances to select basins for further analysis. In spite of the fact that similarity in basin shape and drainage density does not seem to be reflected in this similarity study, the purpose of using the width function as a parameter for similarity is to compare the hydrologic

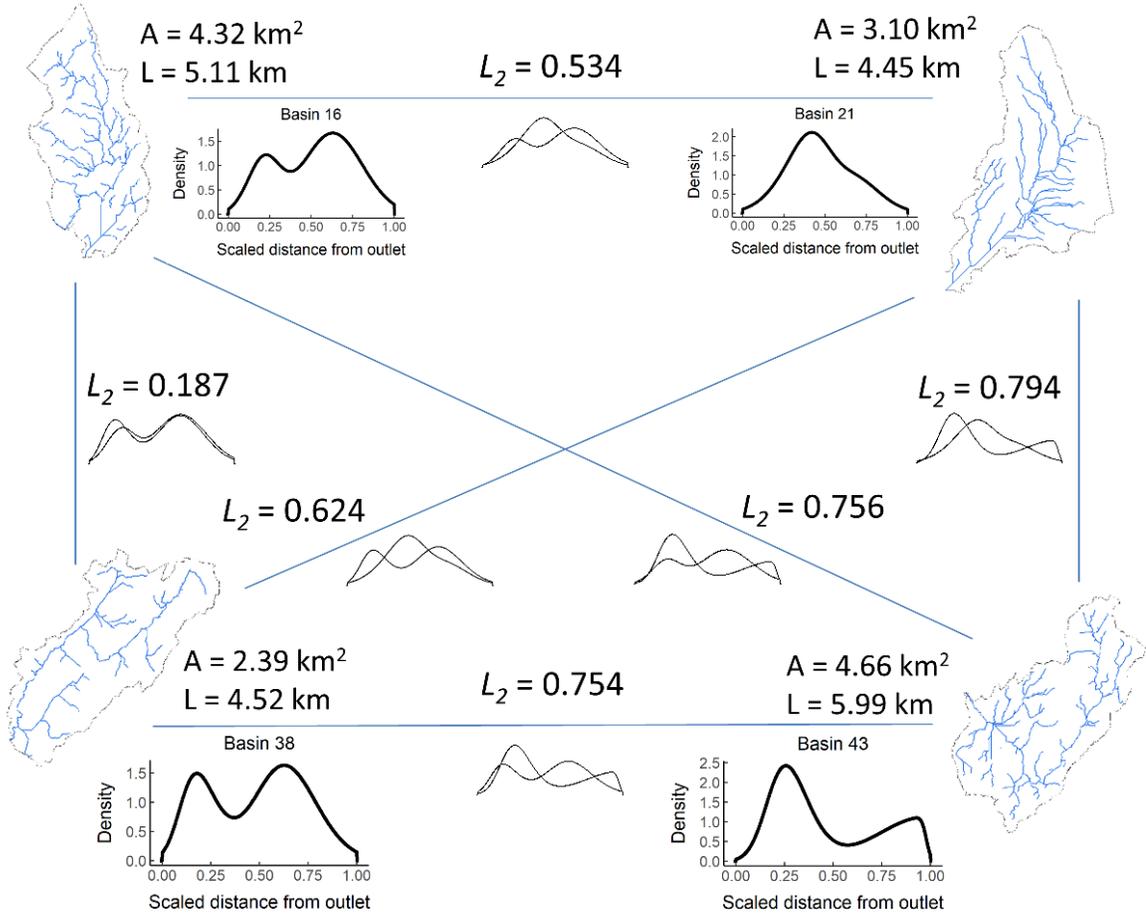


Figure 3.11: Cross-comparison of four basins, based on width functions. Divergence measure, L_2 , was used as an indicator for the similarity between basins, with a lower L_2 distance indicating a higher degree of similarity (see Table 4.1 for details).

response of terrain. As such, the width function affords similarity assessment through a comparison of dynamic response characteristics of watersheds.

3.4.2.4 Computation Time

In this section, the computation time for parameters estimation for the mixture of truncated SN distribution using MLE is explored. The parameter estimation was done in R programming language (R Core Team, 2019) using the "mixsmsn" package (Prates et al., 2013). The script was run on the University of Maine's Advanced Computing Cluster (3.4 GHz, 12 cores, 256 GB RAM). The results are shown in Figure 3.12. The order of processing time ranged from 10^{-1} seconds to 10^3

seconds. There is a strong correlation between the number of elevation points for modeling and the processing time, with the processing time increasing exponentially with the increase in the number of points. The processing time generally increased with the increase in the number of components used in a mixture. However, occasionally an increase in the number of components does lead to a decrease in processing time. This could, however, simply be due to a choice of better initial values.

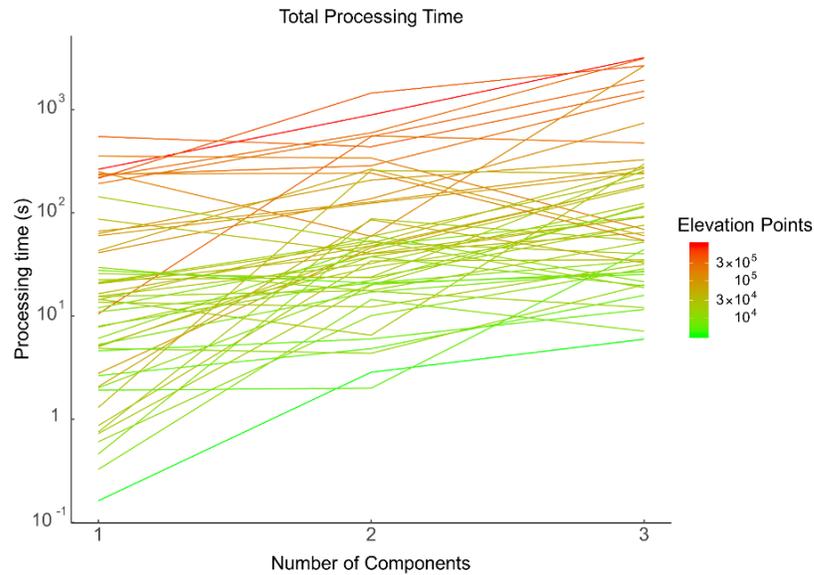


Figure 3.12: Total processing time for parameter estimation. As expected, parameter estimation time generally increases with an increase in the number of elevation points and number of mixture components.

3.5 Discussion

Estimation of hydrologic response in the presence of limited data poses numerous challenges and is a field of ongoing research (Sivapalan et al., 2003). Lumped models that utilize gross estimates such as basin area, average slope, location of the watershed centroid, and catchment indices, to name a few, are the most commonly used measures (Tegegne et al., 2017; Pérez-Sánchez et al., 2019). A widely available data for this study is elevation information, including

high-resolution terrain information using LiDAR (Schwarz, 2010; Ussyshkin and Theriault, 2011). GIUH is a physically-based response characteristic of the watershed that utilizes the drainage network pattern and can be used to estimate the hydrologic response in an ungauged basin (Kumar et al., 2002; Ellouze-Gargouri and Bargaoui, 2012; Sahoo and Saritha, 2015; Swain et al., 2015). Under idealized conditions, the width function represents the catchment’s GIUH (Moussa, 2008). Using 53 sub-basins from the Upper Sheepscot watershed, this study demonstrated a systematic and robust analytical approach to modeling a normalized width function using a finite mixture of truncated skew-normal distributions.

Peters-Lidard et al. (2017) argue that the next paradigm of hydrology focuses on data-intensive science in the presence of large datasets and advances in parameter estimation and modeling. We harnessed the high-resolution LiDAR-based terrain dataset to further the understanding and applicability of the width function in modeling the GIUH. As such, we showed that the mSN distribution can reliably represent the GIUH in its graphical form. This can substantially improve the mathematical formulation of GIUH and allow us to better model hydrologic response in data-scarce environments. We further examined the trade-offs of increasing the number of components of the mixture distribution. As the number of parameters increases, the resulting distribution can more accurately capture the width function shape. While this comes at the added expense of reduced tractability, it can still be useful in representing certain width functions with higher modality in their shapes. We also demonstrated the use of BIC for model selection.

The functional approach provides an efficient measure for storing width function information in the form of three to twelve parameters in addition to the basin size information, as opposed to a large amount of elevation data that would be required for recreating the original width function. This is especially important when using high-resolution data, with the average number of elevation points being 82,000 for

the sub-basins in this study. We also examined the computation time in estimating the model parameters and showed that with the current advancement in supercomputing technology, the computation time does not appear to be a limiting factor. Once these parameters are estimated, the information regarding the hydrologic response provided by high-resolution DEMs with large file sizes can be efficiently stored as a set of parameter values that can be easily used to reproduce the width function for further analysis. Due to the robustness of this approach, it can be applied to large regions, including datasets on a national or global scale.

As an example of the use of the mathematical formulation of width function, we illustrated an application of the formulation as a metric for the hydrologic similarity between drainage basins. The usual methods of assessing hydrologic similarity include the transfer of limited hydrologic information such as hydrologic indices, streamflow values, etc. from ungauged to gauged catchments (Burn and Boorman, 1993; McIntyre et al., 2005). These lumped data carry limited information about the hydrologic response of a catchment and can be subject to large uncertainty (Wooldridge and Kalma, 2001; Wagener and Wheeler, 2006; Tetzlaff et al., 2009; Chang et al., 2014). A reliable metric for the selection of gauged catchments for the transfer of hydrologic information is the spatial proximity to the ungauged catchment. However, this measure prevents the identification of hydrologic similarity among catchments that might be located far from each other (Patil and Stieglitz, 2012). Nonetheless, in the absence of a universal basis for catchment classification, the use of catchment characteristics such as runoff coefficient and baseflow index, hydrogeomorphological dimensionless numbers such as bifurcation ratio and drainage density, hydro-climatic region, soil properties, etc. remain common measures for assessing hydrologic similarity (Wagener et al., 2007). We propose the width function as another potential metric for similarity assessment as a unique extension of the GIUH approach. This allows for assessing the similarity

between basins using a more dynamic measure, i.e. the travel time distribution at the outlet. This method uses a distribution curve comparison-based approach for the evaluation of similarity in hydrologic response rather than a single-valued index.

3.6 Concluding Remarks

The width function represents the GIUH under idealized conditions that the surface flow velocity remains constant and there are no losses. As such, this study provides a framework for the functional formulation of the width function, i.e. the GIUH under ideal conditions, in its current graphic representation, allowing for more robust applications in studying the watershed's response characteristics. The primary results are summarized as follows:

1. This study provides a robust framework for modeling the key component of GIUH, i.e. the width function using a finite mixture of truncated skew-normal distribution. Our findings indicate that the *SN* distribution more reliably captures the key characteristics of the width function shape compared to Beta and Kumaraswamy distributions.
2. The finite mixture-based approach can be used to model the multimodality of the width function resulting from sub-populations. The approach provides versatility in capturing the desired level of detail through the choice of the number of mixture components. Furthermore, we showed an approach for the selection of the number of components using BIC and based on a graphical study of goodness-of-fit.
3. This study also illustrates a potential application of the functional estimation of the width function in assessing the similarity in hydrologic response between watersheds. By using divergence metrics, we provide a framework for comparing the shape of width functions formulations.

4. The functional approach provides an efficient way to capture the response characteristics from large terrain datasets such as the LiDAR-based high-resolution digital elevation data and to efficiently store them as a set of mSN -based width function parameter estimates.

Our results pave the way for a number of fruitful research directions. First, a functional representation of the width function would allow analytical studies of the GIUH response at the watershed scale. Secondly, the estimation approach allows for analysis and modeling using high-resolution and larger-scale (national or global) datasets. The third direction is one of developing a suite of watershed metrics that represent dynamical processes as improved measures for modern computer-based learning approaches, such as artificial intelligence and machine learning (Reichstein et al., 2019).

CHAPTER 4

CHARACTERIZATION OF DRAINAGE BASIN HYPSONOMETRY: A GENERALIZED APPROACH

4.1 Chapter Abstract

The elevation-area relationship within drainage basins has a strong bearing on manifold biophysical processes. The understanding of erosional and hydrodynamical processes also benefits from the quantitative characterization of elevation distribution or basin hypsometry. This study presents a comprehensive review of quantitative approaches to hypsometric analysis, ranging from simple indices to functional formulations (for example, Strahler). Detailed analyses of the goodness-of-fit reveal that the current suite of metrics is deficient when tested over a large number of drainage basins (in this study, Congo, Narmada, and Yukon River basins), especially in the head and the toe sections of the hypsometric curve. A new three-parameter formulation is introduced in this study and tested over a large set of drainages ($n = 419$). We find that this form adapts well to diverse elevational profiles. Application areas, such as large-scale analyses of drainage basin similarity and snowmelt modeling, use hypsometric parameterization and stand to benefit from the new functional characterization discussed here. Efficient estimation of hypsometric curves based on the new formulation also paves the way toward computational research at whole-earth scales utilizing large datasets involving millions of drainages.

4.2 Introduction

The elevational profile of a drainage basin represents the dynamic evolution of the landscape resulting from tectonic processes and erosion driven primarily by

climate, and at the same time serves as a backdrop for biophysical processes such as river runoff, floral and faunal diversity, and microclimates (Gallen et al., 2011; Dudov, 2017; Freeman et al., 2018). The elevation-area relationship is often summarized using histograms (frequency distributions) and functional representations (Strahler, 1952) based on the analysis of gridded digital elevation model (DEM) data. The use of mathematical functions allows efficient reduction of the gridded data to a small set of parameters. To this end, a hypsometric curve (Horton, 1932) quantifies the cumulative area above various elevation levels, and provides graphical and functional representation of the area-altitude distribution within a drainage basin. In general, quantitative measures of elevation have been applied to diverse problem areas: (a) hypsometric integral (HI), a measure of the relative location of mean elevation in reference to the elevational range of a drainage basin, has been linked to the erosional maturity of the landscape (Strahler, 1952; Moglen and Bras, 1995; Pedrera et al., 2009; Guha and Jain, 2020), (b) geomorphic signatures on other planets have been analyzed based on hypsometric curves (Coradini et al., 1980; Rosenblatt et al., 1994; Grant and Fortezzo, 2003a,b; Nicholson et al., 2013), (c) hypsometry has a pronounced effect upon the flood peaks and the time-to-peak in river basins (Howard, 1990) and has relevance for watershed management (Sreedevi et al., 2005), and (d) the link between basin morphometry and hydrologic response is the basis for the geomorphological instantaneous unit hydrograph approach (Rodríguez-Iturbe and Valdés, 1979; Gupta et al., 1980; Gallen et al., 2011) applied widely to understand hydrologic scaling and prediction in ungauged river basins. Thus, a general functional form of the hypsometric curve has the potential to facilitate statistical and numerical analyses applied to the above-noted contexts.

The need for quantitative representation of hypsometry has been well recognized in the literature. Langbein (1947) first used the percentage hypsometric curve in

fluvial geomorphology to analyze erosional potential in watersheds. A number of other studies on erosional topography employed the hypsometric curve (Strahler, 1952; Miller, 1953; Schumm, 1956; Coates et al., 1958; Strahler, 1957), however, this was hindered by the lack of a quantitative formulation (Harlin, 1978). Strahler (1952) developed a simplified two-parameter transcendental hypsometric function to characterize the elevation-area relationship in drainage basins, but its usage remains limited in the literature. A commonly used index in the hypsometric analysis is HI, which is somewhat limited as a diagnostic of evolutionary changes and hydrologic signatures in drainage basins (Harlin, 1978; Willgoose and Hancock, 1998). The loss of morphometric information stemming from the use of HI can be significant and result in ambiguities; for example, two drainage basins with distinct hypsometric curves may have nearly identical HI values. All of this underscores the need for a robust functional form that is pliable in diverse geological and climatic settings. Harlin (1978) developed a polynomial hypsometric function and introduced skewness and kurtosis measures to characterize various attributes of the hypsometric curve. Subsequently, various attempts have been made to devise hypsometric functions to accurately fit the hypsometric curve (Sarkar and Patel, 2011; Vanderwaal and Ssegane, 2013). The polynomial functional form remains the most widely used approach (Harlin, 1984; Luo, 2000, 2002; Vivoni et al., 2008; Huang and Niemann, 2008b; Gallen et al., 2011; Ssegane et al., 2012; Vanderwaal and Ssegane, 2013; Liffner et al., 2018). Although the polynomial representation adapts well to the curvilinear hypsometric curve, the endpoints (boundary conditions) are not constrained.

The present study seeks to provide a comprehensive review of the quantitative approaches to elevation-area relationships in drainage basins. In view of the above-noted limitations within current formulations of hypsometric functions, we present a new three-parameter model that is tested based on comparative analyses

for its generality and efficacy in modeling hypsometric curves in diverse environments, with three river basins as examples: the Congo River basin (Africa), the Narmada River basin (India), and the Yukon River basin (the USA and Canada). Finally, the new formulation is applied to a large set of sub-watersheds in these basins to assess hypsometric similarity as an index of analogous hydrologic response.

4.3 Data and Use Cases

4.3.1 Data

Case studies were conducted in three drainage basins from different climatic regions: the Congo (Africa) with an equatorial tropical climate, the Narmada (India) with a subtropical climate, and the Yukon (Canada) with primarily a subarctic climate. Elevation data were extracted from GTOPO30, a global DEM developed by the United States Geological Survey (USGS) (USGS, 1996). The spatial resolution is 30-arc seconds and the vertical accuracy is 30 m. GTOPO30 was derived from several raster and vector sources of topographic information, and was developed over a three-year period through a collaborative effort led by staff at the USGS's Center for Earth Resources Observation and Science (EROS). Despite the constraint of coarse spatial resolution, GTOPO30 represents a data product with extensive accuracy checks that have likely removed blunders and systematic errors. GTOPO30, therefore, potentially offers a means for identifying and correcting systematic errors of magnitudes greater than 1 km within satellite-derived DEMs that have had few accuracy checks (USGS, 1996).

Channel stream networks were delineated from the DEM through standard Geographical Information System (GIS) procedures using ESRI ArcGIS 10.5.1. The steps include filling the sinks in the DEM (elevation grids with undefined drainage directions), determining the direction of flow at each grid based on the direction of

steepest descent, computing the accumulated flow at each grid, and using an accumulation threshold to identify stream grid cells. The thresholds were chosen to create a dense stream network (resulting in fourth-order streams) that can be used to delineate a sufficient number of sub-basins. Having a large number of sub-basins within each watershed allows for testing the fits for diverse hypsometric curve shapes. Outlet points were placed at confluences of first-order streams and streams of higher order to ensure no sub-basin overlap. Accumulation thresholds of 854 km², 396 km², and 1618 km² were chosen for Congo, Narmada, and Yukon rivers, and a total of 221, 72, and 126 sub-basins were delineated, respectively.

4.3.2 Congo River Basin

The Congo River basin (CRB) is located in the central region of Africa (Fig. 4.1a) between latitudes 9° 13' 31" N and 13° 27' 29" S, and longitudes 12° 4' 58" E and 34° 0' 28" E. The Congo River begins at the Chambeshi River, flows through Lake Mweru, and joins the Lwalaba River (Balek, 2011). It is the second-largest river system in the world with regard to both the water discharge (with a mean annual flow of around 40,600 m³) and the drainage basin size (3.7×10^6 km²). The elevation ranges from zero at the outlet at the ocean to around 4000 m above sea level in the east. The average slope is about 0.7°. A shallow depression occupying almost half the area of the basin, named the "Cuvette Centrale", lies in the heart of the CRB (Bernard, 1945; Laraque et al., 2009; Kadima et al., 2011) has a major effect on the relief, geology, and climate of the entire drainage basin (Laraque et al., 2009). The CRB is underlain and completely surrounded by a Precambrian basement (De Wit et al., 2015). Sediment accumulation and tectonic inversions have occurred in the CRB since the Neoproterozoic era (Kadima et al., 2011). The main drivers of the hydro-climatic dynamics are the coupled ocean-atmosphere modes of the El Nino Southern Oscillation and the Indian Ocean Dipole (Becker

et al., 2018). The CRB has an equatorial tropical climate with two rainy seasons in March-April and October-November (Laraque et al., 2009), whereas the Cuvette receives rainfall throughout most of the year (Mushi et al., 2019). The CRB comprises of about 18% of the world's tropical forests (Achard et al., 2002) and acts as a carbon sink, storing 50 billion tons of carbon (Verhegghen et al., 2012). This makes the CRB susceptible to hydroclimatic changes that need to be better understood (Becker et al., 2018). Climate change and anthropogenic landuse changes have led to changes in basin hydrology and pose threats to water resource availability, ecosystems, and greenhouse gas emissions (Julie et al., 2008; Collier et al., 2008; DeFries and Rosenzweig, 2010; Gibbs et al., 2010; Malhi and Grace, 2000; Tshimanga, 2012; Brooks et al., 2011; Molinario et al., 2015).

4.3.3 Narmada River Basin

The Narmada River basin (NRB) is located in central India (Fig. 4.1b) between latitudes $21^{\circ} 22' 0''$ N and $23^{\circ} 46' 30''$ N, and longitudes $73^{\circ} 4' 0''$ E and $81^{\circ} 45' 30''$ E. The Narmada River originates from the Amarkantak Plateau in Madhya Pradesh, flows through Maharashtra and Gujarat, and into the Arabian Sea (Pandey and Khare, 2018). The NRB drains $20,561 \text{ km}^2$, with elevations ranging from nearly zero to over 1000 m above sea level, and an average slope of 1.1° . The upper reaches are bounded by hills, with a fertile plain region between the hilly tracts and in the lower reaches. The basin is comprised of bedrock ranging as far back as the Proterozoic era. Deccan basalts form the dominant lithology of the region, followed by Quaternary soils, thick sedimentary outcrops of Vindhyan and Gondwanas, and scattered patches of other rocks (Subramanian et al., 2006; Gupta et al., 2011). The lower basin has substantial surface exposures of Mesozoic rocks, Deccan basalts, and Tertiary sediments (Pandey and Khare, 2018). The NRB has a subtropical climate with a hot, dry summer and a cool winter (Mondal et al., 2015).

The Narmada River is the fifth largest river in India and passes through three states that face water shortages during non-monsoon seasons, making proper water use from the river crucial to the inhabitants (Ray and Goel, 2019). The Narmada is a rain-fed river, with heavy rainfall occurring between June and October. Climate change is expected to cause seasonal or regular water-stressed conditions in the basin (Gosain et al., 2006; Gupta et al., 2011).

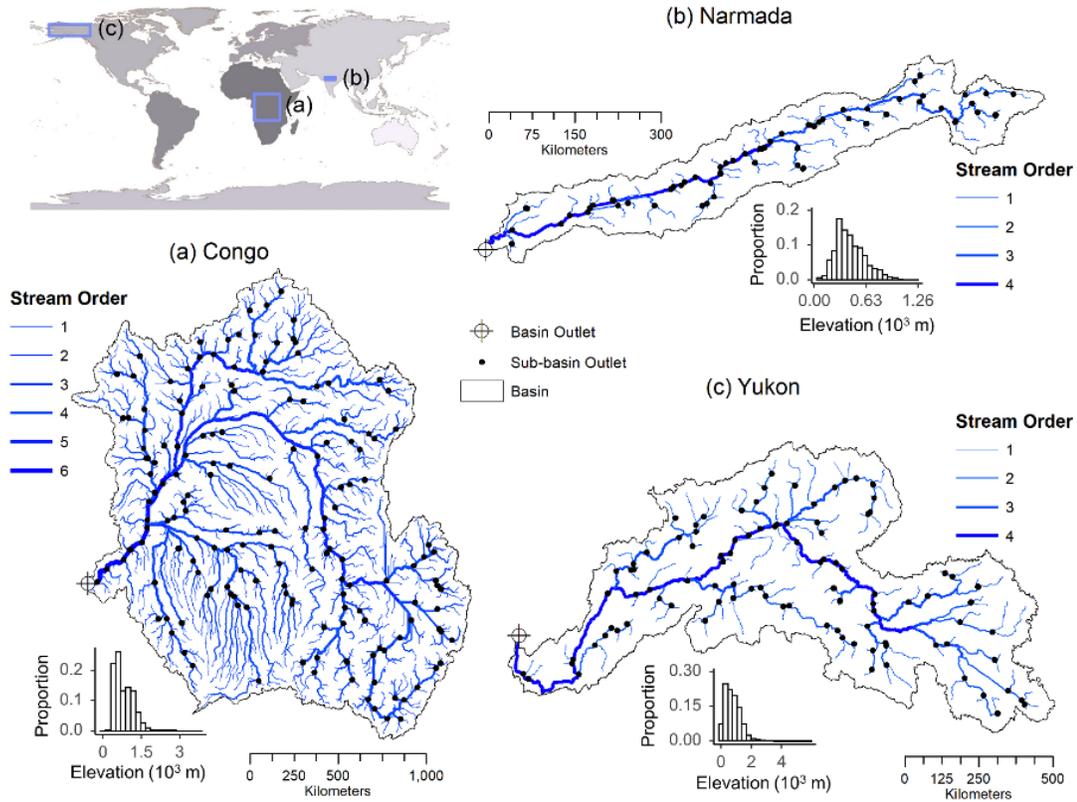


Figure 4.1: Regional map showing the study regions: (a) the Congo River basin, (b) the Narmada River basin, and (c) the Yukon River basin. The respective drainage areas are estimated to be $3.75 \times 10^6 \text{ km}^2$, $9.54 \times 10^4 \text{ km}^2$, and $8.36 \times 10^5 \text{ km}^2$, respectively. The elevations range from 1 m to 3777 m, 8 m to 1202 m, and 1 m to 5791 m above the NGVD, respectively. Histograms of elevation are shown beside each basin. This map also shows the channel network delineated from a DEM and outlet points manually selected to delineate sub-basins for the study. A total of 221 sub-basins were delineated for the Congo, 72 for the Narmada, and 126 for the Yukon.

4.3.4 Yukon River Basin

Yukon River basin (YRB) spans the western Canadian and the central Alaskan regions (Fig. 4.1c), with latitudes between $58^{\circ} 58' 30''$ N and $69^{\circ} 5' 0''$ N, and longitudes between $164^{\circ} 20' 0''$ W and $129^{\circ} 12' 0''$ W. The Yukon River begins at the Llewellyn Glacier in Canada, flows through the Teslin River, continues through Alaska, and ends in the Bering Sea. The YRB is the fourth largest drainage basin in North America with an area of about $857,300 \text{ km}^2$. The elevation ranges from nearly zero to above 5500 m above sea level. The average slope is 3° . The YRB has a complex geography, with various types of consolidated rocks in the mountain ranges around the basin and unconsolidated sediment deposits in the lowlands. These rocks range in age from Precambrian to Holocene (Brabets et al., 2000). The primary runoff patterns in the basin are lake runoff, snowmelt runoff, and glacier runoff. The YRB has a subarctic to arctic climate and is underlain by permafrost, which acts as a barrier to deep percolation, leading to a high ratio between the maximum and minimum river flows (Ge et al., 2013). The majority of streamflow occurs in the summer from snowmelt, rainfall, and glacial melt (Brabets et al., 2000). The YRB has 20 ecoregions, with the most dominant being the interior forested lowlands and uplands, and the interior highlands (Gallant et al., 1995; Group et al., 1995). The Yukon River is important to the Bering Sea ecosystem as it provides most of the freshwater runoff in the eastern part of the Bering Sea (Lisitsyn, 1969). Because of its location, the YRB region has undergone significant climate change with wide-ranging effects: average temperature, snowmelt patterns, and streamflow runoff (Chapman and Walsh, 1993; Brugman et al., 1997; Brown and Braaten, 1998; Nijssen et al., 2001; Manabe et al., 2004; Shulski and Wendler, 2007; Walvoord and Striegl, 2007; Brabets and Walvoord, 2009).

4.4 A Review of Area-Elevation Relationships

4.4.1 Hypsometric Curve

The hypsometric curve is a non-dimensional area-altitude distribution curve (Horton, 1932; Langbein, 1947) that demonstrates how mass is distributed as a function of elevation (Harlin, 1984). The simplest hypsometric curve is plotted in absolute units, with the elevation in meters on the ordinate (y -axis), and the area in square kilometers lying above the given elevation in the abscissa (x -axis). On the other hand, the percentage or scaled hypsometric curve is plotted as the proportion of area above a given elevation (relative area above elevation) against the ratio of elevation and relative relief (relative elevation). This allows for the comparison of basins of different sizes and altitudes (Strahler, 1952). Although the curve can take a variety of shapes depending on the distribution of landmass, the endpoints are constrained (Fig. 4.2) and have been used as boundary conditions (BC) in the functional estimation presented in the later sections.

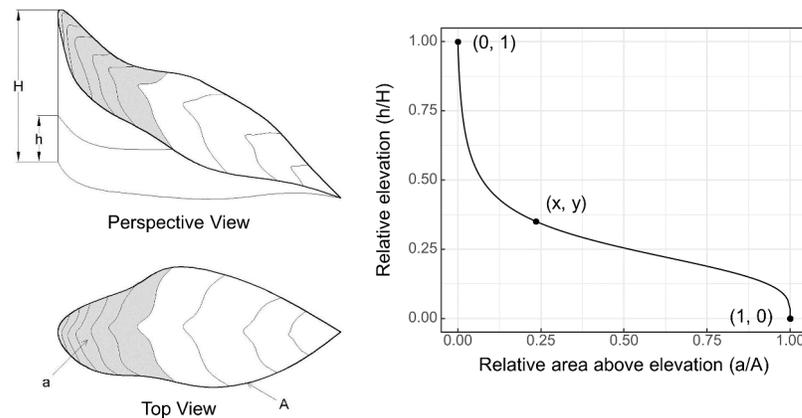


Figure 4.2: The scaled hypsometric curve. Relative elevation is plotted against the proportion of area above this elevation. The span of the curve has two associated boundary conditions: the upper left-hand corner ($x = 0$, $y = 1$) and the lower right-hand corner ($x = 1$, $y = 0$).

Traditionally, hypsometric curves were plotted by finding the areas between successive pairs of contour lines determined by a planimeter (Horton, 1932). Haan

and Johnson (1966) devised a method for obtaining the hypsometric curve based on a random sampling technique. With the rapid improvement of computational power and the availability of digital elevation data, hypsometric curves can be plotted from digital elevation data with high accuracy. The hypsometric curve has traditionally been used for evaluating the geomorphic maturity of catchments (Strahler, 1952; Moglen and Bras, 1995; Pedrera et al., 2009). Hypsometric curves in young (mature) drainage basins have a convex (concave) shape with a relatively large (small) proportion of uneroded topography present. Willgoose and Hancock (1998) divide the hypsometric curve into three regions: the 'head' on the upper left-hand side, the 'toe' on the lower right-hand side, and the 'body' in-between the two, with the shape of the toe related to the stream network and branching in the basin.

Harlin (1984) found a correlation between the watershed time-to-peak and various statistical parameters of the hypsometric function. They found that the moments of the function were related to the headward drainage development, the mid-basin slope, and the rate of change in upper and lower basin slope, and proposed that the use of hypsometry alongside basin area and/or relief can increase the predictive power and transferability of rainfall-runoff models. Furthermore, Howard (1990) found that basin size and average channel gradients provided predictive power in estimating the basin hydrologic response. Willgoose and Hancock (1998) linked the hypsometry with landform processes, showing that hypsometry not only reflects landscape runoff and erosion processes, but is also strongly dependent on catchment geometry, with the hypsometric curve being a sensitive indicator of the planar form of the drainage network. Marani et al. (2001) noted that basins with a large proportion of low elevation areas produce significant interaction with the groundwater when it is relatively deep. Furthermore, basins with a larger variance in the elevation distribution are more efficient in dissipating upward fluctuations of the water table. Vivoni et al. (2008) found that surface and

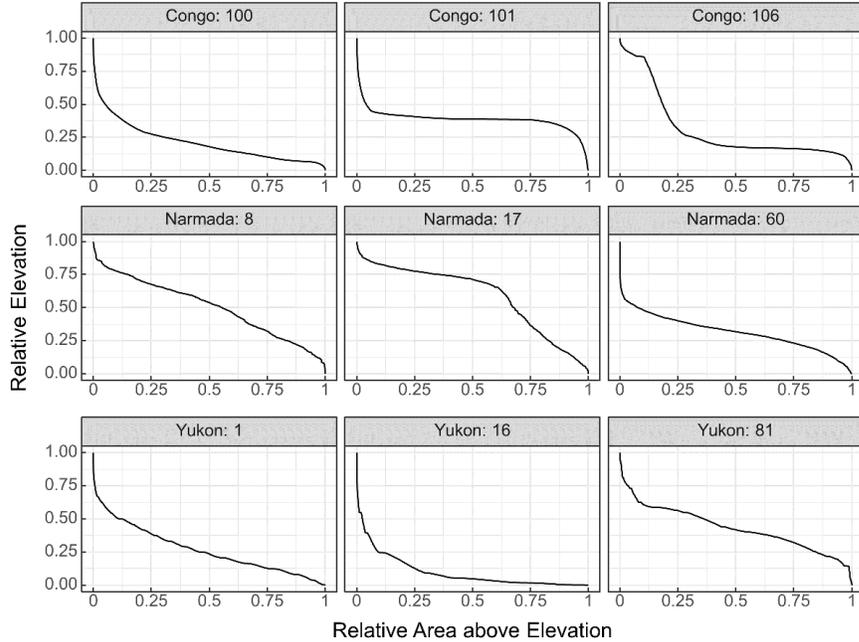


Figure 4.3: Sample hypsometric curves for selected sub-basins. The hypsometric curves have a variety of shapes depending on the distribution of landmass at different elevations.

subsurface runoff components of a basin are a function of the basin hypsometric form. Generally, relatively less eroded basins (with convex hypsometric curves) show higher total runoff that are dominated by subsurface processes, whereas relatively more eroded basins (with concave hypsometric curves) show less total runoff with a larger proportion of surface response. They also linked the basin hypsometry with the spatial distribution of soil moisture and runoff response mechanisms.

Fig. 4.3 shows empirical hypsometric curves for three sample sub-basins in the three study basins, illustrating a variety of shapes. Sub-basins 100 and 106 of the CRB, sub-basin 60 of the NRB, and sub-basins 1 and 10 of the YRB exhibit concave upwards shapes, indicating a mature stage, whereas sub-basin 17 of the NRB has a concave downward shape, indicating a young stage. Furthermore, a wide range of head and toe shapes are apparent, with a prominent concave-down toe curvature in the CRB sub-basin 101 and an absence of concave-down toe curvature in the CRB sub-basin 100, the NRB sub-basins 8 and 17, and the YRB sub-basins 1 and 16.

4.4.2 Hypsometric Integral

The most widely used parameter in the hypsometric analysis is the hypsometric integral (HI), first applied to landforms and regions by Imamura (1937) and Péguy (1942), and popularized by Strahler (1952) who defined it as the area under the hypsometric curve (Eq. (4.2) in Table 4.1). Because of the difficulty in finding the integral value, Chorley and Morley (1959) developed a method for approximating the HI using a simple geometric form, the ratio of the height of the contour dividing the plane projection of the basin in half ($H_{0.5}$) to the total height of the basin (H). Subsequently, Pike and Wilson (1971) proved the HI to be mathematically identical to another metric, the Elevation-Relief Ratio (RR) defined by Wood and Snell (1960) as the relative proportion of upland area to lowland area within a given region (Eq. (4.3) in Table 4.1). Cohen et al. (2008) developed a methodology for calculating the spatial distribution of the HI within a catchment by treating each DEM pixel as a local catchment outlet and averaging at the sub-catchment scale to reduce noise.

The HI is directly linked to the hypsometric curve, with a convex curve exhibiting a higher HI and a concave curve exhibiting a lower HI. As a consequence, a higher HI is an indicator of a young catchment, whereas a lower HI is an indicator of a mature catchment (Strahler, 1952, 1964). Walcott and Summerfield (2008) showed that the analysis of the spatial variation of the HI in basins of different size ranges can be helpful in assessing the factors that control landscape development, such as the relative role of crustal displacement and variations in lithological resistance. The HI is also an indicator of the erosional status of watersheds (Singh et al., 2008). Because the HI is an easily calculated, single-valued index, it can be conveniently used in studying the geomorphic maturity of basins, including erodability across Martian catchments (Nicholson et al., 2013). Furthermore, a lower HI in concave basins indicates lower total runoff with a higher surface

contribution, whereas a higher HI is linked with groundwater-discharge-dominant basins (Vivoni et al., 2008; Luo, 2000). Huang and Niemann (2008b) further suggested that a high HI is produced by space-time patterns of streamflow generation only during early development stages, with a high HI not expected to occur in later states unless lithological or geomorphological processes such as sapping, channel wall slumping, bank failure, etc. cause it to be preserved. Moreover, Cohen et al. (2008) showed that their methodology for calculating the spatial distribution of the area-slope equation and the HI within a catchment not only correlated well with landscape concavity, but also demonstrated its utility in potentially identifying the spatial extent of soil types and geological features.

Two basins can resemble one another with respect to geometric aspects such as local relief and average slope despite having an appreciably different HI (Strahler, 1952). Width to length ratios of catchments tend to have a stronger influence on the hypsometric curve but little influence on the HI (Willgoose and Hancock, 1998). Pérez-Peña et al. (2009) noted that the HI is independent of catchment relief. Liyanagamage and Hewa (2012a) found no significant correlation between the HI and flow statistics such as the mean, coefficient of variation, lag-one autocorrelation coefficient, and Q50-Q10 range flow values. Furthermore, using the HI leads to a question of uniqueness, and the use of the hypsometric curve has a greater potential for quantitative landform analysis compared to the use of the HI alone (Harlin, 1978).

The distribution of HIs in the sub-basins in this study is illustrated in Fig. 4.4. HIs in the CRB range between 0.136 and 0.893, with a mean value of 0.414, and exhibit a slightly right-skewed distribution. HIs in the NRB range from 0.114 to 0.621, with a mean value of 0.344. HIs in the YRB range between 0.066 and 0.488, with a mean value of 0.288, and exhibit a strongly left-skewed distribution. Whereas the CRB and the NRB seem to have a mix of young and mature sub-basins, the

YRB seems to have relatively younger sub-basins based on the HI distribution. Moreover, these might also indicate that sub-basins in the YRB have lower total runoff with higher surface contributions compared to sub-basins in the CRB and the NRB.

4.4.3 Hypsometric Functions

We first present a comprehensive summary of the functional forms available in the published literature in Table 4.1. Hypsometric functional forms, when fitted to elevation data, provide important information regarding the basin. For example, the area under the hypsometric curve is used to define the HI. The inflection points of the curve are morphologically significant as they mark the level at which the rate of

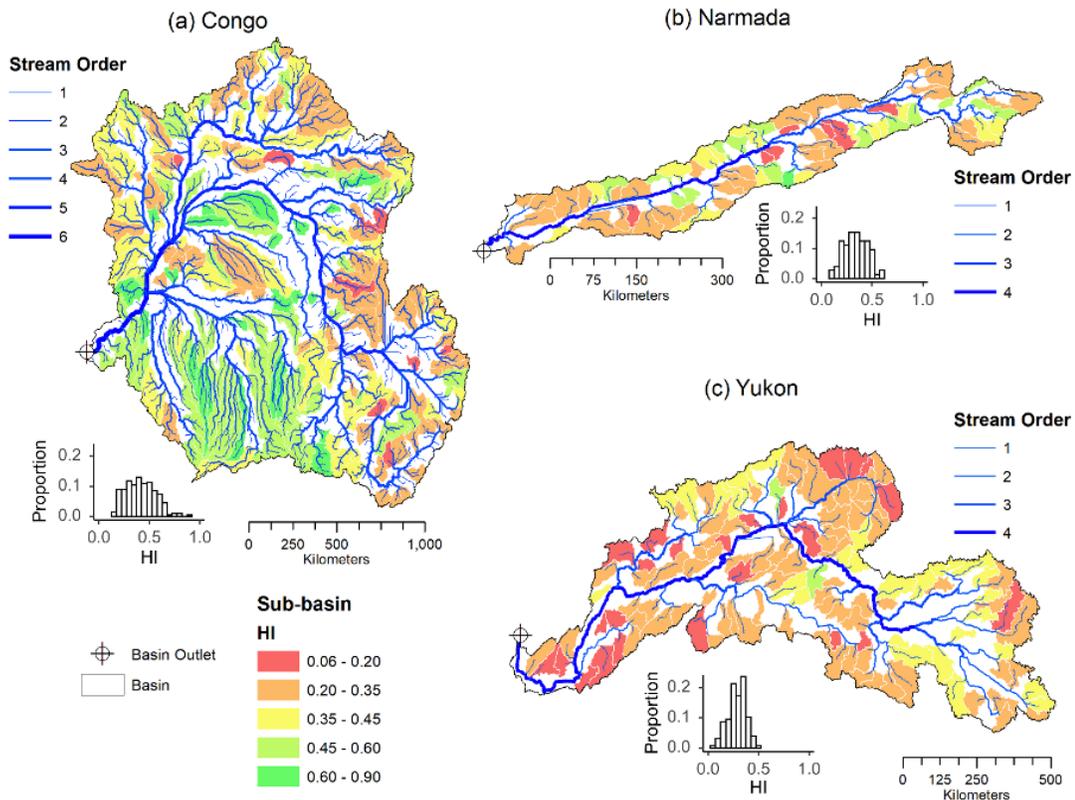


Figure 4.4: Map showing the distribution of hypsometric integral (HI) values in sub-basins in the three study regions: (a) the Congo River basin, (b) the Narmada River basin, and (c) the Yukon River basin. Histograms showing the distribution of HI values are shown alongside each basin map.

decrease in mass upwards changes from a rapid to a diminishing rate (Strahler, 1952). Moreover, the shape and size of the toe correspond to the increasing elevation contributions of hillslopes in the lower elevational regions of the catchment and reflect the form of lateral contribution and the degree of branching of the drainage pattern (Willgoose and Hancock, 1998).

Strahler (1952) formulated a functional form to describe hypsometric curves using a three-variable functional form for the unscaled hypsometric curve, and a two-variable functional form for the scaled curve (Eq. (4.1) in Table 4.1). The Strahler function addresses the commonly present up-concavity in the head and the down-concavity in the tail, while allowing the sinuosity to vary greatly (Strahler, 1952). With larger values of two parameters, r and z , the curve has a simple concave-up form, whereas, at lower values, the curve takes an S-shape, with a concave-up head and a concave-down tail. Although the Strahler function provides an adequate fit for a broad range of common hypsometric curve shapes, the formulation fails to capture the shapes of the head and the tail in some cases because of the limitations of the two-parameter formulation as demonstrated in Fig. 4.5. The r parameter equally affects the curvature of the head as well as the tail, whereas the z parameter more significantly affects the curvature of the tail. Hence, the lack of a parameter primarily controlling the curvature of the head restricts the versatility of the function. Harlin (1978) introduced the use of statistical moments of the hypsometric curve to extract more information than what is possible using just the HI. He described hypsometric curves for several mid-western basins in the US using a Polynomial function (Eq. (4.4) in Table 4.1). The third moment of the function about the x-mean provides hypsometric skewness and the fourth moment provides the hypsometric kurtosis. The third and fourth moments of the first derivative of the function provide density skewness, and density kurtosis, respectively. These four parameters are related to the amount of headward erosion

Table 4.1: A chronological review of quantitative approaches for the estimation of hypsometry in drainage basins. For hypsometric functions, y denotes the normalized elevation, and x denotes the normalized area above the elevation. Boundary conditions to constrain the functions to left top ($x = 0, y = 1$) and right bottom ($x = 1, y = 0$) positions have been provided for functions that were used in this study.

Name	Formulation / Technique	Parameters	References
Hypsometric Curve	$x = 1 - \text{cdf}(y)$		Horton (1932)
Strahler	$y = \left[\frac{r}{1-r} \right]^z \left[\frac{1}{(1-r)x+r} - 1 \right]^z$ (4.1)	$r \in (0, 1], z > 0$	Strahler (1952)
Hypsometric Integral	$\text{HI} = \int_0^1 x dy$ (4.2)		Strahler (1952)
Geometrical Approximation	$\text{HI} \approx 97.7 \frac{H_{0.5}}{H}$		Chorley and Morley (1959)
Elevation-Relief Ratio	$\text{RR} = \frac{\text{Elevation}_{\text{mean}} - \text{Elevation}_{\text{min}}}{\text{Elevation}_{\text{max}} - \text{Elevation}_{\text{min}}}$ (4.3)		Wood and Snell (1960)
Random Sampling	$x \approx 1 - \text{cdf}(y^*), y^* \subset y$		Haan and Johnson (1966)
Polynomial	$\text{HI} = \text{RR}$ $y = a_0 + a_1x + a_2x^2 + a_3x^3$ (4.4)	$a_0, a_1, a_2, a_3 \in (-\infty, \infty)$ BC: $a_0 = 0, a_3 = 1 - a_1 - a_2$	Pike and Wilson (1971) Harlin (1978), Harlin (1984), Luo (2000), Luo (2002), Luo and Harlin (2003), Huang and Niemann (2008b), Gallen et al. (2011), Ssegane et al. (2012), Vanderwaal and Ssegane (2013), Liffner et al. (2018)
Yield-density	$y = \frac{1}{a + b^x}$	a, b	Sarkar and Patel (2011)

Table 4.1 Continued

Logistic	$y = \frac{a}{1 + be^{-cx}}$	a, b, c	Sarkar and Patel (2011)
Rational	$y = \frac{a + bx}{1 + cx + dx^2}$	a, b, c, d	Sarkar and Patel (2011)
Sigmoidal	$y = \frac{ab + cx^d}{b + x^d}$	(4.5) $a, b, c, d \in (-\infty, \infty)$ BC: $a = 1, c = -b$	Sarkar and Patel (2011), Vanderwaal and Ssegane (2013)
Modified Polynomial	$y = \frac{a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4}{(1-x)^{a_4} + a_5x^{a_6}}$	(4.6) $a_0, a_1, a_2, a_3, a_4, a_5, a_6$	Ssegane et al. (2012), Vanderwaal and Ssegane (2013)
Double Exponential	$y = a_1e^{a_2x} + a_3e^{a_4x}$	(4.7) $a_1, a_2, a_3, a_4 \in (-\infty, \infty)$ BC: $a_1 = 1 - a_3,$ $a_3 \in (-\infty, 0) \cup (1, \infty),$ $a_4 = \ln\left(-\frac{(1-a_3)e^{a_2}}{b}\right)$	Vanderwaal and Ssegane (2013)
Generalized Hypsometric	$y = \left[\frac{r(1-x^m)}{(1-r)x^m + r} \right]^z$ or (4.8) $r \in (0, 1], z > 0, m > 0$	(4.9) $\beta (= \frac{1-r}{r}) > 0, z > 0,$ $m > 0$	This study

cdf: Cumulative Density Function, HI: Hypsometric Integral, RR: Elevation-Relief Ratio, BC: Boundary Conditions,

$H_{0.5}$: Height of the contour dividing the plane projection of the basin in half, H : Total height of the basin

in the basin’s upper reaches, the erosion in both upper and lower reaches, the rate of change in the upper and lower basin slopes, and the change in the mid-catchment slopes, respectively (Harlin, 1978; Luo, 2000). Harlin (1978) and Liffner et al. (2018) suggested that third-order (O3) Polynomial (Eq. (4.4) in Table 4.1) sufficiently describes hypsometric curves in a majority of cases. Huang and Niemann (2008b) used a Polynomial form with normalized elevation as the independent variable instead of normalized area. Ssegane et al. (2012) found that a seven-parameter Modified O3 Polynomial model with a rational term (Eq. (4.6) in Table 4.1) provided the best fit for watersheds in three physiographic regions of the Appalachian Plains, Piedmont, and Ridge and Valley. They found an R^2 value of over 0.999 in more than 90% of the cases.

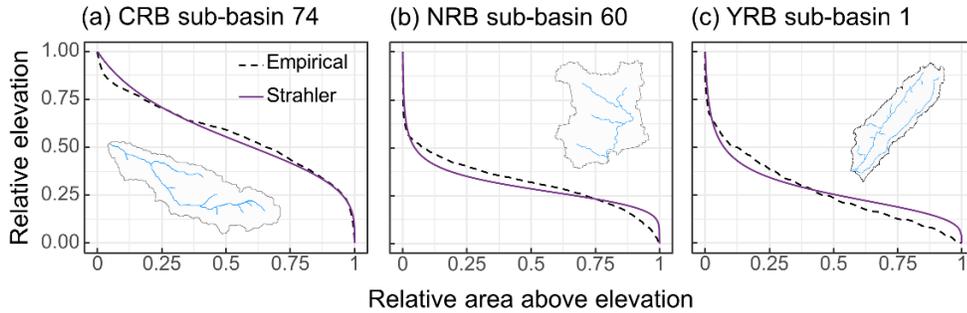


Figure 4.5: Strahler fit for three sample sub-basins, (a) sub-basin 74 of the Congo River basin (CRB), (b) sub-basin 1 of the Narmada River basin (NRB), and (c) sub-basin 1 of the Yukon River basin (YRB). The sub-basins with their stream networks are shown alongside the hypsometric curves.

Sarkar and Patel (2011) tested five functions to fit watersheds in the Dulung River basin, India, and found that the Sigmoidal function (Eq. (4.5) in Table 4.1) provided the best fits for the Monadnock phase watersheds in their study, with the coefficient of determination between 0.98 and 0.99. However, in researching watersheds in the eastern belt mountains of Pahang State in Malaysia, the first author noted that the Sigmoidal model did not sufficiently fit the watersheds, and found that a Double Exponential function (Eq. (4.7) in Table 4.1) performed better for basins with strongly concave-up hypsometric curves. Vanderwaal and Ssegane

(2013) tested an O3 Polynomial function (Eq. (4.4) in Table 4.1), a Modified O3 Polynomial function (Eq. (4.6) in Table 4.1), a Sigmoidal function (Eq. (4.5) in Table 4.1), and a Double Exponential function (Eq. (4.7) in Table 4.1) in 32 Monadnock phase watersheds from three countries (Malaysia, United States, and India), and found that the latter three performed better than the Polynomial function. However, they noted that each equation failed occasionally in giving logical values for hypsometric attributes, and suggested the importance of using multiple functions and averaging.

4.5 A Generalized Hypsometric Function

4.5.1 Generalized Formulation

We propose a three-parameter Generalized Hypsometric function (Eqs. (4.8) and (4.9) in Table 4.1) that builds upon the Strahler function with one additional parameter. As shown in Fig. 4.5, the Strahler formulation appears less adaptable to curvatures linked to the head and toe of the hypsometric curves. The additional parameter, m , affords greater versatility to the function and allows it to attain a better fit for more diverse hypsometric curve shapes without violating the established boundary conditions. Eq. (4.8) presents a form consistent with the Strahler function, allowing for the comparison of parameters, whereas Eq. (4.9) presents a more compact form of the same equation. The proposed function allows for the efficient estimation of parameters using a nonlinear fitting algorithm. Details regarding the computational procedures are described in Appendix B.1.

4.5.2 Parameter Interpretation

We illustrate the effect of each parameter on the shape of the hypsometric curve by plotting a set of synthetic plots (Fig. 4.6). When each parameter is equal to one, the function takes the form of a straight line from the top left ($x = 0, y = 1$) to the

bottom right ($x = 1, y = 0$). Fig. 4.6a-c illustrate families of curves by varying r , z , and m values, respectively, while keeping the other parameters constant at one. Decreasing the r values from 1 to 0.01 increases the degree of upward-concavity (Fig. 4.6a). As z values decrease below one, the hypsometric curve takes a concave-down shape, and as the values increase above one, the curve takes a concave-up shape (Fig. 4.6b). It should further be noted that the z parameter more significantly affects the curvature of the toe. On the other hand, as m values decrease below one, the hypsometric curve becomes concave-up, and as the values increase above one, the curve becomes concave-down (Fig. 4.6c). In contrast to the z parameter, the m parameter more significantly affects the curvature of the head. Thus a combination of z and m parameters allows for finer controls over the curvature of the head and the toe individually to an extent, allowing the r parameter to control the overall shape of the body. This is further demonstrated in Fig. 4.6d, which shows two different S-shaped hypsometric curve forms obtained with a mixture of values for the three parameters. Setting values of the z and m parameters above one leads to a concave-up toe and a concave-down head, respectively, whereas a concave-down toe and a concave-up head results from setting z and m parameter values below one. This further illustrates the flexibility in shapes offered by the three-parameter model beyond what the simple Strahler function offers as described in Section 4.4.3. The curvature of the hypsometric curve is indicative of the developmental stage of a basin (Strahler, 1952). Consequently, different combinations of parameter values can suggest their developmental stage. A combination of $r < 1$, $z > 1$, and $m < 1$ leads to a concave-up hypsometric curve, possibly indicating a mature developmental stage, whereas a combination of $z < 1$ and $m > 1$, with r close to one, leads to a concave-down curve, indicating a young developmental stage. Furthermore, different combinations can also suggest the distribution of fluvial processes in the catchment. A lower m value indicates a high degree of erosion in the upper catchment region,

whereas higher z values indicate high erosion in the lower catchment region. As such, low r and m values and a high z value suggest the dominance of fluvial processes. On the other hand, a combination of high z and m values indicates fluvial processes in the upper region and diffusive processes and mass accumulation in the lower region. As a consequence of the three parameters of the Generalized Hypsometric function affording finer controls over the head, body, and the toe of the hypsometric curve, and thereby the mass distribution in different regions of the catchment, they can offer deeper insights into the topographic features and the dominant erosional/ depositional processes in the catchment compared to single-valued metrics like the HI.

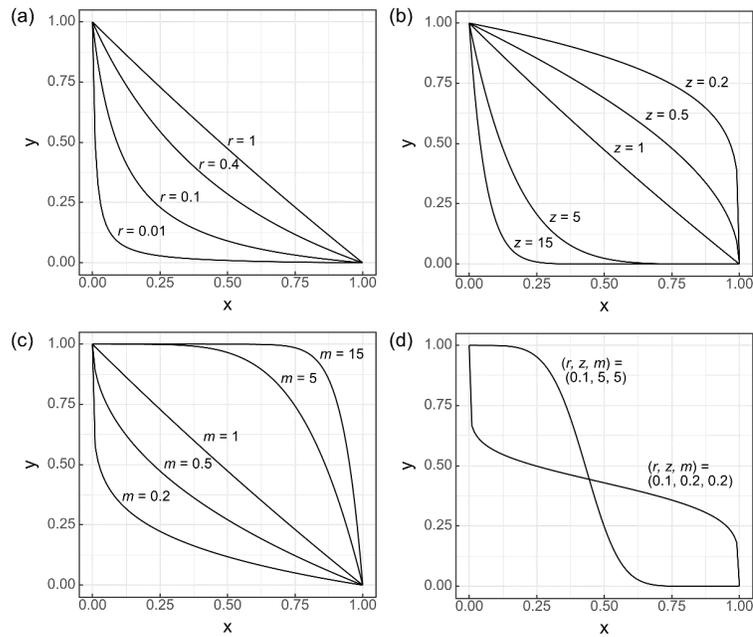


Figure 4.6: Synthetic plots of the Generalized Hypsometric Function with a range of values for the (a) r parameter, (b) z parameter, (b) m parameter, and (d) for a combination of r , z , and m parameter values.

Next, we compare the parameter values between the Strahler and the Generalized Hypsometric functions. The Strahler function can be seen as a special case of the Generalized Hypsometric function with the m parameter equal to one. To this end, the distribution of the m parameter around unity and the change in the

distribution of other parameters illustrate the benefits of the added parameter. Fig. 4.7 shows a cross-comparison of the three-parameter estimates for each sub-basin in this study. The z parameter estimates follow similar trends for both functions (Fig. 4.7b). On the other hand, the r parameter estimates do not show any correlation between the two functions, with numerous estimates simply taking a value of one in the Generalized Hypsometric fit (Fig. 4.7a). Finally, the m parameter estimates take a variety of values in the Generalized Hypsometric function (4.7c). The most

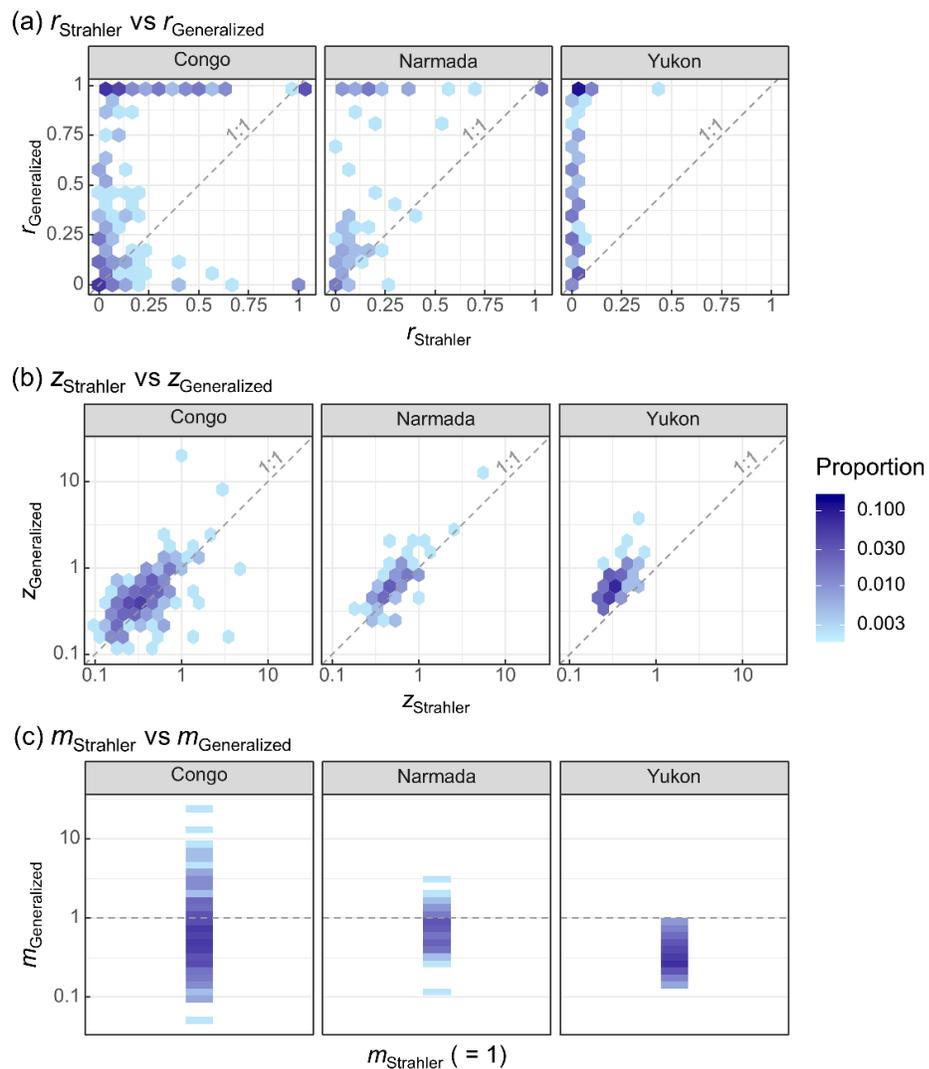


Figure 4.7: Distribution of (a) r , (b) z , and (c) m parameter values for Strahler and Generalized Hypsometric function fits for each sub-basin. $m = 1$ for Strahler is represented by a horizontal dashed line at unity.

notable case is that for the YRB sub-basins. Although the r parameter estimates have a low variance in the case of the Strahler fits, they are more evenly distributed in the Generalized Hypsometric fits. In addition, unlike other basins in the study where the m parameter estimates have values both larger and smaller than one, all estimates for the YRB are smaller than one. The YRB is largely comprised of sub-basins that have high HIs. As such, the Strahler fit sets the r parameter close to zero to form the overall concave-up shape of the head and z parameter at values less than one to form a concave-down toe (Fig. 4.3). Instead, the Generalized Hypsometric fit uses the m parameter to form the pronounced concave-up head (Fig. 4.3), allowing more variability in the r parameter to control the body. Notably, the m parameter values are closely distributed around one for the NRB, indicating that the added parameter adds little in this region, while the large variance around one in the CRB indicates that the m parameter contributes to both the concave-up and concave-down head shapes.

Glaciers have significant effects on the runoff systems in the YRB. All m values being less than one, which results in concave-up hypsometric curves, could be the result of glacial erosion in the Yukon. Furthermore, low m values indicate a high degree of erosive processes by glaciers at higher elevations, leading to a shoulder in the hypsometric curve (Montgomery et al., 2001). Similarly, the diverse landforms in the CRB are reflected in the large variance in parameter values. The presence of significantly high m values results from a flat head shape followed by a steep drop in the body or the tail region of the hypsometric curve (Fig. 4.6c) and could, therefore, be indicative of large canyons and waterfalls with a sudden drop in elevations in the CRB. Furthermore, specific combinations of the parameters reflect the presence of plateaus. For instance, the northern region of the CRB consists of plateaus with elevations around 610 m (Leturmy et al., 2003), and these are represented by low

values of r and z (Fig. 4.8). This combination creates an S-shaped curve with steep head and toe regions and a relatively flat body.

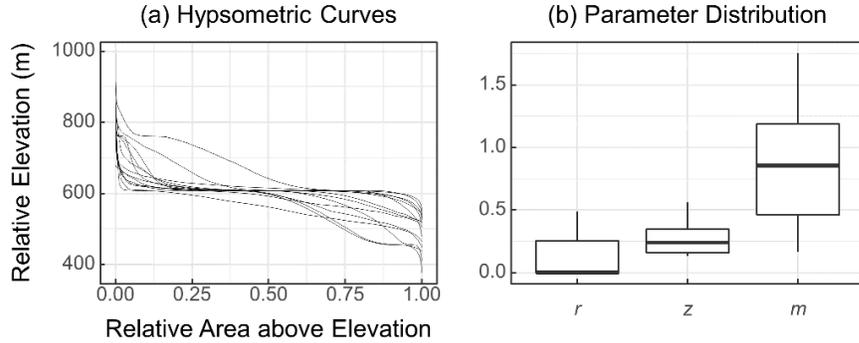


Figure 4.8: (a) Hypsometric curves of northern plateaus of the Congo River basin and (b) their parameter distribution for Generalized Hypsometric fits. The y-axis of the hypsometric curves has been plotted in absolute units to illustrate that the plateaus have elevations of around 610 m.

4.5.3 Inter-comparison of Functional Forms

We now compare the various functional forms that have been shown to be promising in literature, along with the Generalized Hypsometric function, to determine which function best captures the diverse hypsometric curve shapes. These include the Strahler function (Strahler, 1952), the Polynomial function (Harlin, 1978), the Sigmoidal function (Sarkar and Patel, 2011), and the Double Exponential function (Vanderwaal and Ssegane, 2013). The latter three functions were modified to impose the BCs leading to modified equations (Eqs. (B.4), (B.5), and (B.6)) to maintain consistency and to increase their interpretability to describe various aspects of the hypsometric curve. We used root mean square error (RMSE) to evaluate the goodness of fit of these functions (see Section B.6 for details).

Fig. 4.9 shows each of these functions plotted over the empirical hypsometric curve of three sample sub-basins. CRB sub-basin 74 has a concave-down hypsometric curve with a pronounced toe curvature and a moderate head curvature (Fig. 4.9a), while NRB sub-basin 60 has a concave-up curve with a pronounced

head curvature and a moderate toe curvature (Fig. 4.9b), and YRB sub-basin 1 has a concave-up curve with a pronounced head curvature and an insignificant toe curvature (Fig. 4.9c). This plot illustrates some of the strengths and weaknesses of these functions. The O3 Polynomial function is non-monotonic. Because a hypsometric curve is an inverted cumulative density function (cdf), the Polynomial fit violates the rule of monotonicity, and as such can be a limiting factor in its application. This effect is pronounced in the case of NRB sub-basin 60 and YRB sub-basin 1. Furthermore, although this function performs well in providing various metrics such as the moments of the hypsometric curve, the non-monotonicity might lead to a poor fit for the head or the toe of the hypsometric curve. The fit fails to

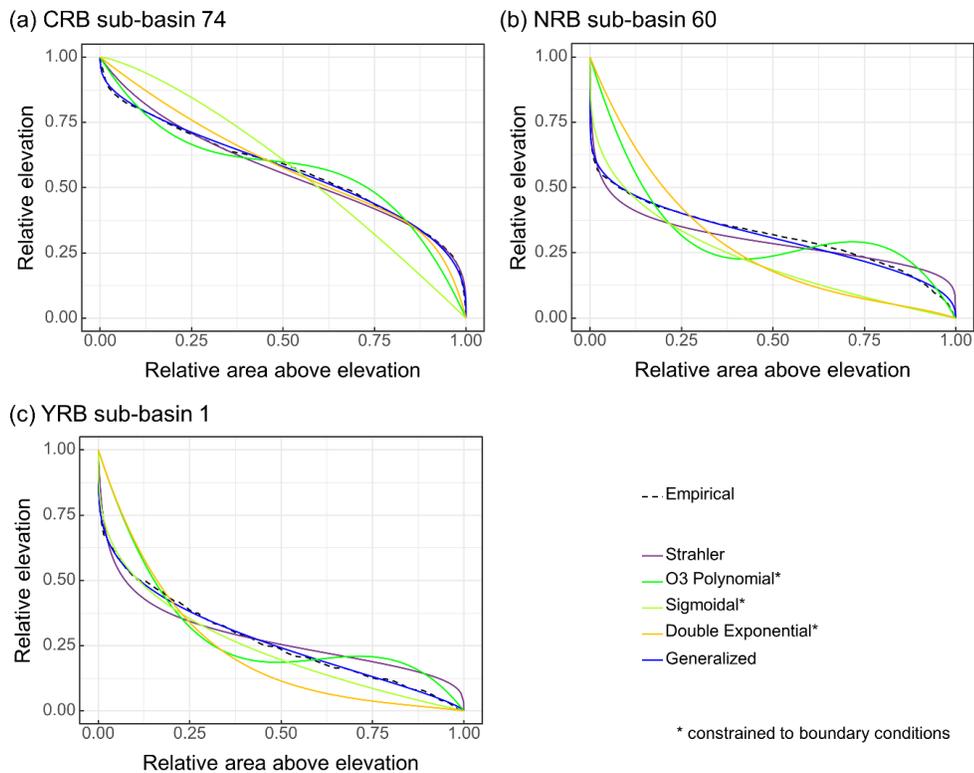


Figure 4.9: Comparison of fits for the Strahler function, the third-order Polynomial function, the Sigmoidal function, the Double Exponential function, and the Generalized Hypsometric function for three sample sub-basins, (a) sub-basin 74 of the Congo River basin (CRB), (b) sub-basin 1 of the Narmada River basin (NRB), and (c) sub-basin 1 of the Yukon River basin (YRB). The constrained forms of Polynomial, Sigmoidal, and Double Exponential functions were used.

properly reproduce the curvature of the head and the toe of both CRB sub-basin 74 and NRB sub-basin 60, while it only manages to partially capture the curvature of YRB sub-basin 1 (Fig. 4.9). The Sigmoidal function fails to properly capture the head or the toe, and has a poor fit overall for CRB sub-basin 74 (Fig. 4.9a). The Double Exponential function captures the toe curvature to an extent but fails to capture the curvature of the head. Although the Strahler function performs better in capturing the tail but misses the head, it still manages a better overall fit than the Double Exponential function. In the case of the NRB sub-basin 60 (Fig. 4.9b), the Double Exponential function provides a poor overall fit and is unable to replicate the shape of the head or the tail. The Sigmoidal fit performs better in capturing the head but appreciably misses the body and the toe. The Strahler function provides a better fit compared to the aforementioned functions. Similar to that for the NRB, the Double Exponential fit fails to provide a good fit for YRB sub-basin 1 (Fig. 4.9c). The Strahler function, on the other hand, captures the head curvature but the functional form forces it to also form a pronounced toe curvature unlike in the empirical curve. The Sigmoidal function manages to capture the shape of the head and provides an overall better fit than the aforementioned functions. In all three cases, the Generalized Hypsometric function provides the best overall fit, while better replicating both the head and the toe curvatures.

A detailed cross-comparison between the Generalized Hypsometric function and other functional forms (presented in Table 4.1) permits a thorough assessment of the relative strengths of the function over a large number of basins (Fig. 4.10). The RMSE is used as a metric to assess the relative goodness of fit. The RMSE values for the Generalized Hypsometric function were found to be consistently lower than those for other functional forms (as evidenced by the location of the points above the 1:1 line). Among the three drainage basins, fits for the YRB show the smallest RMSE, indicative of a better fit for mature drainage basins. In the Generalized

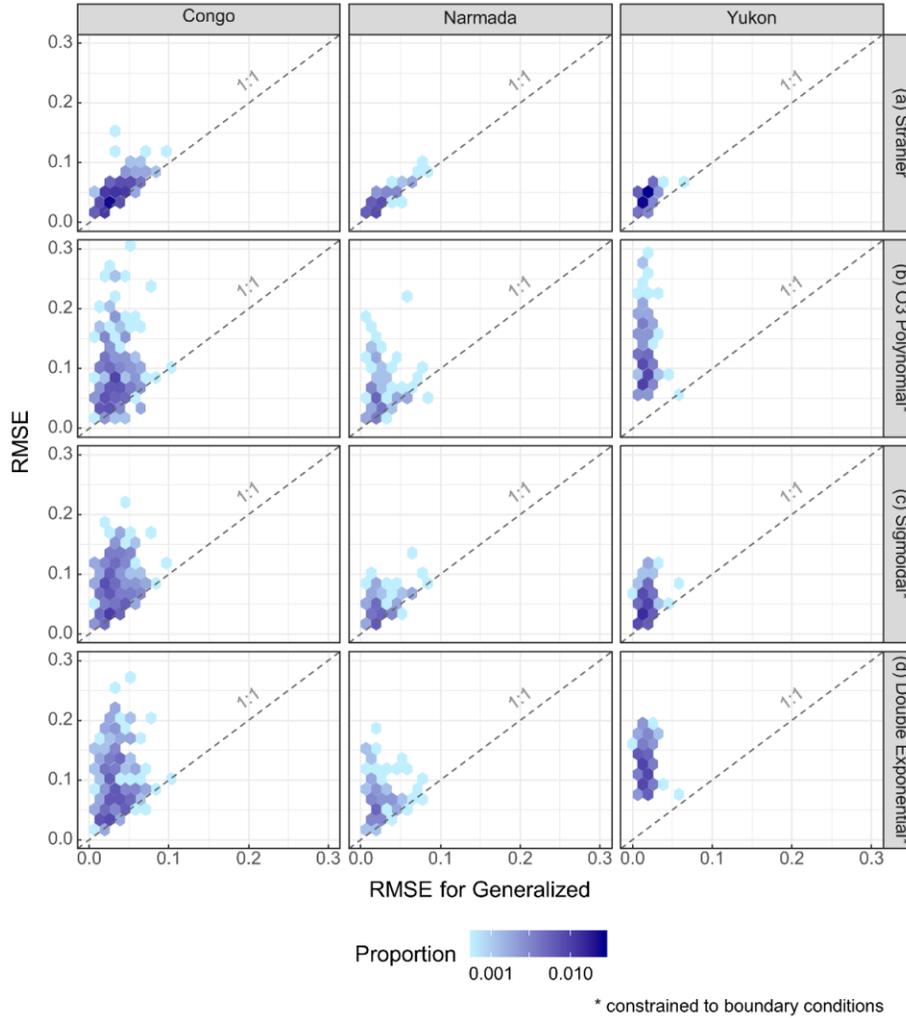


Figure 4.10: Comparison of the root mean square errors (RMSE) for the Generalized Hypsometric function with (a) the Strahler function, (b) the third-order Polynomial function, (c) the Sigmoidal function, and (d) the Double Exponential function. Hexagonal plots were used to illustrate the scatter of the values, with the color of the hexagon indicating the proportion of total observations at each region of the plot. The constrained forms of the Polynomial, Sigmoidal, and Double Exponential functions were used.

formulation, combinations of r and m parameter values allow for fits that adapt to concave-up shapes (see Section 4.5.2 for details). It is also noteworthy that the O3 Polynomial function shows the largest RMSE with a widespread, while Strahler function has the next lowest RMSE.

4.5.4 Potential Applications

Hypsometric curves hold information about geomorphic maturity (Strahler, 1952; Moglen and Bras, 1995; Pedrera et al., 2009) and hydrologic signatures (Willgoose and Hancock, 1998; Marani et al., 2001; Vivoni et al., 2008) of catchments. Metrics derived from the hypsometric curve such as the area under the curve, inflection point, curvature of the toe, statistical moments, etc., have been used to study a variety of geomorphologic and hydrologic properties of catchments (Strahler, 1952, 1964; Harlin, 1978; Willgoose and Hancock, 1998; Singh et al., 2008; Vivoni et al., 2008; Huang and Niemann, 2008b). A functional form that sufficiently fits the hypsometric curve aids the analytical assessment of catchment hypsometry and can be used to efficiently store important hypsometric information of the catchment in terms of a small set of parameters that can be used to recreate the curve.

First, we propose the use of functional form in the assessment of regionalization and similarity between different basins. Using data-rich catchments to transfer information to catchments without these data is a useful technique to predict physical phenomena such as hydrologic response in ungauged river basins (Burn and Boorman, 1993; Tung et al., 1997; McIntyre et al., 2005; Wagener et al., 2007; Reichl et al., 2009; Archfield and Vogel, 2010; Oudin et al., 2010; Patil and Stieglitz, 2011, 2012; Razavi and Coulibaly, 2013; Athira et al., 2016; Brunner et al., 2018). The hypsometric curve can be used as a metric to assess physical basin similarity (Booij et al., 2007; Ssegane et al., 2012; Hailegeorgis et al., 2015). Booij et al. (2007) used the HI as one of the 14 physical characteristics in their study in the reduction of predictive uncertainty in flows in 56 basins in the United Kingdom using regression relationships between calibrated hydrologic model parameters and physical basin characteristics. The HI was found to be related to the non-linearity and recession coefficient of the quick flow routine and the maximum percolation in the baseflow routine. Ssegane et al. (2012) assessed the statistics of hypsometric curves to

differentiate hydrologic behavior across three Mid-Atlantic physiographic regions within the USA using the Modified Polynomial function to fit the hypsometric curve (Eq. (4.6)). Hailegeorgis et al. (2015) found that the similarity of hypsometric curves was among the metrics that performed well in a regionalization study for 26 river basins in mid-Norway. Here, we propose the use of the Generalized Hypsometric function as a means of assessing similarity between the hypsometry of catchments. Because the three parameters of the function more effectively capture the erosive processes and distribution of landmass in different regions of the catchment compared to other hypsometric functions and single-valued metrics like the HI, this function should serve as an efficacious metric for finding analogous basins with similar hypsometry. For the comparison of different hypsometric curves, a discordance index (DI) was defined as the absolute area between two hypsometric curves, i.e., the integration of the absolute value of the difference between two hypsometric functions (see Section B.2 for details). The DI ranges between zero and one, with a value of zero indicating identical hypsometric functions.

Fig. 4.11 illustrates the similarity assessment of two pairs of hypsometric curves. Sub-basins 207 and 208 of the CRB have visually similar hypsometric curves and subsequently, similar HIs and a low DI (Fig. 4.11a). However, sub-basins 130 and 179 of the CRB have visually dissimilar hypsometric curves despite having similar

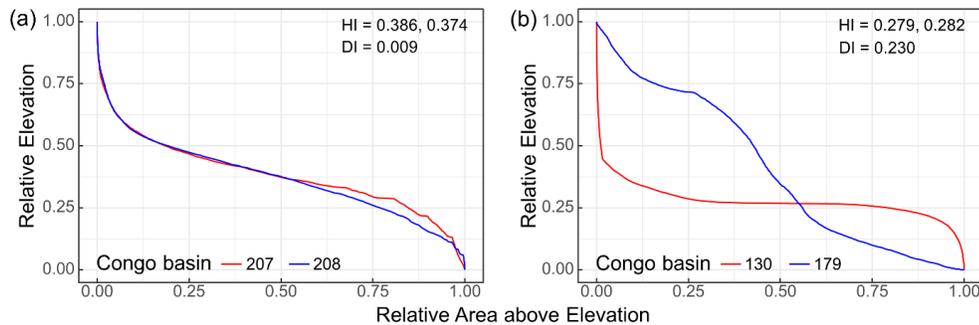


Figure 4.11: Basin similarity assessment for (a) sub-basin 207 and 208, and (b) sub-basins 130 and 179 for the Congo River basin (CRB). Hypsometric integral (HI) values of the sub-basins and discordance index (DI) values of the pairs are shown.

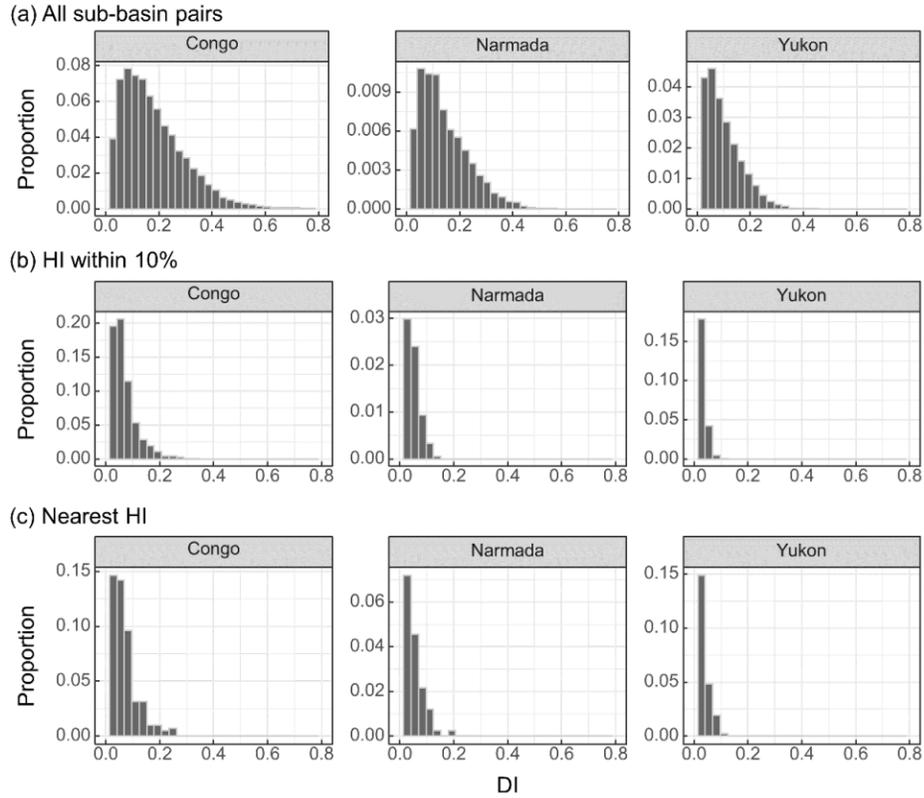


Figure 4.12: Distribution of discordance index (DI) values for (a) all pairs of sub-basins (with duplicates removed), (b) sub-basin pairs with hypsometric integral (HI) values within 10%, and (c) sub-basin pairs with closest HI values.

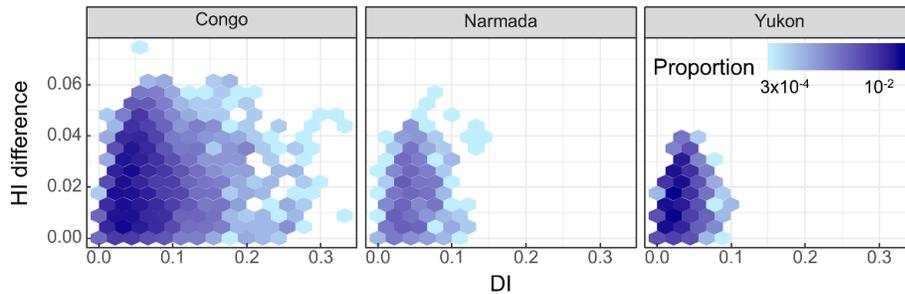


Figure 4.13: Hexagonal plot showing the joint distribution of discordance index (DI) values and the difference in hypsometric integral (HI) values for sub-basin pairs with less than 5% difference in the HI. The color of hexagons reflects the proportion of the total observations.

HI values (Fig. 4.11b). This indicates that the use of the HI as the primary metric to find analogous basins based on basin hypsometry could lead to erroneous results because these two sub-basins have different overall shapes and significantly different toe

curvature. On the other hand, the high DI value correctly reflects the difference in hypsometric curve shapes. This illustrates the shortcoming of a single-valued factor such as the HI to assess the similarity in the hypsometric curves as they do not capture the entirety of the shape, and are hence non-unique to the hypsometric curve. This is further highlighted in Figs. 4.12 and 4.13. Fig. 4.12a shows the distribution of DI values for each sub-basin pair (all unique combinations of two sub-basins) across the three basins in this study. All three histograms are right-skewed, indicating a high degree of similarity in sub-basin hypsometries (which is to be expected because they are all sub-basins within their respective study regions). It can further be noted that the YRB seems to have the lowest DI values on average, which is justified since the Yukon has a large number of sub-basins with similar concave-up hypsometric curves. Figs. 4.12b and 4.12c show histograms of the DI for sub-basin pairs with HIs within 10% and for sub-basin pairs with the closest HI values, respectively. Both these plots indicate the occurrences of appreciable variance in DI values in spite of similar HI values. It should be noted that the DIs are considerably lower compared to Fig. 4.12a, but there are still cases where low HI values fail to guarantee low DI values, meaning that the HI metric alone fails to find sub-basins with similar overall hypsometric curve shapes. Furthermore, Fig. 4.13 illustrates the variance in the DI between sub-basin pairs with similar HI values. Here, the differences in HIs are plotted against the DIs of each sub-basin pair whose HIs are within 5%. These pairs represent the sub-basins judged to be similar in nature with respect to hypsometry using the HI index. The appreciable variance in the DI, however, indicates that not all pairs have similar hypsometric curves. The difference in the entirety of the shapes can be important based on the purpose of the study. For instance, two hypsometric curves can have similar HIs while having different shapes of the toe, which can constitute an appreciable difference in hydrologic response (Willgoose and Hancock, 1998). It is

interesting to note, however, that in the case of the YRB, the HI performed considerably better than for the CRB and the NRB, with the HI performing more poorly in the CRB. We suggest that a non-parametric index such as the DI that is based on the entire hypsometric curve is a better metric for comparing hypsometric curves compared to a single-valued metric such as the HI when the entirety of the shape of the curve is important to consider, and this is facilitated by a functional representation of empirical hypsometry.

Next, we discuss the use of hypsometric curve in understanding the hydrologic response in watersheds. In hydrologic models, elevation zones are commonly used to account for the dependency between snow-related processes and altitude at basin scale (Leung and Wigmosta, 1999; Valéry et al., 2014; Wayand et al., 2015). For example, the snow accounting routine, CemaNeige (Valéry et al., 2014), uses elevation zones alongside a physical representation of snow accumulation and cold content to accurately model snowpack and snowmelt within a lumped watershed model. The model computes hydrologic response on daily time scales, and requires temperature, precipitation, and historical streamflow data for calibration. Hypsometric information, in the form of 101 elevation points and associated area over the watershed, is needed to account for elevation-wise snow accumulation and melt. Although meteorological and hydrological data are readily available, hypsometric analysis requires the processing of DEM data. The generalized hypsometric function presented in this study can be applied to elevation data to compute watershed-specific parameters that fully specify the needed information for configuring the CemaNeige model for a particular watershed. As shown in this study, the three parameters allow a complete specification of hypsometric function, and in this case can be used to derive a dataset comprised of calculated parameters on regional and national scales (for example, for the stream gauging network managed within the USGS National Water Information System). Thus, this

eliminates an often tedious step of DEM analyses for hypsometry, and allows for the use of the three parameters and the elevation range for the watershed as a pre-computed dataset.

Another potential application area for the elevation-area relationship is in the Geomorphological Instantaneous Unit Hydrograph (GIUH) formulation in hydrologic analysis (Rodríguez-Iturbe and Valdés, 1979; Valdes et al., 1979; Gupta et al., 1980). The GIUH offers a physically-based description of the runoff travel time distribution of a catchment based on channel network characteristics and velocity distribution. Mathematical formulation of link-based GIUH has been developed (Bras, 1990); to be fully operable, functional forms for the channel network topology and velocity distribution are needed. A recent study Bajracharya and Jain (2020) presented a truncated skew-Normal distribution approach to estimate the width function, which is a useful proxy for the channel network topology. The Generalized Hypsometric function readily accommodates a wide variety of hypsometric curve shapes and paves the way toward a fully analytical implementation of the velocity distribution within the link-based GIUH approach.

4.6 Summary

Drainage basin hypsometric curves exhibit significant diversity. A comprehensive review of the functional estimation approaches revealed a reliance on transcendental and polynomial functions (see Table 4.1 for details), and that both categories offer credible estimates with two to seven parameters. The Strahler hypsometric function was the first attempt to represent the area-elevation relationship, and its unique features, up-concavity in the head and the down-concavity in the tail, as well as variations in sinuosity (Strahler, 1952). The Hypsometric integral (HI), another mathematical representation of hypsometry (Horton, 1932; Strahler, 1952), was shown to be consistent with the Strahler formulation. Pike and Wilson (1971) later

showed this to be mathematically identical to a simpler formulation based on mean relative elevation. Several other formulations to describe the hypsometric formulation have been studied in the literature (Harlin, 1978; Sarkar and Patel, 2011; Vanderwaal and Ssegane, 2013; Ssegane et al., 2012). In particular, the polynomial approach presented by Harlin (1984) offers ready estimates of the hypsometric curve and its higher moments (skewness and kurtosis), which provide a further description of the hypsometry beyond HI; this, however, remains limited in its ability to conform to boundary conditions ($x = 0, y = 1$ and $x = 1, y = 0$). A majority of approaches present functions with two or three parameters, and thus are attractive from the standpoint of statistical estimation and interpretation of parameters. The published literature on this topic is limited in efforts to link the estimated parameters with particular features of the hypsometric curve, as well as limited testing of functions in drainage basins with diverse hypsometry. The formulation of Strahler (1952) meets the boundary condition constraints and its two parameters can be broadly linked to the shape of the hypsometric curve. However, the range of variations within hypsometric curves demands higher flexibility, especially towards the head and the toe regions of the curves.

In this study, we presented a new functional formulation, the Generalized Hypsometric function with one additional parameter that greatly expands the range of hypsometric curve shapes while at the same time, accommodating the boundary conditions. The added parameter allows control over the shape of the hypsometric curve, with the parameters r , z , and m primarily influencing the body, toe, and the head of the curve, respectively. Consequently, these parameters reflect the mass distribution and erosional/ depositional processes across different regions of a basin. This function was tested on 419 sub-basins from the Congo, Narmada, and Yukon River basins to assess its adequacy in diverse hypsometries. The Strahler function can be seen as a special case of the Generalized Hypsometric function, with the m

parameter equal to unity. With values highly distributed as low as below 0.1 and as high as above 10 in the CRB, the added parameter, m clearly helps in shaping the curvature of the head, while the low variance around unity in the NRB indicates that the added parameter does not provide much utility in its sub-basins. Most interestingly, m only takes values below unity in the Yukon while r values have a greater variance between zero and one, compared to Strahler fits where r values are mostly closer to zero. With low HIs in the sub-basins, the m parameter controls the concave-up shape of the head, allowing the r parameter to control the overall shape of the body, as opposed to only controlling the concavity of the head in the case of the Strahler function. Furthermore, this combination of parameters suggests that glaciers have eroded high elevation regions leading to concave-up hypsometric curves. Comparative analyses using other functional forms were conducted by first revising model formulations to satisfy boundary conditions. As such, a systematic comparison of the error distributions revealed consistent improvements in fit when the Generalized Hypsometric function was used in the analysis. Furthermore, our analysis revealed that the Generalized Hypsometric function better captured the curvature of the head and the toe compared to other functions in a large majority of sub-basins in this study.

Finally, we provided a few potential avenues for the application of the functional form in hydrologic response modeling and in drainage basin regionalization and similarity assessment. With three parameters along with the elevation range capturing the hypsometric information of a drainage basin, they can be used to derive a dataset on regional and national scales. This eliminates the often tedious step to process DEM data to derive hypsometric information required for various hydrologic modeling processes such as snow accumulation and melt processes based on the elevation distribution of a basin. Furthermore, this can aid in the large-scale assessment of basin similarity. A well-fit hypsometric function can also aid in the

analytical formulation of the hydrologic response component of the GIUH. Consequently, this study can pave the way for future research into the improved characterization of quantitative basin hypsometric information, one that permits efficient reduction of data, is computationally efficient, and can be extended to global scales.

CHAPTER 5

HYDROLOGIC SIMILARITY BASED ON WIDTH FUNCTION AND HYPSOMETRY: AN UNSUPERVISED LEARNING APPROACH

5.1 Chapter Abstract

The prediction of hydrologic conditions in watersheds has manifold applications, ranging from flood disaster preparedness to water supply and environmental flow management. In watersheds with scarce or no flow data, it is difficult to make accurate hydrologic predictions. Past work has used similarity in single-valued properties of the terrain (for example, drainage area, mean slope) as the basis to relate flow conditions in gauged watersheds to the ungauged ones. The resulting predictions show modest accuracy and have a weak physical basis. In this study, we develop a physics-informed machine learning approach to extract features that represent the hydrologic dynamics—width function and hypsometric curve. These two geomorphometric measures are computed using functional forms fitted to estimates derived from digital elevation data. Furthermore, dynamically-similar groups are identified based on results from unsupervised clustering and divergence measures. Our approach paves the way toward a flexible and scalable machine learning approach that can be used to assess hydrologic similarity and improve prediction, one informed by the physics of surface flow generation and transport in watersheds. A case study involving 72 sub-watersheds in the Narmada River Basin (India) is used to illustrate the new methodology.

5.2 Introduction

Flow regimes and hydrologic response in watersheds are tied to the spatial pattern and distribution of a number of biogeophysical variables, including but not

limited to topography, soil, vegetation, and built structures. In watersheds where streamflow data is available, hydrologic models can be readily deployed and calibrated for the purposes of hydrologic prediction. However, in ungauged or data-scarce watersheds, current approaches to flow estimation rely on (a) statistical regression models that use measurable watershed attributes, such as area, slope, and stream length; and (b) comparative hydrology that considers watershed characteristics to assess hydrologic similarity to select analogous gauged watersheds as proxies (Blöschl et al., 2013). Hydrologic similarity defines the similarity in the response of basins to precipitation (Sawicz et al., 2011). This can be useful for the classification of basins and transferability of information. Lumped catchment characteristics and dimensionless indices such as Strahler ratios, drainage density, average slope, etc. are the commonly used metrics to assess hydrological similarity (Horton, 1932, 1945; Strahler, 1957; Bras, 1990; Rodríguez-Iturbe and Rinaldo, 2001). However, the simplification of complex catchment properties into numerical parameters can result in a loss of information that arises from collapsing complex system characteristics into a single number (Wooldridge and Kalma, 2001; Wagener and Wheeler, 2006; Tetzlaff et al., 2009; Chang et al., 2014). Distribution curves such as the topographic index, height above nearest drainage, and reduced dissipation per unit length index (Loritz et al., 2019) often improve the spatial description of river basin characteristics. Newer conceptions regarding hydrologic similarity focus on hydrologic response and therefore emphasize the use of dynamical measures of the stream network and watershed terrain (Bajracharya and Jain, 2020, 2021), thus serving as a rationale for this study. Dynamical measures—width function and hypsometric curve—can be readily estimated using the available global digital terrain datasets. The computational burden, while significant, can be reduced by functional estimation and machine learning approaches (Bajracharya and Jain, 2020, 2021).

In hydrological sciences, machine learning has been used in applications such as precipitation analysis (Sun and Tang, 2020), rainfall-runoff processes (Hsu et al., 1995; Minns and Hall, 1996; Dawson and Wilby, 1998; Abrahart and See, 2000; Duan et al., 2020; Opiel and Mewes, 2020), groundwater hydrology (Karandish and Šimnek, 2016; Sahu et al., 2020), reservoir hydrology (Bai et al., 2016; Mital et al., 2020), hydraulic networks (Dibike et al., 1999), river basin management (Solomatine and Ostfeld, 2008), flow mapping (Zhu and Guo, 2014), land use analysis (Loukika et al., 2021), and disaster risk management (Whitehurst et al., 2021). Explainable artificial intelligence (XAI) is a subdomain of machine learning that aid in the interpretability of machine learning models by helping users understand how their 'black-box' models operate (Maksymiuk et al., 2020; Althoff et al., 2021). Along these lines, applications to the problem of hydrologic prediction in ungauged or data-scarce environments present an attractive opportunity to meld machine learning approaches with the knowledge of watershed dynamics. In this study, we propose an approach that employs unsupervised classification to group similar basins based on the distribution properties of hydrological basins. This allows an efficient organization of terrain data by subsetting it into coherent groups of similar basins based on dynamical measures.

The aim of this study is to demonstrate a simple, yet flexible approach that harnesses statistical and machine learning to leverage known catchment properties that have been shown to act as controls on streamflow response at watershed scales in order to find hydrologically analogous basins, where the paucity of streamflow records precludes hydrological model calibration and analyses. The geomorphological instantaneous unit hydrograph (GIUH) (Gupta and Waymire, 1983; Mesa and Mifflin, 1986; Bras, 1990) has been shown to provide a theoretical basis for runoff prediction in ungauged and data-scarce basins. GIUH treats basin response as a probability density function of basin holding time, computed from the

stream network geometry, and uses a routing function that governs the water transport through the streams (Mesa and Mifflin, 1986). The width function encapsulates the spatial distribution of the network geometry, and as such, serves as a metric for similarity assessment in our study. This is supplemented by the hypsometric curve to incorporate elevational information as a proxy for the velocity component of the GIUH. It is well known that streamflow velocity depends on slope, which provides an intuitive basis for using the hypsometric curve due to its relation to the mean ground slope (Strahler, 1952). In addition, multiple studies have also directly found that hypsometry affects basin hydrology (Harlin, 1984; Howard, 1990; Willgoose and Hancock, 1998; Marani et al., 2001; Vivoni et al., 2008). As components that make up the GIUH, these metrics provide a direct connection to the hydrological response of river basins. Consequently, the clusters based on width functions and hypsometric curves on their own provide hydrological analogues based on unit rainfall response characteristics and elevation distribution, respectively; while a bivariate grouping can provide a synergistic combination of the streamflow path characteristic alongside elevational profiles. Our approach enables a computationally efficient means for finding hydrological analogues based on the hydrological response of basins that can have large-scale applications, including at national and global scales, with minimal expert supervision, while allowing for some flexibility.

In what follows, we first discuss the study area and the dataset used, Next, we briefly review some common approaches to similarity assessment. We then discuss the background information about GIUH and hierarchical clustering, before presenting our methodology. Finally, we illustrate the results of the width function- and the hypsometric function-based clustering based on a case study of the Narmada River basin.

5.3 Study Area and Data

The Narmada River basin (NRB) is located in central India between latitudes $21^{\circ} 22' 0''$ N and $23^{\circ} 46' 30''$ N, and longitudes $73^{\circ} 4' 0''$ E and $81^{\circ} 45' 30''$ E. The drainage area is $95,000 \text{ km}^2$ (Figure 5.1). The elevation ranges from nearly zero to over 1000 m above sea level, with an average slope of 1.1° . The basin is bounded on the north, east, and south by hills, and on the west by the Arabian sea. The lower middle reaches are comprised of fertile plain lands. The Narmada River is the fifth largest river in India. It originates from the Amarkantak plateau in the form of groundwater seepage, flows mostly through the Deccan Traps, and drains into the Arabian sea (Pandey and Khare, 2018; Gupta et al., 2011). On its way, the river passes through hills of the town of Mandla, the Marble Rock Gorge, and several notable waterfalls. Over 70% of the basin is occupied by Deccan basalt, while other considerable lithological features include Quarternary soils, sedimentary outcrops, and thick patches of other rocks (Gupta et al., 2011). A number of reservoirs have been constructed in the basin for a variety of purposes including water supply, irrigation, and hydropower generation. The Narmada River passes through three states that face water shortages during non-monsoon seasons (Ray and Goel, 2019).

The elevation data for the region was obtained from GTOPO30, a global digital elevation model (DEM) developed by the United States Geological Survey (USGS). It was derived from several raster and vector sources of topographic information (USGS, 1996). The dataset has a spatial resolution of 30-arc seconds and a vertical accuracy of around 30 m. It is based on several sources of elevation information, including various vector and raster datasets, merged together, with a priority given to the data with greater topographic detail and accuracy. With extensive accuracy checks, GTOPO30 data are suitable for numerous regional and continental applications, including the extraction of drainage features for hydrologic modeling (USGS, 1996). The width function and hypsometric curve have been shown to be

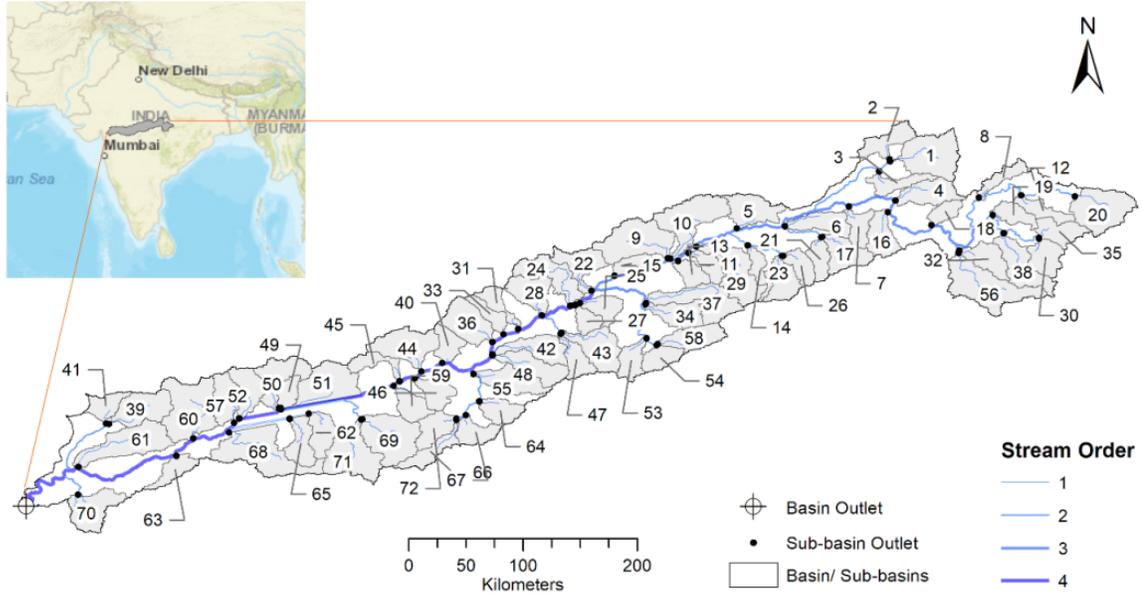


Figure 5.1: Map of Narmada river basin and its location. Delineated sub-basins are shown along with their identifier ids.

less sensitive to DEM resolution by some studies (Mathur, 2018; Nazari-Sharabian et al., 2020), and as such, the GTOPO30 data set provides an adequate level of accuracy for this demonstrative analysis. However, other studies have shown that DEM resolution can affect the accuracy of width function and hypsometric properties (Liffner et al., 2018). The method demonstrated in this study should work with higher resolution terrain data and provide some marginal improvements at the minimum, depending on the purpose of the study.

The stream network was derived from the DEM in ESRI ArcGIS 10.5.1 through standard Geographic Information System (GIS) procedures. First, elevation grids with undefined drainage directions, known as sinks, were filled; then the flow direction was determined based on the direction of steepest descent; followed by the computation of accumulated flow at each grid. A threshold of 396 km^2 was used to delineate the stream grids. This threshold was chosen to ensure a dense stream network, resulting in fourth-order streams. This allowed for the delineation of a considerable number of sub-basins to test the fits for diverse width function and

hypsothetic curve shapes. Finally, outlets were placed at the confluences of the first order and higher-order streams to create 72 non-overlapping sub-basins.

5.4 Methods

5.4.1 Hydrological Similarity

Comparative hydrology is an approach to the prediction in ungauged basins (PUB) that examines a large number of catchments to distinguish patterns of hydrological behavior using common catchment and climatic characteristics. While there is no universal basis for the hydrological classification of catchments (Blöschl et al., 2013), they are self-organizing systems whose hydraulic behavior results from adaptive geomorphological processes (Sivapalan, 2006) and there are discernible patterns that form the foundations for understanding their hydrological nature. In general, catchments can be considered hydrologically similar if they have similar responses to climatic variability (Blöschl et al., 2013). Proximity is a commonly used, reliable metric for determining similar catchments, however, this measure is limited in that it does not allow for the use of catchments that are not closer to each other (Patil and Stieglitz, 2012). Since climate strongly impacts catchment characteristics and hydrological behavior, the hydro-climatic region where a catchment is located provides another basis for catchment classification (Budyko et al., 1974; L'vovich, 1979; Abrahams, 1984; Milly, 1994; Sankarasubramanian and Vogel, 2002; Woods, 2006; Yadav et al., 2007). Similarly, readily observable spatial patterns in the catchment structure that affect the temporal response can be used as signatures to determine possible co-evolution of basin dynamics (Blöschl et al., 2013), and can be utilized to transfer hydrological information from data-rich catchments to ungauged basins to predict physical phenomena such as hydrologic response (Burn and Boorman, 1993; Tung et al., 1997; Aryal et al., 2002; McIntyre et al., 2005; Wagener et al., 2007; Reichl et al., 2009; Archfield and Vogel, 2010;

Oudin et al., 2010; Patil and Stieglitz, 2011, 2012; Razavi and Coulibaly, 2013; Athira et al., 2016; Brunner et al., 2018). A commonly used technique involves the transfer of lumped characteristics such as catchment shape and size, Strahler ratios, drainage density, average slope, etc. that are used to explain hydrogeomorphological characteristics (Horton, 1932, 1945; Strahler, 1957; Bras, 1990; Rodríguez-Iturbe and Rinaldo, 2001). An issue with this is the possibility of the loss of information in simplifying complex catchment properties into a single number (Wooldridge and Kalma, 2001; Wagener and Wheeler, 2006; Tetzlaff et al., 2009; Chang et al., 2014). Alternatively, distribution curves, having been closely linked to runoff generation and flow processes, can be used to assess hydrological similarity. Examples of this include the use of the distribution of topographic index, height above nearest drainage, reduced dissipation per unit length index (Loritz et al., 2019), the distribution of riparian and hillslope effects on streams, the riparian-area change along the stream network (McGlynn and Seibert, 2003), the hypsometric curve (Booij et al., 2007; Ssegane et al., 2012; Hailegeorgis et al., 2015; Bajracharya and Jain, 2021), and the width function (Moussa, 2008; Bajracharya and Jain, 2020). Furthermore, various mathematical models that link catchment structure to hydrological response based on underlying physics or statistical relationships have been used to explore catchment similarity and to develop similarity parameters (Hebson and Wood, 1982; Sivapalan et al., 1987; Larsen et al., 1994; Milly, 1994; Reggiani et al., 2000; Aryal et al., 2002; Woods, 2003).

5.4.2 Dynamical Representation of Watershed Morphometry

5.4.2.1 Width Function

The width function represents the travel distance distribution of a stream network (Gupta et al., 1986; Bras, 1990). For a given drainage basin, the width function, $N(x)$, denotes the areal extent between x and $x + dx$, where x represents

the total distance along the flow path to the outlet (Veneziano et al., 2000), termed here as the hydrological distance. As we do not distinguish between the hillslope and channel network distance in this study, the width function becomes synonymous with the area function. Under the assumption of constant velocity, the width function represents the probability distribution of travel times or the instantaneous unit hydrograph, reflecting the topological features of a basin’s stream response (Lashermes and Foufoula-Georgiou, 2007; Moussa, 2008). The width function is strongly linked to the peak and shape of the hydrograph (Kirkby, 1976; Gupta and Waymire, 1983; Troutman and Karlinger, 1984, 1989). As the width function can be obtained from terrain elevation data and captures important characteristics of the basin’s hydrologic response, it is a powerful tool for prediction in ungauged basins (Moussa, 2008).

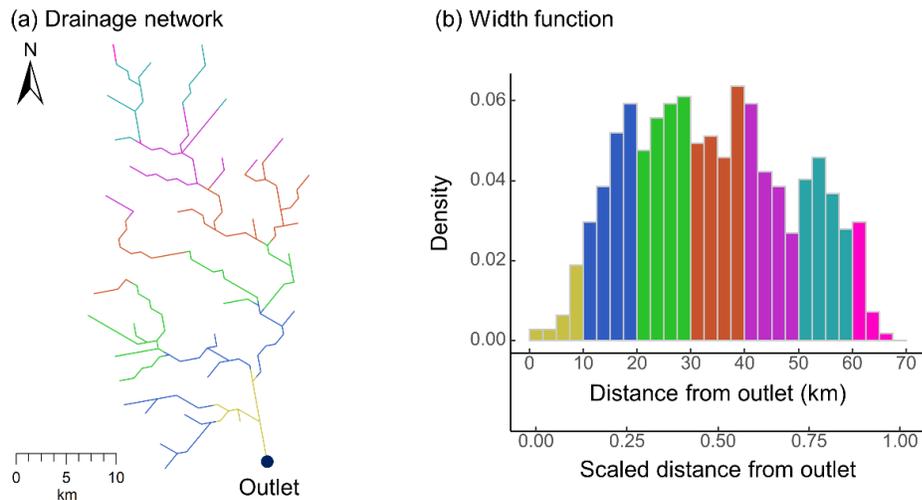


Figure 5.2: Schematic illustrating the definition of width function. (a) Stream drainage network of sub-basin 57, with color gradations based on flow path distances to the outlet denoting hydrological distances, and (b) width function with corresponding color gradation. The upper x -axis shows the hydrological distance in absolute units (km), while the lower x -axis presents the corresponding scaled hydrological distance. The binwidth is 2.5 km.

5.4.2.2 Hypsometric Curve

The hypsometric curve reflects the area-altitude distribution of a basin (Horton, 1932; Langbein, 1947) and reflects the distribution of landmass as a function of elevation (Harlin, 1984). The shape of the hypsometric curve indicates the geomorphic maturity of catchments, with a concave-up shape indicating relatively mature basins with a high degree of erosive activity, and a concave-down shape indicating relatively young basins with a large proportion of uneroded topography or creep-dominated hillslopes (Strahler, 1952; Moglen and Bras, 1995; Pedrera et al., 2009; Willgoose, 2018; Guha and Jain, 2020). Furthermore, the hypsometric curve can be related to the mean ground slope accounting for the length of the contour belt (Strahler, 1952) and studies have linked the hypsometric curve with various drainage basin features such as the hydrograph time-to-peak, head-ward drainage development, regional basin slopes (Harlin, 1984), average channel gradient (Howard, 1990), stream network branching (Willgoose and Hancock, 1998), groundwater interaction, water table fluctuation (Marani et al., 2001), watershed management (Sreedevi et al., 2005), and surface and subsurface runoff properties (Vivoni et al., 2008). Willgoose and Hancock (1998) further divided the hypsometric curve into three regions: the 'head' (upper left-hand side), the 'toe' (lower right-hand side), and the 'body' (between the head and the toe), and linked the shape of the toe to stream branching characteristics of the basin. Furthermore, hillslopes with active fluvial erosion and creep exhibit concave down head and concave up tail (Willgoose, 2018). In long skinny catchments and hillslopes with parallel flow lines, the hypsometric curve reflects the hillslope long profile and can be used as an indirect test of the slope-area relationship, while in more rounded catchments, the stream network branching also affects the shape of the hypsometric curve (Willgoose, 2018). The hypsometric integral (Strahler, 1952; Pike and Wilson, 1971), defined as the area under the hypsometric curve, is often used to represent

the hypsometry of basins; however it has been shown to have a limited correlation with flow statistics (Liyanagamage and Hewa, 2012b). Harlin (1978) argued that the hypsometric curve has a greater potential for quantitative landform analysis relative to hypsometric integral alone. A summary of the existing literature linking the importance of hypsometry in hydrological response is provided in Table 5.1.

The hypsometric curve can be plotted in absolute units, with elevation in meters and area in square kilometers, or in relative units, with relative elevation plotted against the relative area above said elevation (Figure 5.3). The latter, termed as the percentage hypsometric curve, allows for the comparison of basins of different altitudes and sizes (Strahler, 1952).

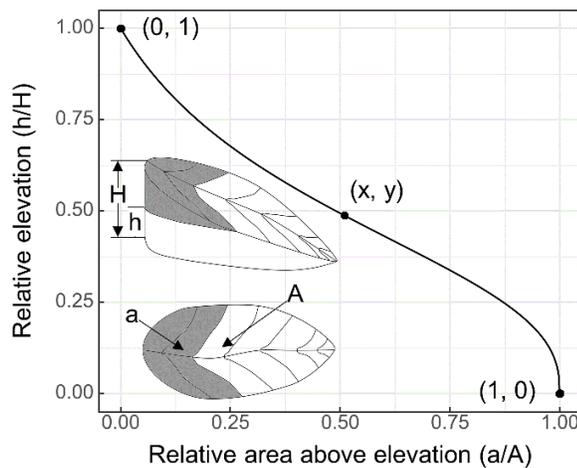


Figure 5.3: The scaled hypsometric curve showing the relative elevation plotted relative to the proportion of area above this elevation.

5.4.2.3 Geomorphological Instantaneous Unit Hydrograph

The geomorphological instantaneous unit hydrograph (GIUH), introduced by Rodríguez-Iturbe and Valdés (1979) and Gupta et al. (1980), uses the morphometric structure of the river basin to compute the unit hydrograph at the river basin outlet. This formulation is based on the probability distribution of travel times of excess rainfall through their natural travel paths across the basin and into the

Table 5.1: Influence of hypsometry on hydrological response.

Reference	Significance
Harlin (1984)	<ul style="list-style-type: none"> • Correlation between the watershed time-to-peak and hypsometric function parameters, • Relation between moments of the hypsometric function and the headward drainage development, mid-basin slope, and rate of change in upper and lower basin slope, and • Potential for increase in the predictive power of rainfall-runoff models through the use of hypsometry alongside basin area and/ or relief.
Howard (1990)	<ul style="list-style-type: none"> • Basin size and average channel gradients have predictive power in estimating basin hydrological response.
Willgoose and Hancock (1998)	<ul style="list-style-type: none"> • Hypsometry is a predictive indicator of the planar form of the drainage network.
Marani et al. (2001)	<ul style="list-style-type: none"> • Basin elevation distribution is linked to groundwater interactions and water table fluctuations.
Vivoni et al. (2008)	<ul style="list-style-type: none"> • Hypsometry relates to the surface and subsurface runoff components of the basin, and • Hypsometry is related to the spatial distribution of soil moisture and runoff response. mechanisms

outlet. Gupta et al. (1986) developed the link-based GIUH that utilizes width function as the probability densities of the hydraulic response of channels:

$$h(t) = \int_0^{x_{max}} g(x, t)N(x)Z^{-1}dx \quad (5.1)$$

where $h(t)$ denotes the instantaneous unit hydrograph as a function of time, t ; $g(x, t)$ denotes the hydraulic response function at a distance, x , from the outlet of the basin; and the term $N(x)Z^{-1}$ represents a non-dimensionalized width function with the hydrological distance, x scaled between 0 and 1. Figure 5.4 outlines a schematic for the conceptual basis for the two components of the link-based GIUH

formulation and its utility in the prediction of the hydrological response of the basin with rainfall data. The width function component provides the travel times across the basin at constant velocities, while the hydraulic response component provides the spatial variation of velocities.

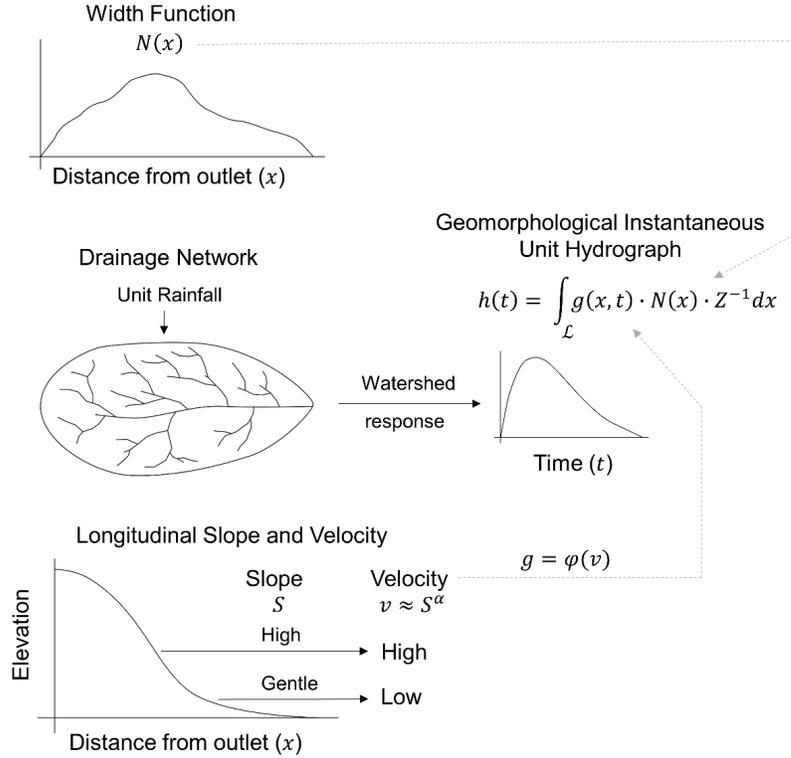


Figure 5.4: Components of the link-based Geomorphological Instantaneous Unit Hydrograph (GIUH). The scaled width function ($N(x) \cdot Z^{-1}$, where Z is the area under the width function) component provides the distribution of the flow distance, while the hydraulic response component, $h(x)$, accounts for the streamflow velocity along the flow paths.

As an example, when the hydraulic response function is a simple translation, it can be computed as:

$$g(x, t) = \delta \left(t - \frac{x}{V} \right) \quad (5.2)$$

where δ is the Dirac delta function and V is the velocity of translation (Bras, 1990). While the width function can be obtained from the terrain information, the determination of the hydraulic response component can be more challenging, as the

streamflow velocity across the basin is a non-measurable parameter (Al-Wagdany and Rao, 1997). Streamflow velocities depend on multiple factors such as surface roughness and slope. Various geomorphological formulations of the velocity parameter have been developed (Al-Wagdany and Rao, 1997; Chen et al., 2019). Chen et al. (2019) developed the formulations for the time of flow concentration and parameter velocity based on the length of the main stream and the mean catchment slope. In this study, we illustrate a simpler proxy for flow velocity based on the elevation-area relationship for watersheds. The hypsometric curve can be linked to the mean ground slope (Strahler, 1952), and the dependence of streamflow velocity on channel gradient is well understood. Furthermore, Luo and Harlin (2003) illustrated how the hypsometric function can be directly used to infer the time of concentration and travel times. As such, we posit the use of hypsometric function as a proxy for the hydraulic component of the link-based GIUH alongside the width function component as metrics for similarity assessment of hydrodynamic response.

5.4.3 Functional Estimation

5.4.3.1 Width Function

The width function is most commonly represented by a histogram with the hydrological distance in the x -axis and the frequency or density of the areal extent of streams in the y -axis (Figure 5.2). Bajracharya and Jain (2020) demonstrated the use of a truncated skew-Normal (SN) mixture model to analytically represent the width function with the x -axis normalized by scaling between 0 and 1, and demonstrated its utility in finding hydrologically analogous drainage basins using divergence measures such as the L_2 distance (Tsybakov, 2008). The SN distribution is a three-parameter probability distribution formed by adding a skewness element to the Normal distribution.

For a continuous random variable, X , the SN distribution is represented as:

$$f(x; \xi, \omega^2, \alpha) = \frac{2}{\omega} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left(\alpha \frac{x - \xi}{\omega}\right), x \in (-\infty, \infty) \quad (5.3)$$

where $\phi(x)$ denotes the standard Normal density function of x , $\Phi(x)$ denotes the cumulative distribution function (*cdf*) of the standard Normal, and ξ , α , and ω are the location, scale, and shape parameters, respectively. The domain of the SN distribution is then truncated to $[0, 1]$ using a correcting factor to guarantee the validity of the normalization condition (Thomopoulos, 2017):

$$g(x) = \begin{cases} \frac{f(x)}{F(1)-F(0)} & , x \in [0, 1] \\ 0 & , x \in (-\infty, 0) \cup (1, \infty) \end{cases} \quad (5.4)$$

where $F(x)$ denotes the cumulative density function. Finally, a finite mixture model of n truncated SN distributions is represented as:

$$h(x) = \sum_{i=1}^n w_i g(x; \xi_i, \omega_i^2, \alpha_i) \quad (5.5)$$

where w_i denote the non-negative mixing proportions that sum to one.

Furthermore, the L_2 distance used by Bajracharya and Jain (2020) to measure similarity between two width functions is computed as:

$$L_2 = \sqrt{\int (N_1 - N_2)^2 dx} \quad (5.6)$$

where N_1 and N_2 represent the two width functions under consideration. A value of zero indicates identical width functions, while greater values reflect a larger difference. Additional choices of divergence measures are listed in Bajracharya and Jain (2020).

5.4.3.2 Hypsometric Curve

Various functional forms have been developed to represent the hypsometric curve (Strahler, 1952; Harlin, 1978; Sarkar and Patel, 2011; Vanderwaal and Ssegane, 2013; Bajracharya and Jain, 2021). Bajracharya and Jain (2021) developed a three-parameter model named the Generalized Hypsometric function by modifying the equation developed by Strahler (1952). The model places an emphasis on the curvatures of the head, body, and toe. The function is defined as:

$$y = \left(\frac{1 - x^m}{1 + \beta x^m} \right)^z \quad (5.7)$$

where β , z , and m denote the three parameters. Furthermore, Bajracharya and Jain (2021) illustrated the use of hypsometry to find analogous basins using the discordance index (DI), defined as the total absolute area between two hypsometric curves. The DI between two hypsometric functions, $f_1(x)$ and $f_2(x)$ is defined by:

$$DI = \int_0^1 |f_1(x) - f_2(x)| dx \quad (5.8)$$

A value of zero indicates identical hypsometric functions whereas greater values indicate larger deviations.

5.4.4 Hierarchical Clustering

Clustering is a descriptive unsupervised data mining technique for creating subsets by grouping similar data together based on some measure of similarity or dissimilarity (Veyssieres and Plant, 1998; Rokach and Maimon, 2005). The clustering structure is represented by a set of subsets, $C = C_1, \dots, C_k$ of S , such that $S = \bigcup_{i=1}^k C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$. Hierarchical clustering is a clustering method that creates clusters by recursive partitioning, resulting in a dendrogram structure that represents the nested grouping of instances and similarity levels at which the groupings change. The recursive algorithm could be bottom-up, starting

from every element in their individual cluster, with similar elements then grouped into a single cluster in each successive step (agglomerative clustering); or top-down, starting from all elements grouped in a single cluster, followed by the most dissimilar elements being separated into another cluster at each iteration (divisive hierarchical clustering). Various methods have been developed based on the manner in which the similarity measure is calculated and optimized, most of which are variants of single-link, complete-link, and minimum-variance algorithms (Jain et al., 1999). These algorithms consider the distance between two clusters to be equal to the shortest, longest, and average distance between a member of one cluster to a member of the other, respectively. Single-link methods are more versatile (Rokach and Maimon, 2005) but are susceptible to the "chaining effect", where a few points form a bridge between two poorly separated, but distinct clusters lead to them being merged at an early stage (Guha et al., 1998). On the other hand, complete-link methods usually produce more compact clusters (Rokach and Maimon, 2005). Finally, average-link clusters may cause the splitting of elongated clusters and the merging of portions of neighboring elongated clusters (Guha et al., 1998).

In this study, we used the "agnes" function (Kaufman and Rousseeuw, 2009) from the "cluster" package (Maechler et al., 2021) in *R* programming language (R Core Team, 2019) for the clustering analysis. This function provides the agglomerative coefficient (ac) which measures the amount of clustering structure. For a set of observations, ac is the average of $1 - m(i)$, where $m(i)$ is the ratio of dissimilarity of each observation, i , to the first cluster, it is merged with the dissimilarity of the final merger of the algorithm. ac varies between zero and one, with larger values indicating more balanced clustering structures and values closer to zero indicating less well-formed structures. For the given dataset, the Ward method (Ward, 1963), a type of minimum-variance algorithm, was found to have a better ac value compared to the other methods.

The width functions and hypsometric curves were first transformed to their functional forms to facilitate the efficient computation of dissimilarity matrices (Figure 5.5). Width function clustering was done with the fitted SN functions, using the L_2 distance as the dissimilarity measure. This leads to width function analogues that share similarities in hydrological responses based on stream network structures. Similarly, hypsometric clustering was done with the fitted Generalized Hypsometric functions, using the DI as the dissimilarity measure. These clusters are likely to share common hypsometric signatures in terms of erosional/depositional properties. While hypsometric curves are more closely related to the erosional status of the basin, studies have indicated links between hypsometric curves and hydrodynamic properties of basins (Harlin, 1984; Willgoose and Hancock, 1998; Marani et al., 2001; Vivoni et al., 2008; Sridhar et al., 2018) due to the topographic controls on stream generation and flow.

The gap statistic was used to determine the optimal number of clusters (Tibshirani et al., 2001). For a dataset of size n , with k clusters based on a distance measure, the gap statistic is defined as

$$Gap_n(k) = E_n^*[log(W_k)] - log(W_k) \quad (5.9)$$

where E_n^* represents the expected value for a sample size of n from the reference distribution and W_k is the pooled within-cluster sum of squares around the cluster means, defined as $W_k = \sum_{r=1}^k \frac{1}{2n_r} \sum D_r$. Here, D_r represents the sum of pairwise distances for all points in a given cluster, r . This statistic measures the deviation of the observed W_k from its expected value under the null hypothesis. The optimal number of clusters, \hat{k} , can be chosen based on various algorithms, including the global maximum method, which maximizes $Gap_n(k)$, signifying the farthest deviation from uniform points distribution. Due to the lack of clear group demarcations in both width function and hypsometric curve shapes, we chose \hat{k}

based on local maxima of the derivative of the gap function - where the increase in $Gap_n(k)$ first tails off. There is a level of subjectivity in the choice of the number of clusters, with more groups leading to more homogeneity within the group members but a smaller number of members per group. While more robust measures for the choice of the number of clusters are available in the literature (Jung et al., 2003; Zhou et al., 2016), in this study, we have opted for a simpler method that provides a greater degree of flexibility to the user to allow for variations in the choice based on the kind of study based on intuition and expertise.

We also demonstrated the process of outlier detection to reduce intra-cluster variance with a simple algorithm based on similarity measures with the nearest neighbors. We used a minimum threshold approach where members exceeding a minimum similarity index with a selected number of nearest neighbors were classified as outliers and removed from the study. However, care was taken not to omit members with important and distinct physical characteristics. Finally, the sub-basins with common width function clusters and hypsometric function clusters were identified.

5.5 Results

5.5.1 Width Function Clusters

5.5.1.1 Hierarchical Clustering

First, the optimal number of clusters was determined using the gap statistic. Figure 5.6 shows the gap statistic as a function of the number of clusters (k). The graph shows that a larger number of clusters results in a higher gap statistic, and consequently, a better clustering. The continued increase in gap statistic with an increasing number of clusters indicates that the different cluster regions are not sharply delineated. However, a large number of clusters impedes the interpretability of the width function shapes in each cluster. As such, the choice of optimal k

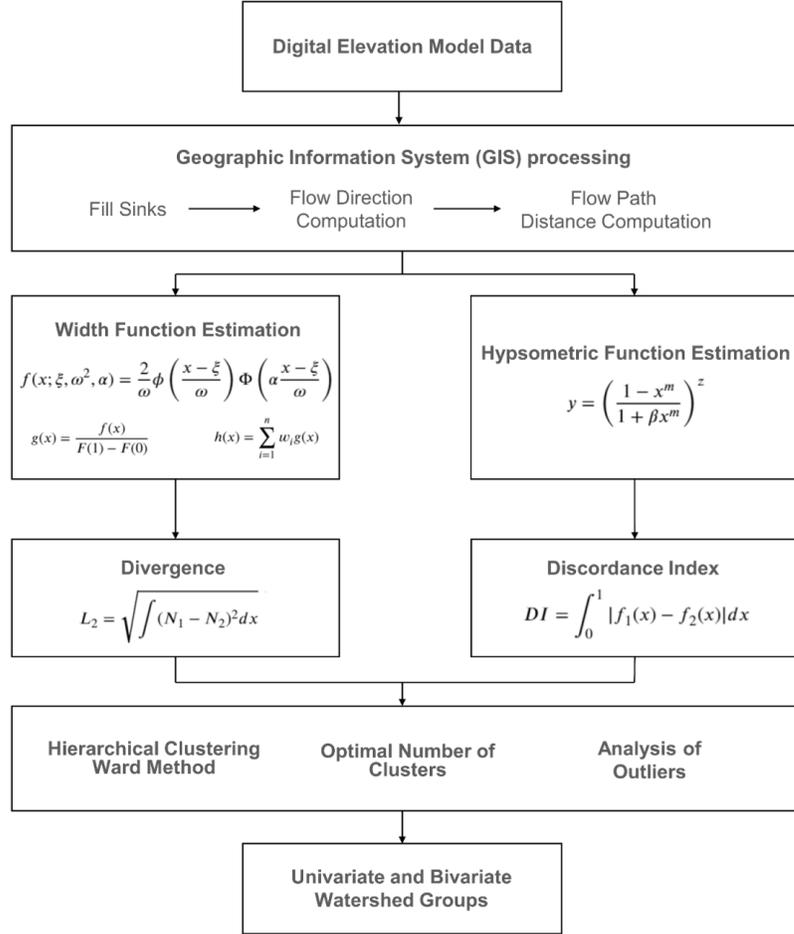


Figure 5.5: Flowchart of the study methodology. Truncated skew-Normal mixture model for estimating the width function, the Generalized Hypsometric function, L_2 divergence measure for width function-based distance matrix, and discordance index (DI) for hypsometric function-based distance matrix have been shown.

involves some subjectivity. We based the choice on where the rate of increase in the gap statistic first sharply decreases. The change in the gap statistic has a sharp decrease when $k > 6$, and as such, the optimal number of clusters for the width functions was chosen as six. The width functions in each cluster are shown in Figure C.1. While there are some considerable variances in the width function shapes within each cluster, different clusters do exhibit noticeably different overall shapes.

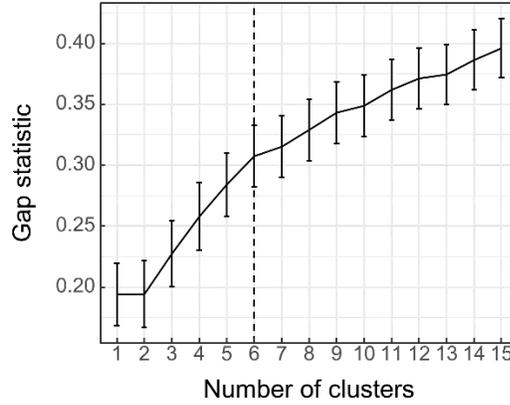


Figure 5.6: Determination of the optimal number of width function clusters using gap statistic. Gap statistics is generated through bootstrapping. The standard error of the simulated gap values is shown as whiskers. The optimal number of clusters was chosen based on the change in the rate of increment of the gap statistic.

5.5.1.2 Analysis of Outliers

Outliers can cause chaining effects, leading to dissimilar objects being drawn into the same cluster (Everitt et al., 2011). Removal of outliers can help reduce intra-cluster variance. However, different outlier detection algorithms can lead to different data points being classified as outliers. Moreover, outlier detection can mistakenly classify small clusters as outliers and remove valuable information from the data. Thus, outlier detection involves a degree of subjectivity. Here we use a simple algorithm to analyze, detect, and remove outliers based on similarity measures with nearest neighbors. Figure 5.7 shows the L_2 distance to fifteen closest

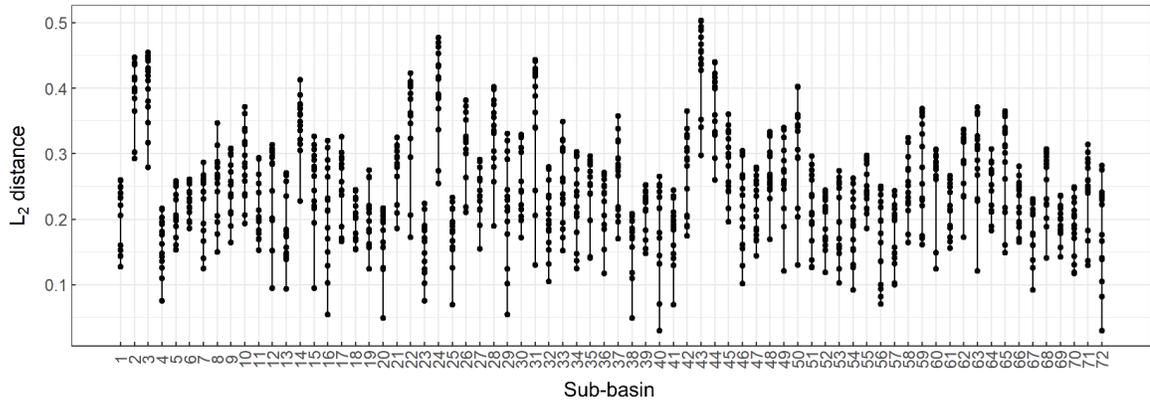


Figure 5.7: L_2 distances between 15 closest neighbors for each width function.

neighbors for each width function. Based on this measure, a threshold can be chosen subjectively to delineate outliers based on specific goals. In this study, width functions with the L_2 distance greater than 0.45 for up to 15 closest neighbors were marked as outliers. This led to only three width functions being classified as outliers. Intra-cluster uniformity can be further improved by lowering this threshold. While rigorous methods for the removal of outliers exist in the literature (Almeida et al., 2007; Fan et al., 2013; Krleža et al., 2021), we employed this basic outlier detection algorithm as a proof of concept, one that is easy to understand and can be readily applied.

5.5.1.3 Analysis of Clusters

After the removal of the outliers, the width functions were reclassified into six clusters (Figure 5.8). With the removal of only three outliers, there are minimal improvements in intra-cluster uniformity, as seen by the removal of two notable outliers in cluster 3. To closely examine the properties of each cluster group, representative width functions in each cluster have been highlighted in Figure 5.8. Representative width functions were chosen based on the lowest L_2 distances with the mean width functions within each cluster, i.e. the width function that is most closely matched with the mean width function. Mean width functions were calculated by averaging y values between all members of a given cluster at each x value. Cluster 5 has a slightly higher peak in the first SN component, while all other clusters have higher peaks in the second SN component, which could indicate a difference in hydrograph peak locations. Among them, cluster 3 does not have a prominent peak, whereas cluster 6 has a prominent peak towards the right end of the width function. Furthermore, the shape of the left rising side and the right falling side of the curves differ between clusters. For instance, the right side of the curves for clusters 2 and 6 are steeper compared to other clusters. It should be

noted that while the overall shapes of the curves are similar within clusters, there is still a considerable degree of heterogeneity in the size and location of the peaks.

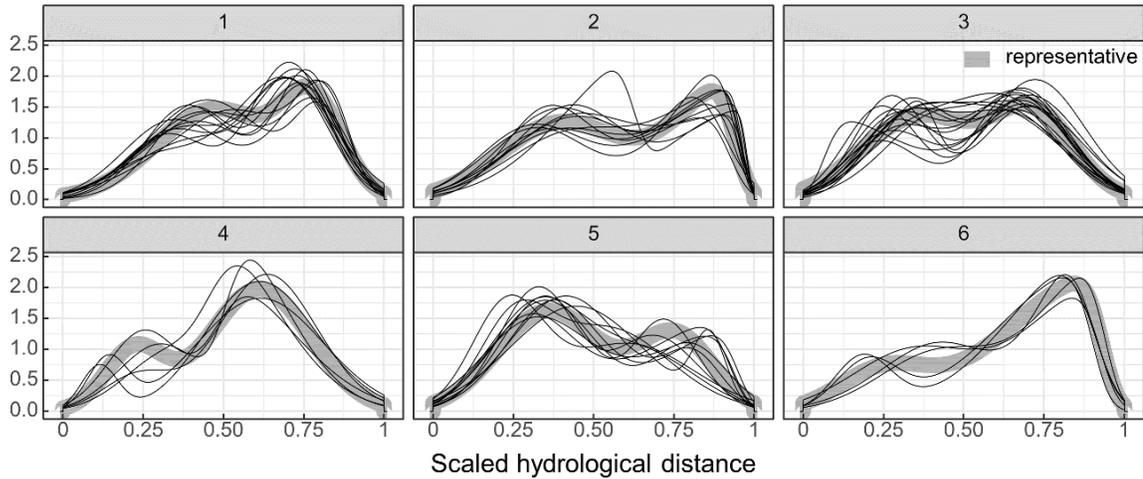


Figure 5.8: Width functions in each cluster after removing the outliers. The representative width functions for each cluster are shown as thick grey lines.

Hierarchical clustering can be best denoted using dendrograms. The dendrogram notation of the width function clusters is shown in Figure 5.9, along with the mean width functions and the location of the sub-basins. Figure 5.9 (b) further highlights the diversity in the shape of the width functions in each cluster. The width function shapes seem mostly independent of the location of the sub-basin along the watershed as well as the sub-basin areas (Figure 5.9 (c)). Cluster 4 seems to be concentrated in the mid-region of the watershed and cluster 6 seems to be concentrated mostly in the bottom half, whereas all other clusters are spread across different regions. Interestingly, a number of sub-basins within the same cluster groups appear alongside each other.

5.5.2 Hypsometric Function Clusters

5.5.2.1 Hierarchical Clustering

Similar to width functions, the gap statistic was used to determine the optimal number of hypsometric function clusters (k) by evaluating the change in gap

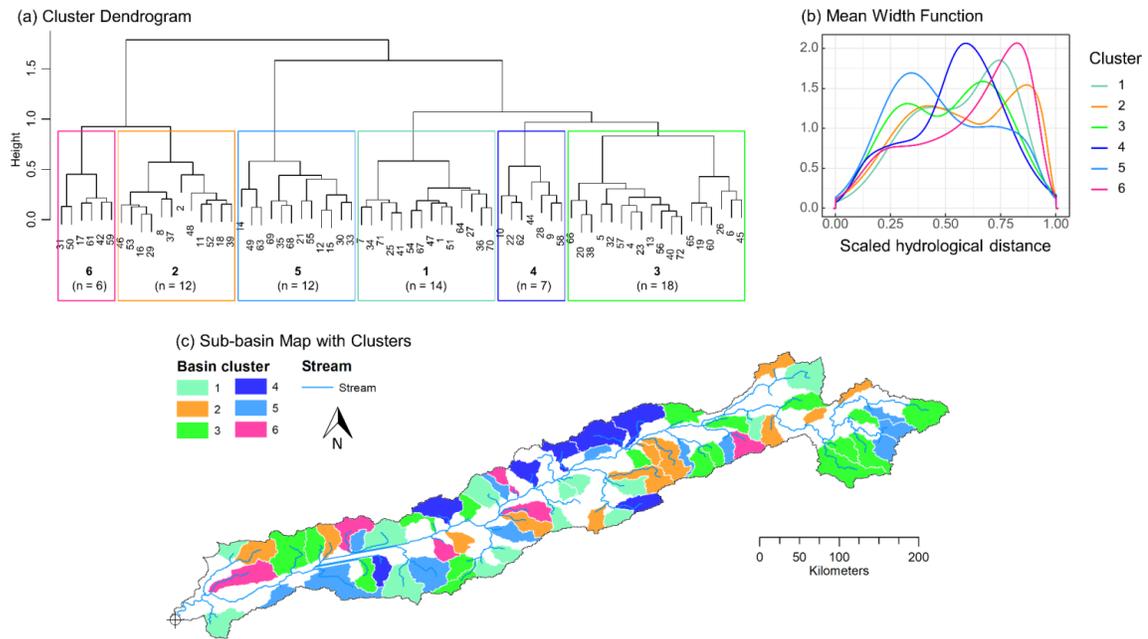


Figure 5.9: (a) Dendrogram of watershed width functions using hierarchical clustering using Ward’s method. (b) Mean width functions for each cluster. (c) Map of sub-basins grouped by cluster.

statistic with the change in k (Figure 5.10). The change in the gap statistic sharply decreases when $k > 8$, and hence, the optimal number of clusters was chosen as eight. It should be noted that when there is a slight decrease in the gap statistic at $k > 2$. However, based on the diversity of the hypsometric functions as seen in Figure 5.11, only two groups would not be adequate. The classified hypsometric functions along with the corresponding representative curves are shown in Figure 5.11. There is a clear distinction in the shapes of the hypsometric curves in each cluster. Clusters 1, 2, 3, 6, and 8 comprise of concave up shapes, while cluster 7 comprises of concave down shapes. The similarity in hypsometric curves could indicate similarity in geomorphological characteristics within the clusters.

Furthermore, hypsometric curves in clusters 4, 5, and 8 have prominent tail regions following inflections in the curve, whereas other clusters lack prominent tail regions. However, it is notable that there is some appreciable variability in the head and toe curvatures within each group. For instance, while the majority of curves in cluster 2

do not have an inflection point near the tail, there are a few curves with prominent tail regions. On the other hand, some hypsometric curves with no prominent tail curvatures are classified into clusters 4 and 5.

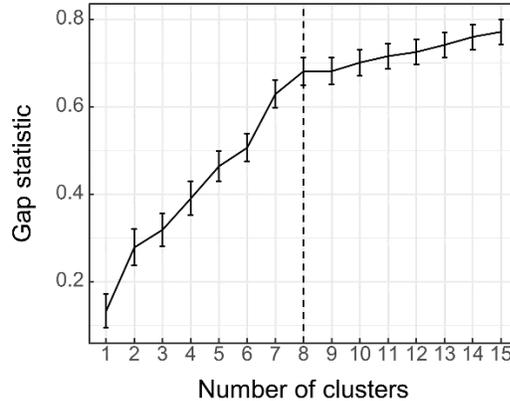


Figure 5.10: Determination of the optimal number of hypsometric function clusters using gap statistic. Gap statistics is generated through bootstrapping. The standard error of the simulated gap values is shown as whiskers. The optimal number of clusters was chosen based on the change in the rate of increment of the gap statistic.

A similar outlier analysis algorithm was applied to these clusters, with a DI of 0.65 chosen as the threshold. However, this led to both sub-basins in cluster 7 being classified as outliers. While this is computationally valid, cluster 7 is the only cluster comprised of concave down curves. Consequently, this group carries an important geomorphological distinction as compared to other clusters, and as such, should not be classified as an outlier or be removed from the study. This indicates a shortcoming of the earlier outlier analysis algorithm, and indicates that a degree of subjective choice may be necessary for the outlier analysis so as to not omit important cluster groups. However, lowering the number of nearest neighbors being considered to just one leads to no member being classified as outliers. This matches visual inspection since the intra-cluster variance in each group is already low. As a result of this, no outlier was removed.

Cluster dendrograms are shown in Figure 5.12 (a), along with the mean hypsometric curves for each cluster group (Figure 5.12 (b)), and their locations

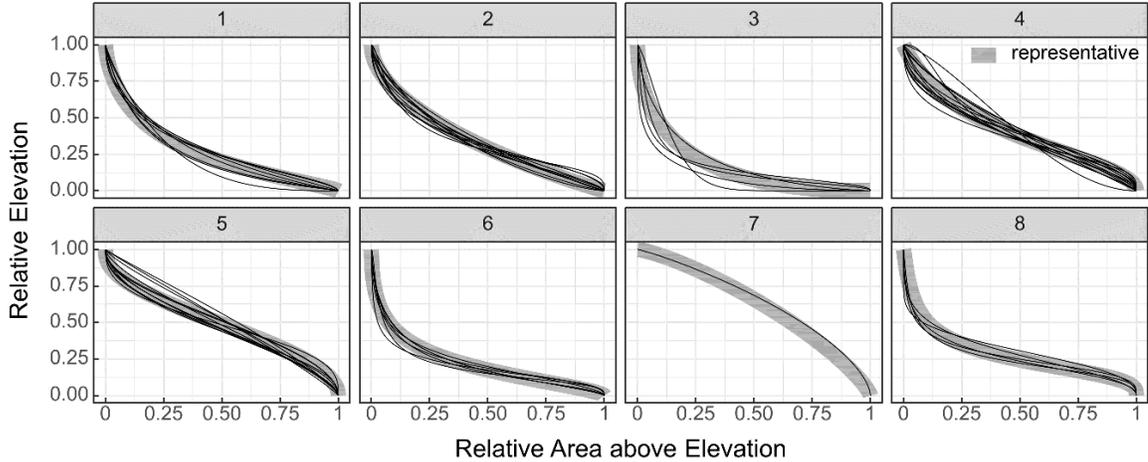


Figure 5.11: Basin hypsometric curves. Corresponding representative curves for each cluster are shown as thick grey lines.

(Figure 5.12 (c)). Mean hypsometric curves are computed by averaging the relative elevations of each cluster member along the relative areas above the elevations. The mean hypsometric curves indicate a gradual change from concave up to concave down shapes along the clusters. There is no clear relationship between the hypsometric curve shapes and the locations of the sub-basins along the watershed or the size of the sub-basins. Sub-basins in cluster 4 are concentrated in the lower half of the watershed, while those in cluster 3 are concentrated in the upper half. Notably, as in the case of width function clusters, numerous sub-basins that fall within the same hypsometric function cluster groups appear adjacent to each other.

5.5.3 Joint Analysis of Hierarchical Clustering of Width Functions and Hypsometric Curves

Next, we discuss the potential to combine the width function and the hypsometric clustering to represent watershed analogs that take into account both, the planar stream network geometry as well as the elevational characteristics of the basin (Figure 5.13). This provides a framework for bivariate clustering that incorporates multiple metrics that supplement each other. For instance, sub-basins 14, 21, 30, 33, 35, and 63 fall in hypsometric cluster 4 and width function cluster 5,

with these members indicating mildly mature hypsometry and width functions with the peak considerably skewed to the right. As such, these sub-basins could potentially be analogues with similar hydrological response properties. Sub-basins 17 and 53 have concave-down hypsometric curves (hypsometric cluster 7), but have considerably different width function shapes (width function clusters 2 and 5), indicating that the hydrological response behavior of these two sub-basins might be considerably different. As such, width function and hypsometry can provide complementary properties, which results in a fuller description of basin processes. In Figure 5.13, we explore the spatial relationships between members in the bivariate groups. Group 5-2, with an early width function peak and a relatively linear hypsometric curve, is predominantly formed at the upstream region of the watershed. Group 4-5, with a highly steep falling limb of the width function and a relatively linear hypsometric curve, exhibited relatively smaller accumulation areas. However, in general, the spatial relationship within the highlighted bivariate groups is slightly weaker compared to individual clusters discussed earlier. Proximity- and regression-based regionalization analyses emphasize contiguity and closeness as means for hydrological similarity because closer basins tend to have similar characteristics such as average slope, climatic conditions, etc. However, our study shows that the use of dynamic measures can provide similar analogues that are not necessarily proximate, yet share properties that directly influence hydrological response.

5.6 Discussion

New understandings and a sound physical basis for the prediction in ungauged and data-scarce river basins are of great theoretical and practical importance. To that end, this study provides an additional approach through the use of unsupervised learning and functional data reduction to derive dynamical measures

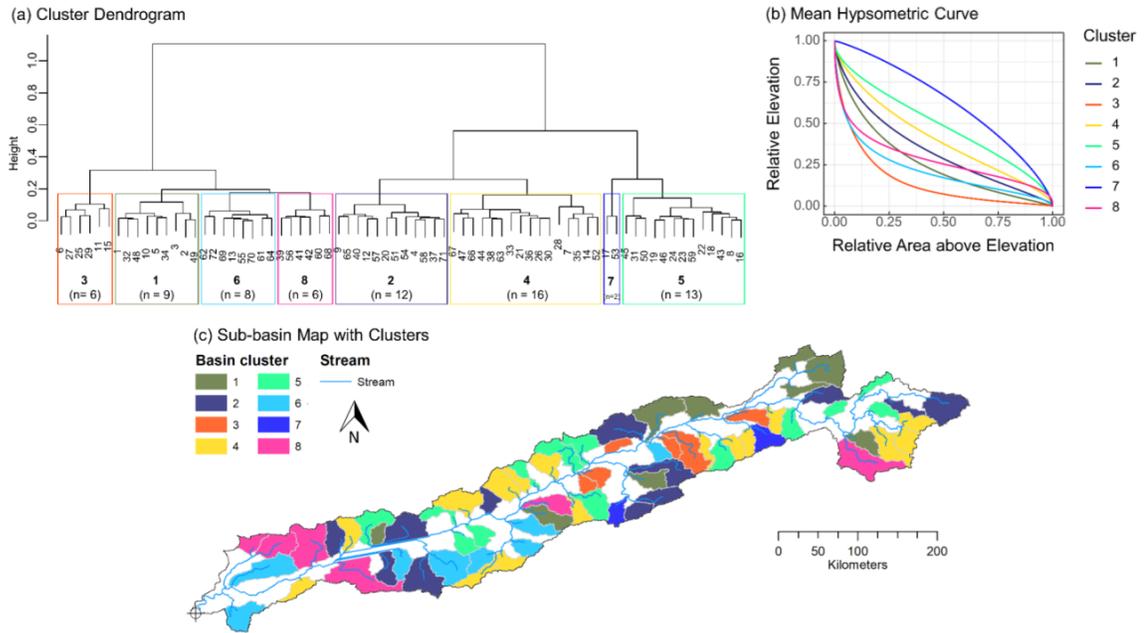
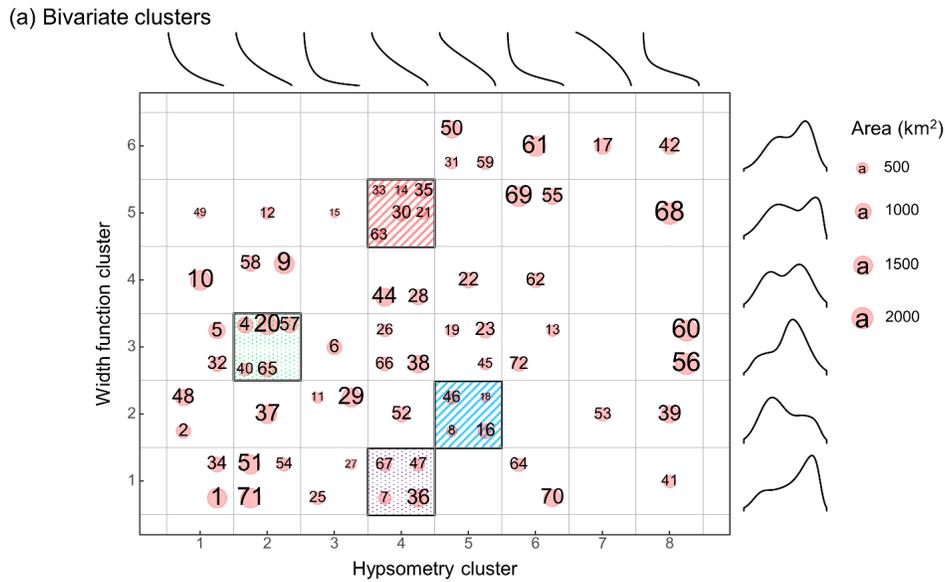


Figure 5.12: (a) Dendrogram of basin hypsometric functions using hierarchical clustering using Ward’s method. (b) Mean hypsometric functions for each cluster. (c) Map of sub-basins grouped by cluster.

of hydrologic response in watersheds. We demonstrated that the classification of basins through clustering when applied using dynamical measures of watershed behavior allows for the partitioning of watersheds into groups with consistent functional forms. We proposed a four-step approach for forming hydrologically similar analogues. This first step involves the functional estimation of two dynamic features, the width function and the hypsometric curve. Next, divergence measures are applied across all basin pairs to form dissimilarity matrices, which are then used for hierarchical clustering. The clusters based on width functions and hypsometries on their own provide groups of basins with similar drainage topology and elevation distribution, respectively. Finally, groups of basins with common width function and hypsometric function clusters can serve as analogous basins with similar hydrological response characteristics. With the wide availability of terrain information, this method can be applied at large scales (national or global) to find a sizeable number of similar hydrological basins at low data and computational costs.



(b) Sub-basin map with clusters

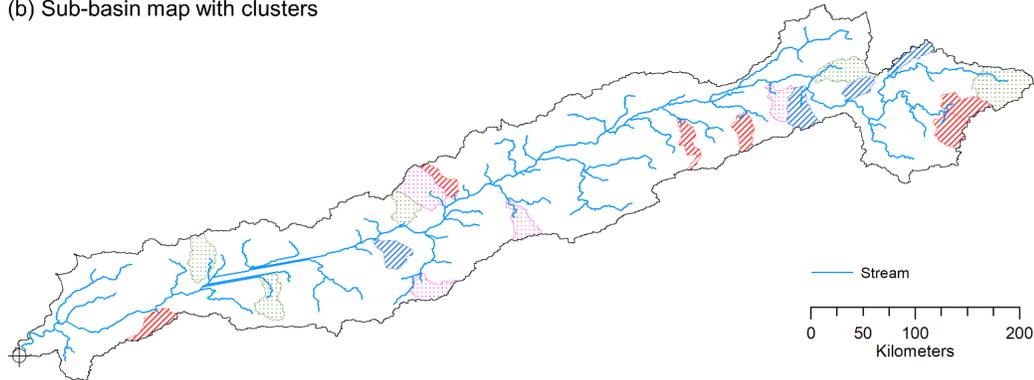


Figure 5.13: (a) Bivariate cluster groups based on the width function and the hypsometric function. The mean curve for each cluster has been shown (width function curves on the right side and hypsometric function curves on the top) to illustrate the nature of the clusters. (b) Map of sub-basins highlighting the bivariate groups with at least four members.

This allows for a large number of catchments to be included in the donor pool and thus, provides a means for the statistical analysis of uncertainty in the hydrological signatures being transferred.

Our study illustrated this framework in the context of the Narmada River basin in India. The following observations and takeaways can be made about the hydrologically similar characteristics across the 72 selected sub-basins for the basin:

1. The majority of width functions exhibit late peaks, with only one out of the six clusters exhibiting an early peak. Interestingly, the early peak cluster seems slightly concentrated near the outlet. Furthermore, a number of sub-basins that share cluster groups appear adjacent to each other.
2. The majority of sub-basins exhibited concave up hypsometric curves, with only two sub-basins showing concave down curves. This could indicate that these sub-basins are surface runoff-dominant and highly eroded. While the hierarchical clustering approach performed well in classifying the overall concavity of the curves, it was slightly less effective in classifying the head and the tail curvatures.
3. There is a level of subjectivity in the choice of the number of clusters. The considerable degree of intra-cluster heterogeneity in the location of the peaks of width functions indicates the need for a relatively large number of clusters for width functions if a high degree of homogeneity is desired. On the other hand, a relatively lower number of hypsometric clusters might be sufficient due to the cumulative nature of the curve which tends to offer a lower variance.
4. Figures 5.9, 5.12, and 5.13 show several sub-basins that share cluster groups but are not adjacent. This points to the shortcoming of relying on proximity alone to find analogues.
5. Two bivariate groups with similar width functions and hypsometric functions were identified with at least four members, one was identified with five members, and one with six members from a total sample size of 72. These represent sub-basins with potentially similar hydrological response characteristics. This can easily be scaled to thousands of watersheds around the world.

The presence of some adjacent pairs of sub-basins with similar width functions and similar hypsometric curves does indicate that the use of neighboring basins as donors has the potential to serve as a means to determine hydrologically similar analogues. However, the lack of a definitive spatial pattern and adjacency of analogous basins, particularly in the case of joint clusters, could indicate that spatial proximity alone might not be a strong predictor of basin hydrological response. Similar basins do not always occur as adjacent neighbors, and while spatial proximity is justifiably a good metric of hydrological similarity in most use cases, dynamic metrics such as the width function can serve as another strong measure in identifying analogues.

Our study allows for some subjective intervention in the clustering process, particularly in the choice of the number of clusters and in the detection and removal of outliers. This conscious choice was made to allow a reasonable degree of flexibility to incorporate the subjective expertise of the dataset and study. These were primarily demonstrated in the choice of the number of clusters as well as the choice to not remove outliers in the case of the hypsometric function-based clustering. While more robust methods are available in the literature (Jung et al., 2003; Almeida et al., 2007; Fan et al., 2013; Zhou et al., 2016; Krleža et al., 2021) and can be used depending on the study, we believe that this level of flexibility allows for an easier application in a large diversity of use cases, which we feel is better than opting for one-size-fits-all solutions.

5.7 Conclusion

Modern data collection techniques such as satellite hydrology and crowd-sourcing tools have led to an explosion in data volume. The future of water sciences hinges on our ability to harness this big data to understand hydrological phenomena based on smart, data-driven computational techniques (Peters-Lidard

et al., 2017; Sit et al., 2020). Our approach focuses on the efficient use of large volumes of elevation data to find hydrological analogues through dynamical properties of terrains and facilitates large-scale applications. This approach utilizes the width function, a major component of the geomorphological instantaneous unit hydrograph, alongside the elevation distribution of watersheds to better reflect the dynamical properties of watersheds in order to assist in assessing hydrological response in ungauged or data-scarce watersheds without the use of hydrological models. Furthermore, the use of functional data reduction and unsupervised learning to derive dynamic metrics for the identification of hydrologically similar analogues at large scales is consistent with the growing recognition in the hydrological community regarding the use of explainable artificial intelligence techniques that build upon conceptual and machine learning models to explain the hydrological phenomena (Maksymiuk et al., 2020; Althoff et al., 2021). In this vein, we presented a flexible approach for utilizing dynamical catchment properties that exert controls over the hydrological response. While we have used the width function due to its function as a building block for the link-based GIUH, and the hypsometric curve as a proxy to the hydraulic response component of the GIUH due to its influence over catchment-scale streamflows, other similar catchment properties can instead be used with this approach to find analogous basins and to aid in similarity assessment. More specifically, this approach provides a means for transferring relevant information about the hydrological response from gauged basins to ungauged basins in a way that reflects dominant flow processes at watershed scale. An application of hydrological similarity study is to assist in improving our understanding of hydrological processes in watersheds (Blöschl et al., 2013) and future works can build upon this study by integrating the width function and elevation-based slope and velocity distributions to create a robust dynamical metric for hydrological response quantification and similarity assessment.

CHAPTER 6

EXAMINING TOPOGRAPHY-BASED METRICS IN RELATION TO HYDROLOGICAL RESPONSE AT EVENT SCALE

6.1 Chapter Abstract

With the growing focus on the prediction in ungauged basins, there is a need to transition from calibration-based models to process-based models that incorporate geophysical dynamics. The hydrological response comprises of an interplay between numerous factors, including topography, land use, geology, soil, and climate. However, we limit this study to topographic metrics as an entry point. In particular, this research assesses the efficacy of distribution-based topographic metrics to predict hydrological response at the event scale by comparing their correlation with five hydrograph characteristics. These metrics include slope, long profile, hypsometry, compound topographic index (CTI), height above nearest drainage (HAND), reduced dissipation per unit length index (rDUNE), and width function. Representative hydrograph properties are derived by averaging the characteristics across historical storms. Width function was found to have a statistical correlation with all five hydrograph properties, the hypsometric curve had a correlation with four, and long profile and CTI had a correlation with one of the properties. Furthermore, these metrics were grouped into three distinct clusters based on principal component analysis. These results suggest that a robust hydrological model based on geomorphology such as a geomorphological instantaneous unit hydrograph should utilize hypsometry and CTI as its components alongside the width function.

6.2 Introduction

The Prediction in Ungauged Basins (PUB) initiative has set out to improve the scientific understanding of hydrological processes in order to improve the modeling and prediction of hydrologic response in ungauged basins. The inadequacy of prevalent hydrological models that rely on parameter fitting through calibration has been recognized by the scientific community (Sivapalan, 2003; Hrachowitz et al., 2013). Consequently, there is a growing need for a shift towards more physical process-oriented methods that are derived from catchment form and incorporate metrics of topography, land use, geology, soil, and climate (Wagener et al., 2007; Gupta et al., 2008; Hrachowitz et al., 2013). This is especially important because the majority of basins in the world are effectively ungauged (Hrachowitz et al., 2013). The most common approach in such basins is to use regression equations that use watershed attributes such as area and slope to determine flow response (Asquith and Roussel, 2009; Gotvald et al., 2012; Dudley, 2015). These equations often carry a large degree of uncertainty (Blöschl et al., 2013).

Significant advances have been made in various aspects of the PUB initiative since its first advent, from a better understanding of the spatio-temporal heterogeneity of physical characteristics and processes to a fundamental shift in focus from empirical, calibration-based models to process-based models (Hrachowitz et al., 2013). While no universal classification of catchments is available yet (Blöschl et al., 2013), the self-organizing nature of catchments has now been recognized (Sivapalan, 2006). The comparative hydrology approach has grown as an important tool for process-based understanding and numerous approaches have been developed that are based on proximity, hydro-climatic region, and catchment indices (Horton, 1932, 1945; Strahler, 1957; Budyko et al., 1974; L'vovich, 1979; Abrahams, 1984; Bras, 1990; Burn and Boorman, 1993; Milly, 1994; Sankarasubramanian and Vogel, 2002; Tung et al., 1997; Rodríguez-Iturbe and Rinaldo, 2001; Aryal et al., 2002;

McIntyre et al., 2005; Woods, 2006; Yadav et al., 2007; Wagener et al., 2007; Reichl et al., 2009; Archfield and Vogel, 2010; Oudin et al., 2010; Patil and Stieglitz, 2011, 2012; Razavi and Coulibaly, 2013; Blöschl et al., 2013; Athira et al., 2016; Brunner et al., 2018). Of these many frameworks, approaches that utilize spatial distributions of hydrologic indices such as the topographic wetness index are of particular interest due to a variety of reasons. Firstly, similarity assessments based on hydrologic indices are not constrained by watersheds that are close to each other (Patil and Stieglitz, 2012). Secondly, distribution metrics potentially capture more information compared to single-valued lumped characteristics such as average slope, drainage density, Strahler ratios, etc., minimizing the loss of information due to simplification (Wooldridge and Kalma, 2001; Wagener and Wheater, 2006; Tetzlaff et al., 2009; Chang et al., 2014). A variety of distribution curves have been linked to runoff generation and flow processes. While the hydrologic response is a complex interplay between a multitude of geophysical and climatic factors, we limit this study to one of these aspects: the landscape topography.

The importance of topography in determining hydrologic response is well established (Horton, 1945). The current landscape topography is the result of a complex geomorphological history, shaping it through tectonic, weathering, and erosional processes. Furthermore, the topography is a key driver of runoff generation and routing. Consequently, there is an abundance of catchment metrics in the literature. To this end, our study aims to: (a) provide a methodology for evaluating the efficacy of distribution-based catchment metrics to assess the hydrological similarity between basins based on outflow hydrographs, (b) develop a meta-organization of these catchment metrics from the standpoint of their formulations and how they relate to basin response, and (c) determine how strongly these metrics correlate with different aspects of flow dynamics such as maximum discharge and hydrograph time to peak. The results of this study can help in the

choice of the set of metrics that can be used as a building block for building a hydrologic response estimation model such as an improved link-based GIUH formulation (Gupta et al., 1986; Bras, 1990). Such models can provide a physical basis for stream runoff prediction in ungauged basins (Gupta and Waymire, 1983; Mesa and Mifflin, 1986; Bras, 1990).

In what follows, we first describe our study area and layout our approach. We then present the results of our study and discuss the implications of our findings. Finally, we discuss some future avenues for this research, especially in the pursuit of developing a link-based GIUH model that can incorporate important metrics to better capture different aspects of flow dynamics.

6.3 Study Area and Data

The U.S. Department of Agriculture (USDA) established a series of experimental watersheds in the mid-1930s to better understand runoff and other basin properties. This program was further expanded to form the Agricultural Research Service (ARS) experimental watersheds. The high-resolution observations from this network have given rise to numerous research methods and models that have guided investments in infrastructure and conservation (Goodrich et al., 2016). The Walnut Gulch Experimental Watershed (WGEW) in Tuscon, Arizona, USA ($31^{\circ}43'N$, $110^{\circ}41'W$) is one of the experimental watersheds within this network and is operated by ARS's Southwest Watershed Research Center (SWRC). This watershed comprises of the upper 148 km² of the Walnut drainage basin (Houser et al., 2000). The watershed elevation ranges from 1220 m to 1938 m, with a mean of 1419 m and a standard deviation of 82 m. The average slope of the watershed is 6.2 degrees, with a standard deviation of 5 degrees. The watershed lies between the Chihuahuan and Sonoran Deserts. The primary land use category is rangeland and the soil types range from silts and clays to well-cemented conglomerates, with the surface soil

texture mostly comprising of gravels and sandy loams (Renard et al., 1993). Consolidated rocks and fan and alluvial deposits dating as far back as Precambrian form the geological makeup of the region. The large-scale land forms and dissected surfaces exhibit the effects of tectonic activities, weathering, and rock erosion. While the watershed geomorphology is complex, it can be described as an actively eroding alluvial-fan surface (Osterkamp, 2008). The drainage pattern is a result of large-scale crustal disturbances that started in Precambrian time and the extensive faulting that began in mid-Cenozoic time (Menges and Pearthree, 1989).

The precipitation in the watershed varies considerably by season and between different years (Renard et al., 1993), with about two-thirds of the rain occurring

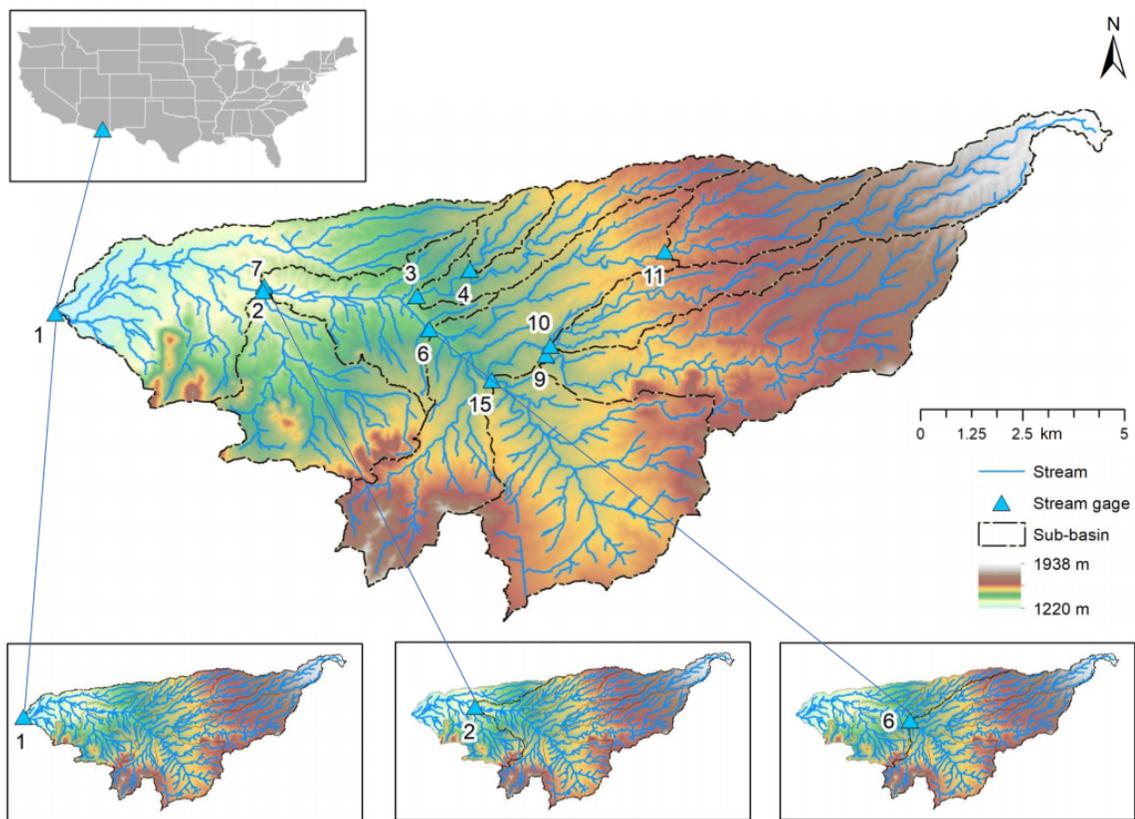


Figure 6.1: Surface topography of Walnut Gulch Experimental Watershed and selected sub-basins. The sub-basins were delineated from 10 selected stream gauge stations. Three of these gauges (1, 2, and 6) lie downstream from other gauges, resulting in larger sub-basins that share boundaries with the upstream sub-basins. These three sub-basins are shown separately in corresponding inset maps.

during the summer monsoon season (Houser et al., 2000). The watershed has a semi-arid climate. Consequently, runoff is ephemeral in nature and the channels are dry most of the year (Stone et al., 2008). They mostly result from short, high-intensity thunderstorms with limited areal extents occurring between July and September (Goodrich et al., 2016). These runoffs are usually short in duration, peak quickly, and occur at near-critical depths (Keppel and Renard, 1962; Renard et al., 1993; Stone et al., 2008). The groundwater depth varies between 45 m and 145 m (Houser et al., 2000).

The hydrological and catchment data was obtained from SWRC database at <https://www.tucson.ars.ag.gov/dap/>. This includes the continuous streamflow data obtained from 19 stream gauges and the digital elevation model (DEM) data. The horizontal resolution of the DEM data is 10 m while the runoff data has a temporal resolution of one minute. Of the 19 stream gauges, nine were not included in this study due to their small size, resulting in insufficient DEM pixels for catchment metric characterization. Of these 10 sub-basins, seven do not overlap each other. Figure 6.1 shows these 10 sub-basins along with the areas that overlap. The areas of the 10 sub-basins range from 2 km² to 148 km², with a mean of 45 km² and a standard deviation of 50 km². The naming convention for the stream gauges is "FL" followed by a number. We simply use this number as the identifier for the stream gauge as well as its contributing sub-basin.

6.4 Approach and Methods

6.4.1 Topographic Metrics

Due to the importance of topography in shaping the basin hydrologic response, there is a large number of indices and metrics in hydrology. As such, it is impractical to include every metric in this research, and we limit our study to some of the promising ones in the literature that have either been used prominently in

hydrological models or have physical significance in shaping underlying hydrologic and hydraulic processes. Furthermore, we only include distribution-based metrics to maintain consistency among the included metrics, meaning some important indices such as the Horton-Strahler ratios and drainage density (Horton, 1945; Strahler, 1957) have not been included. The metrics we included in this study are summarized in Table 6.1.

Table 6.1: Catchment metrics used in this study. Abbreviated forms of these metrics are used in some of the figures in this study, and are shown here. Abbreviated names with asterisks denote average values along the flow path. The primary references for these metrics are also shown where necessary.

Metric	Abbreviation used	References
Slope distribution	Slope, Slope*	
Hypsometric curve	Hypso	Horton (1932); Langbein (1947)
CTI distribution	CTI, CTI*	Kirkby (1975); Beven and Kirkby (1979)
HAND distribution	HAND, HAND*	Crave and Gascuel-Oudoux (1997); Rennó et al. (2008)
rDUNE distribution	rDUNE, rDUNE*	Loritz et al. (2019)
Long profile	Trunk	
Width function	WF	Gupta et al. (1986); Bras (1990)

Different aspects of topography affect how the catchment responds to rainfall input, eventually manifesting in the shape of the outflow hydrograph. Land surface slope affects whether the precipitation infiltrates into the subsurface or runoff. Channel slopes and hillslopes also affect the velocity of travel. Average catchment slope has often been used as a metric for similarity studies or in regression equations for determining hydrologic response (Asquith and Roussel, 2009; Gotvald et al., 2012; Dudley, 2015) To maximize topographic information, we instead use slope distribution as a metric in this study. Elevation distribution is another important catchment property that can reflect streamflow mechanisms in the watershed. The cumulative elevation distribution curve is represented by the hypsometric curve

(Horton, 1932; Langbein, 1947). This curve demonstrates the distribution of mass as a function of elevation (Harlin, 1984). While hypsometric curve has traditionally been used to study basin maturity based on erosional processes (Strahler, 1952; Moglen and Bras, 1995; Pedrera et al., 2009), newer studies have argued for a link between basin hypsometry and catchment hydrological properties (Harlin, 1980, 1984; Howard, 1990; Willgoose and Hancock, 1998; Marani et al., 2001; Luo and Harlin, 2003; Vivoni et al., 2008; Huang and Niemann, 2008a). We use the percentage hypsometric curve in this study because it allows for the comparison of basins of varying sizes and altitudes (Strahler, 1952). This curve plots the relative area above elevation, i.e. the proportion of area above a given elevation, against the relative elevation, i.e. the ratio of elevation and relative relief. Furthermore, hydrological models such as TR-55 use the profile of the longest flow path, called the long profile, to determine the time of concentration (SCS, 1986). We, therefore, use this as another metric in this study.

One of the most commonly used hydrologic metrics is the compound topographic index (CTI), also called the topographic wetness index (TWI), defined as the ratio of the upstream drainage area to the local slope (Kirkby, 1975; Beven and Kirkby, 1979). CTI is formulated as shown in equation 6.1.

$$CTI = \ln \left(\frac{\alpha}{\tan\beta} \right) \quad (6.1)$$

where α is the contributing area, and β is the slope at each point (grid cell). The popular hydrological model, TOPMODEL uses CTI with the assumption that the points in a catchment with the same CTI value respond in a hydrologically similar way (Beven and Kirkby, 1979; Beven, 1997, 2011). CTI intuitively incorporates the tendency of water to accumulate or flow based on higher or lower ground slope and overland runoff contribution, and has subsequently been used to delineate hydrologically important zones such as wetlands (Rampi et al., 2014; Horvath et al.,

2017; Bian et al., 2021). Another popular hydrologic metric is the height above nearest drainage (HAND), defined as the elevation of a point on the basin above the nearest drainage based on its flow path (Rennó et al., 2008). This index was first developed by Crave and Gascuel-Oudoux (1997) and named elevation difference (DZ). To compare HAND with other metrics, we use its natural logarithm. From this point of the study, we simply use HAND to refer to $\ln(\text{HAND})$. HAND assumes that the gravitational potential energy of a given unit weight of water with a reference level set to the elevation of the nearest river acts as the main driver for overland and sub-surface storm flow. As such, HAND distribution represents predominant control on lateral distribution and redistribution of water in a catchment and has been shown to be highly correlated with the depth of the water table (Nobre et al., 2011). Along with numerous applications in hydrological analysis, HAND has been used extensively in similarity studies and has been proven to be a good metric for catchment classification (Gharari et al., 2011; Gao et al., 2014). However, this metric does not consider any form of energy losses along its flow path (Loritz et al., 2019), and Kleidon et al. (2013) found that this metric was not sufficient in assessing similarity with respect to runoff generation due to the dissipative nature of water flow. Considering these limitations, Loritz et al. (2019) proposed a new metric, reduced dissipation per unit length index (rDUNE) which accounts for both the driving potential difference and the accumulated dissipative loss along the flow path following straightforward thermodynamic arguments (equation 6.2).

$$rDUNE = -\ln\left(\frac{HAND}{x}\right) \quad (6.2)$$

where $HAND$ and x represent the HAND and flow length distance to the nearest stream respectively. The majority of incoming potential energy associated with water flow within the hillslope is dissipated and only a fraction of it reaches the

stream network as kinetic energy. The driving potential at the hillslope scale is the geo-potential difference between the upstream catchment area and stream channel. Dissipative losses have a high degree of variability but accumulate along the flow path. Consequently, rDUNE uses the flow distance length to the river as a proxy for these losses due to their proportionality in normal conditions. Loritz et al. (2019) showed that rDUNE is a better metric of similarity than CTI and HAND based on hydrogeological differences in landscapes. We use the distribution of CTI, HAND, and rDUNE over the watershed as three more metrics for similarity assessment.

Another important catchment hydrologic property is the width function, which represents the travel distance distribution of a stream network (Gupta et al., 1986; Bras, 1990). For a given drainage basin, the width function, $N(x)$, denotes the areal extent between x and $x + dx$, where x represents the total distance along the flow path to the outlet (Veneziano et al., 2000). If we assume a constant flow velocity, the width function represents the probability distribution of travel times or the instantaneous unit hydrograph, thereby reflecting the topological features of a basin's stream response (Lashermes and Foufoula-Georgiou, 2007; Moussa, 2008). The width function is strongly linked to the shape of the hydrograph (Kirkby, 1976; Gupta and Waymire, 1983; Troutman and Karlinger, 1984, 1989). Furthermore, the width function forms a building block of the link-based geomorphological instantaneous unit hydrograph (Gupta et al., 1986; Bras, 1990), defined as:

$$h(t) = \int_0^{x_{max}} g(x, t) N(x) Z^{-1} dx \quad (6.3)$$

where $h(t)$ denotes the GIUH, t denotes time; $g(x, t)$ denotes the hydraulic response, x denotes the hydrologic flow distance from the outlet of the basin; and the term $N(x)Z^{-1}$ forms a scaled width function, where x is linearly scaled between 0 and 1. GIUH is particularly suitable for modeling the hydrologic response in data-scarce or ungauged basins (Kumar et al., 2002; Ellouze-Gargouri and Bargaoui, 2012; Sahoo

and Saritha, 2015; Swain et al., 2015). Since the x-axis of the width function is the longest flow path along the watershed, we also modified metrics for slope, CTI, HAND, and rDUNE, whereby their average values are taken along the x-axis of the width function. These modified metrics were tested alongside their simple distributions.

6.4.2 Hydrograph Properties

Runoff response can be quantified by a variety of metrics. In long-term scales, various streamflow indices such as runoff ratio, slope of flow duration curve, and baseflow index can be used to describe the characteristics of runoff. However, at the event scale, the hydrograph is the most accurate indicator of surface runoff. Hydrograph properties such as the maximum discharge and time to peak are some of the most important characteristics in the study of streamflow response. Consequently, we use hydrograph properties as empirical similarity metrics in order to evaluate the catchment-based metrics. However, storms occur on different temporal and spatial scales. Different sized catchments, by their very natures, will produce different hydrographs for the same storm, whereas storms with different intensities, duration, and spatial and temporal distributions will produce different hydrographs in the same catchment. An approach to standardize the response for any given catchment is through the use of a unit hydrograph, defined as the outflow hydrograph resulting from one unit of effective rainfall uniformly distributed over space and time (Sherman, 1932). There are various approaches to the computation of unit hydrographs, including deconvolution through matrix calculation and the use of synthetic hydrographs (Chow et al., 1988). However, the former suffers from a large number of parameters leading to computational instability, thereby restricting its practical application (Singh, 1976; Rao and Tirtotjondro, 1995). While newer methods have been suggested (Yang and Han, 2006; Seong and Sung, 2021), they

are yet to gain widespread adoption. On the other hand, synthetic unit hydrographs generally have associated assumptions and parameters that make them suitable only when selected criteria are met (Young et al., 1986).

In this study, we have strived to keep the assumptions to a minimum when developing empirical hydrological response metrics based on streamflow observations. As such, we use descriptive statistics such as the slope of rising limb, skewness, and kurtosis of the hydrograph similar to the study by Tardif et al. (2009). These metrics describe the shape of the hydrograph and are independent of the hydrograph size. This means that these properties are comparable across different storms in the same catchment, and taking a mean of these values can provide a good indication of the average shape descriptions of the hydrograph. Skewness provides the overall shape of the hydrograph and to an extent, the early or late occurrence of the hydrograph peak. On the other hand, kurtosis can be a good indicator of the flashiness of the hydrograph. Additionally, due to the importance of the maximum discharge and time to peak, we include these properties by scaling them using the time base of the hydrograph, defined as the total hydrograph duration. This method requires individual runoff events to be isolated from the time series and many approaches are available in the literature (Carey and

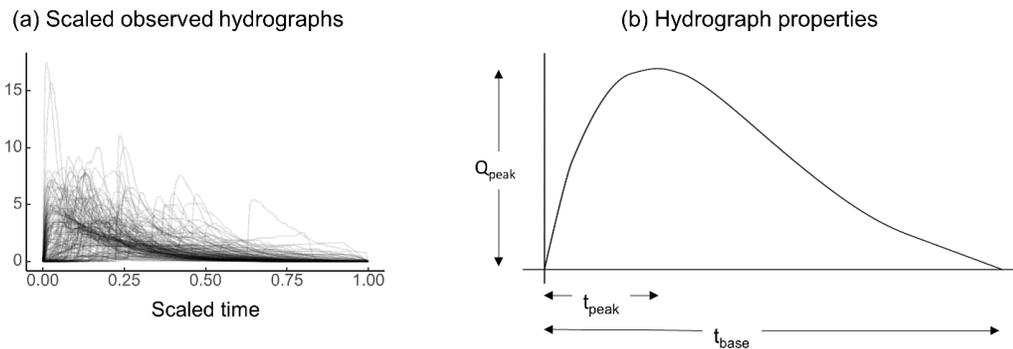


Figure 6.2: (a) Historical storm discharges for sub-basin 6 scaled to unit time scale for visualization. (b) Some key properties of the hydrograph, including the peak discharge (Q_{peak}), time to peak (t_{peak}), and time base (t_{base}).

Woo, 2001; Spence, 2006; McNamara et al., 1998; Tardif et al., 2009). However, due to the ephemeral nature of runoff in WGEW, most individual storm events are already isolated. However, not all individual storm events are isolated, especially in the case of complex storms. However, due to the availability of long-term data, these aberrations should be averaged out as is evidenced in visual inspection. Figure 6.2 provides an example for a sub-basin with all event hydrographs overlaid. Here, the time bases and total areas are linearly scaled to unit values. However, we would like to emphasize that this scaling is for visual demonstration only and is not required for the computation of the aforementioned hydrograph properties. To check the representativeness of the averaged metrics, we took four large storms that occurred over the entire watershed and compared their properties with the averaged ones6.3. These four storms were chosen randomly from the set of large, contemporaneous historical storms with relatively similar rainfall inflows. While there is considerable variance as would be expected, the averaged metrics do seem to reflect the overall trends of the individual storm properties.

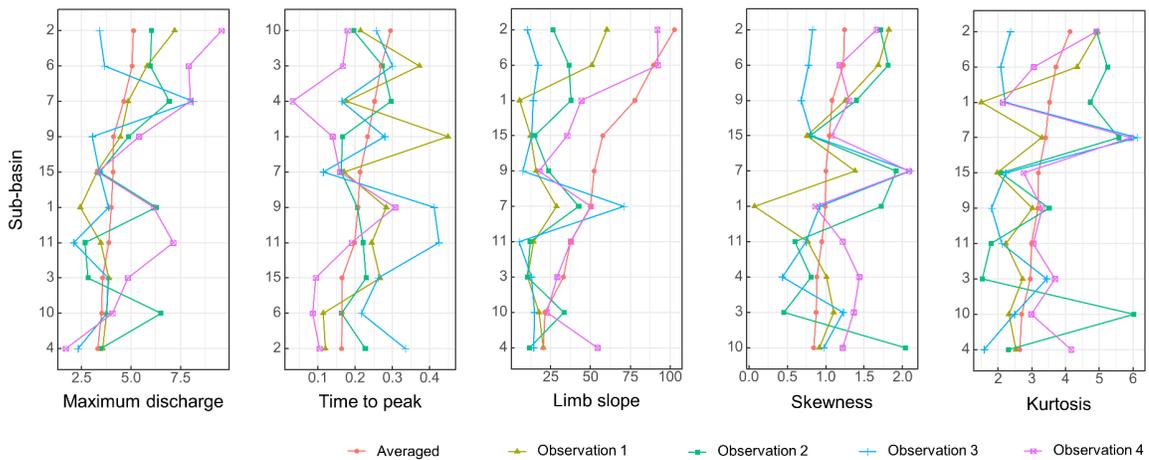


Figure 6.3: Averaged hydrograph properties compared with properties of four storms plotted occurring over the entire watershed. The red lines with circular dots denote the averaged properties, while the observed storm properties are denoted by other colors and shapes. For each property, sub-basins are ordered by the averaged values.

6.4.3 Spatial Analysis and Similarity Assessment

We derived the stream network from the DEM using standard Geographic Information System (GIS) procedures in ESRI ArcGIS 10.5.1. The upstream drainage area contribution threshold for stream delineation was chosen to closely match the National Hydrography Dataset (NHD) stream network and the generated stream network was manually verified with satellite photos. Next, the contributing drainage areas for the 19 stream gauge stations were delineated. Of these, nine sub-basins with less than 0.3 km² area were not included in this study as they did not have enough elevation values to resolve all of the catchment metrics properly. Next, we derived a variety of spatial properties at each grid cell of the DEM, including the slope, upstream contributing drainage area, downstream flow distance to the drainage outlet, CTI, HAND, and rDUNE as well as the longest flow path. For each sub-basin, the distribution of these grid values and the average values along flow distances were then computed.

The statistical analysis was performed in R programming language. Pair-wise sub-basin similarity matrices were constructed for each of the catchment metrics and for the hydrograph properties. Similarity for the latter was simply defined as the difference between corresponding values. However, the catchment metrics are in distribution form and hence, a divergence measure was used to assess the "distance" between two distributions (Mahalanobis, 1936; Bhattacharyya, 1943; Kailath, 1967). (Loritz et al., 2019) used the Jensen-Shannon divergence (Lin, 1991) to estimate the similarity in distributions. However, since not all metrics in this study are in the form of probability distributions, we use the discordance index (DI) (Equation 6.4) , defined as the absolute area between two curves (Bajracharya and Jain, 2021).

$$DI = \int_0^1 |f_1(x) - f_2(x)| dx \quad (6.4)$$

DI is the integration of the absolute value of the difference between two distribution curves. The values range between zero and one, with a value of zero indicating identical distributions.

We then performed a Pearson correlation between the catchment metric-based and the hydrological property-based distance matrices. Since overlapping sub-basins are expected to share catchment properties (Gottschalk et al., 2006; Skøien et al., 2006; Skøien and Blöschl, 2007; Gottschalk et al., 2011; Müller and Thompson, 2015), we excluded them from the correlation analysis. However, for completeness, results for all sub-basins including the nested ones are provided in the Appendix D.1. Statistically significant correlation, as shown by the p-value, is an indicator that the catchment metric is related to the hydrological behavior. In this manner, we mapped the relationships between the catchment metrics with the hydrograph properties.

Finally, we also performed a principal component analysis (PCA) to classify these metrics into clusters. The pairwise DI values for each metric were used as the basis of this analysis. Since DI values vary considerably among these metrics, ranks based on DIs were used instead to ensure that the results were independent of the scale and any monotone transformations (Oja, 2010; Aluja, 2018).

6.5 Results and Discussion

6.5.1 Interrelationships Among Catchment Metrics

Figure 6.4 shows the plots of the catchment metric. The frequency distributions for CTI, HAND, and rDUNE are mostly Gaussian, with slight skewness. CTI and slope distributions are right-skewed while HAND is left-skewed. The average CTI and rDUNE curves along the flow path are noticeably similar in their curvatures, as are average slope and HAND curves. CTI and rDUNE share a slope term in the denominator in their formulations, however, the former uses local slope while the

latter uses downstream slope to the nearest drainage. By definition, CTI and local slope are inversely related, while rDUNE is an energy-based reimagination of HAND (Loritz et al., 2019). Furthermore, there is no direct mathematical relation between local slope and HAND that can be derived from their mathematical formulations. This could suggest that these properties are related through underlying basin geomorphology. The long profile has a slight resemblance with the hypsometric curve, with the head of the hypsometric curve somewhat resembling the upstream profile (except in the case of sub-basin 10). However, the tail region of the hypsometric is distinctly different from the downstream region of the long profile. Strahler (1952) showed a linear relationship between the hypsometric curve and the

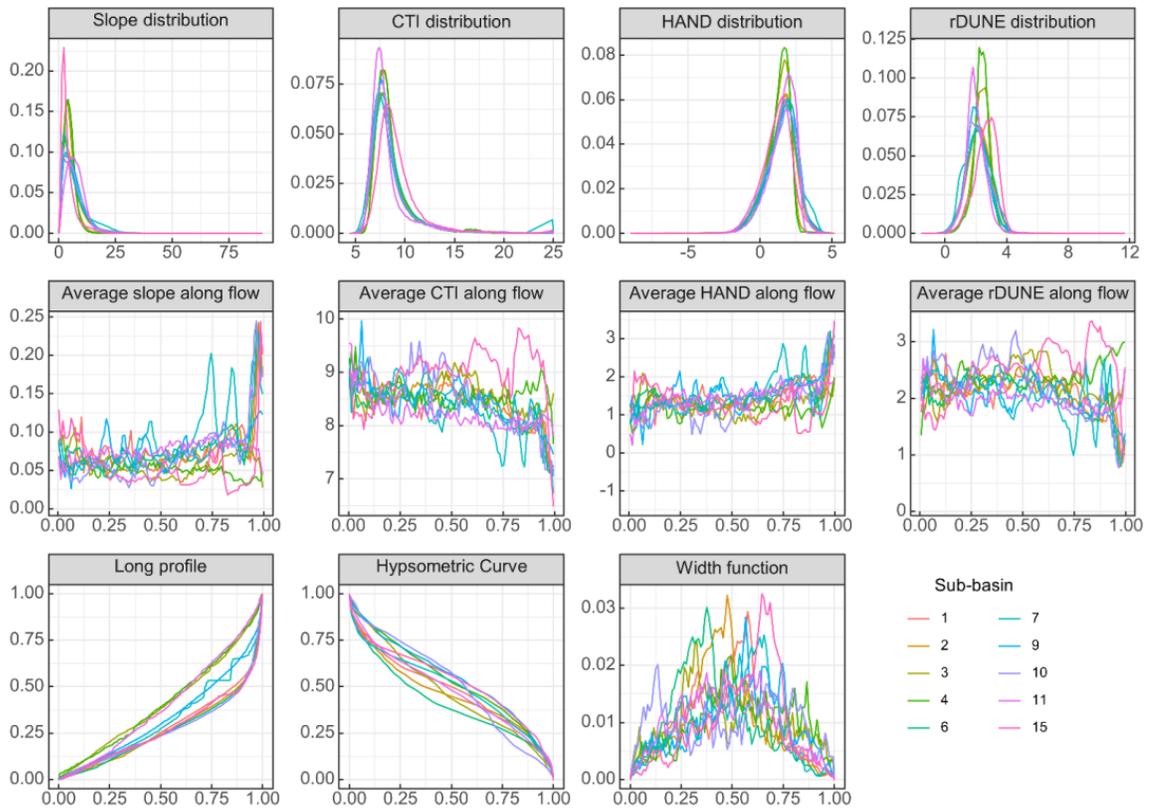


Figure 6.4: Distribution of catchment metrics. Slope, CTI, HAND, and rDUNE are shown as density functions and as an average along the flow path. HAND is shown in a natural logarithmic scale ($\ln(\text{HAND})$). The x-axis of the width function and the x- and y-axes of the long profile and hypsometric curve are linearly scaled to unit length.

physical ground slope of the watershed if the slope of the former is adjusted for the contour length. Width functions show a considerable degree of variation among the sub-basins and most of the width functions are multi-peaked.

Figure 6.5 shows the pairwise distance matrices for four of the 11 catchment metrics and two of the five hydrograph properties. A visual inspection of the pairwise DI values indicates that width function and hypsometric curve have a strong resemblance with the pairwise difference in hydrograph properties. While rDUNE and HAND do not show much correspondence with the hydrograph properties, they show a clear correlation with each other. In fact, four of the catchment metrics, slope, CTI, rDUNE, and HAND show a strong correlation with

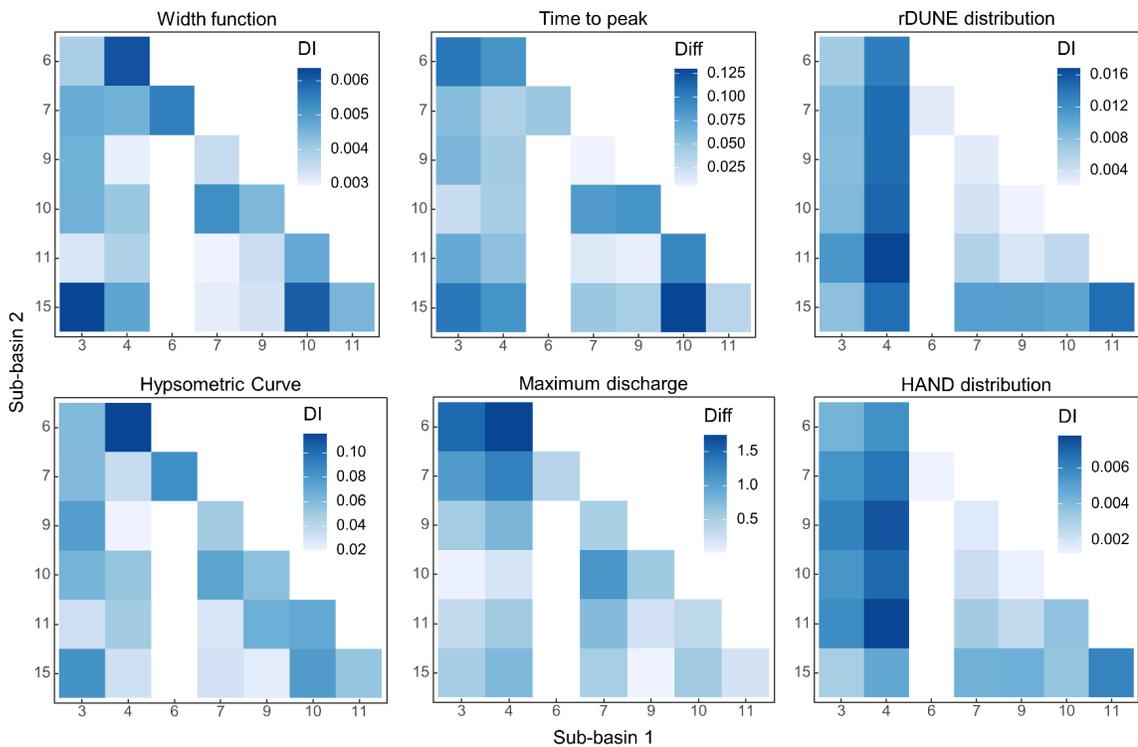


Figure 6.5: Pairwise differences in metrics for sub-basins. For width function, rDUNE, hypsometric curve, and HAND, the difference is computed as discordance index (DI) values. For time to peak and maximum discharge, the difference simply means the arithmetic difference between the values. Sub-basins with shared boundaries have not been included. DI values range between 0 and 1, with a higher value indicating a higher divergence.

each other, further strengthening the notion that these metrics are physically related.

The interrelationship among the catchment metrics was further examined through principal component analysis (PCA). Before running this analysis, we performed Bartlett’s test of sphericity to check whether these metrics are suitable for structure detection. With a Chi-Square test statistic of 247 and a p-value of 4.3×10^{-26} , the metrics were found to be significantly related and thus, suitable for a PCA analysis. The first two principal components (PC) explain 68.5% of the variance (Figure 6.6). The results of the analysis further reinforce our earlier assertions about how these metrics are interrelated. The width function is strongly correlated with the first principal component (PC1), while slope, CTI, HAND, and rDUNE are all strongly correlated with the second (PC). Hypsometric curve and long profile are correlated with both of the principal components. These variables

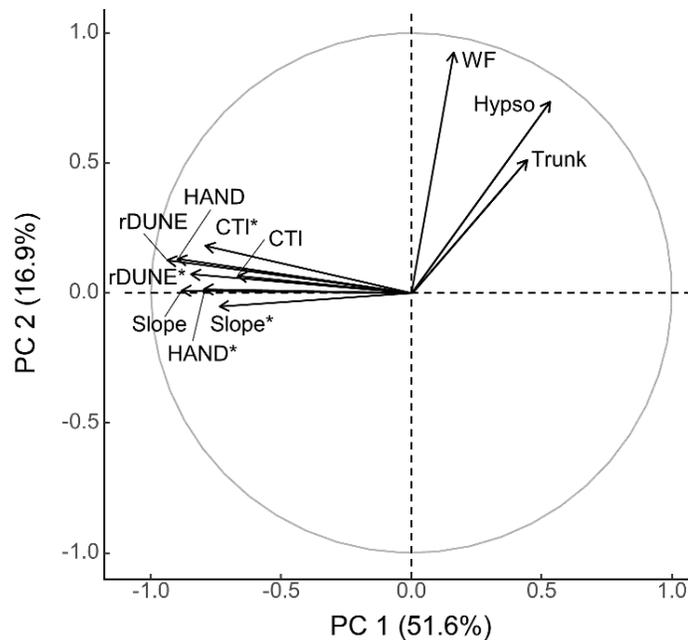


Figure 6.6: Factor map of catchment metrics with the first two principal components (PC) as the axes. These two components represent 68.5% of the variance. Arrows indicate which PC each metric contributes towards and lengths indicate the magnitude of contribution. Overlapping sub-basin pairs were not included in the analysis.

form three distinct clusters in the first factorial plane. While a more detailed analysis of this is required to draw any definitive conclusions, this preliminary analysis indicates that these three clusters carry relative complementary information about the catchment, and can form a basis for a more complete hydrogeomorphic model to analyze catchment processes.

6.5.2 Assessment of Catchment Metrics in Relation to Hydrograph Properties

We, then, compared how well these catchment metrics predict hydrograph shape by plotting the pairwise DIs for the catchment metrics against the pairwise arithmetic difference in hydrograph properties. The plots for the width function are shown in Figure 6.7. Visually, the relationship seems linear, with some outlier points, particularly in the case of maximum discharge, limb slope, and skewness. We further tested the relationship through a Pearson correlation test. We began from the premise that if a catchment metric captures flow dynamics of watersheds, the similarity in the metric should correlate with the similarity in hydrologic response characteristics. Figure 6.8 shows the results for two of these properties, time to peak and maximum discharge. The former correlates with the hypsometric curve and width function while the latter correlates with the hypsometric curve, CTI, and width function. Significance codes for p-value less than 0.05, 0.01, and 0.001 are indicated by shades of blue and the number of '*' symbols. Figure 6.8 (b) indicates that the hypsometric curve and long profile have high degrees of correlation with the maximum discharge. It is, however, prudent to note that historical maximum discharges are subject to a high degree of irregularity, with frequent large storms leading to statistical outliers. Several tropical storms have affected Arizona and these can result in unusually high downpours and runoffs. While the effect of outliers can be removed by choosing to use median instead of mean, we opted to go

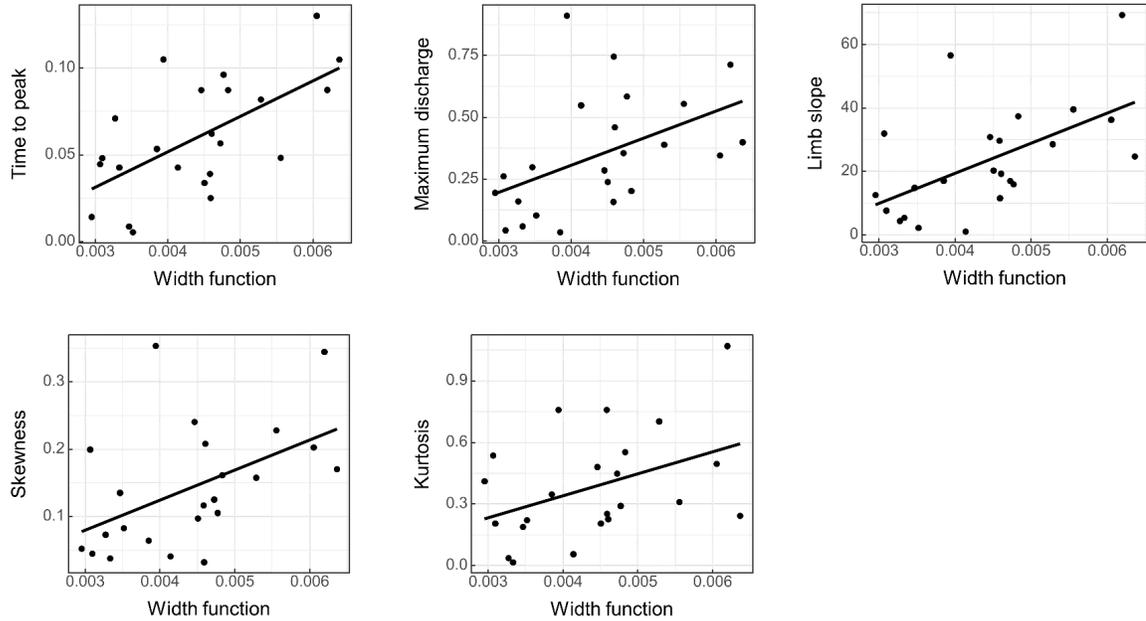


Figure 6.7: Comparison of similarity based on width function and similarity based on hydrograph properties. Pairwise DIs for width function are plotted on the x-axes and the arithmetic difference between hydrograph properties are plotted on the y-axes. A linear regression line is shown to demonstrate the relationship. Overlapping sub-basin pairs were not included in the analysis.

with the latter since these large storms are of particular interest in hydrological modeling, especially in engineering applications. Finally, the results for all hydrograph properties is summarized in Table 6.2. The same significance codes are used to indicate the level of significance of correlation. The width function was found to correlate with all five hydrograph properties, while the hypsometric curve correlates with four of them. Finally, CTI and long profile had a correlation with one of the properties, the maximum discharge. None of the other catchment metrics had statistical significance with the five hydrograph properties in this study.

The results for width function are somewhat expected since its importance in the theoretical GIUH model is well mentioned in a number of studies (Surkan, 1969; Kirkby, 1976; Beven, 1979; Bras, 1990; Naden, 1992; Rinaldo and Rodriguez-Iturbe, 1996; Rodríguez-Iturbe and Rinaldo, 2001; Grimaldi et al., 2010, 2012; Moussa, 2008). However, relative to its potential, its use in hydrologic response prediction in

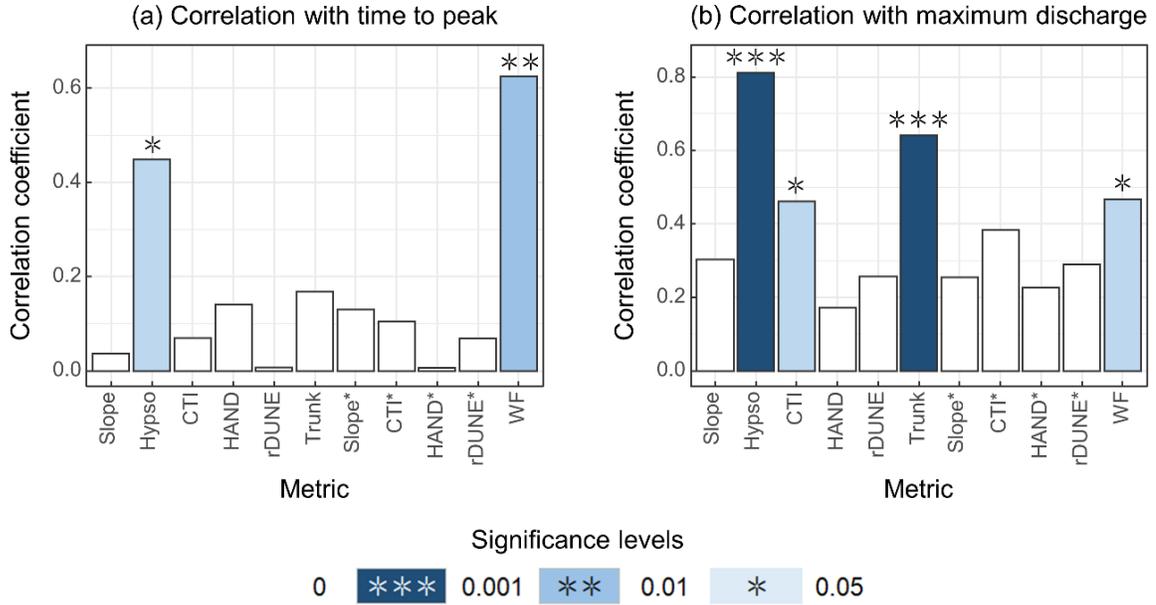


Figure 6.8: Pearson correlation between similarity based on the catchment metrics and based on two hydrograph properties. The properties are (a) hydrograph time to peak and (b) maximum discharge. The statistical significance of the correlation based on p-values is denoted using the significance codes. Darker shades of blue with a higher number of ‘*’ symbols indicate a higher significance. Overlapping sub-basin pairs were not included in the analysis.

Table 6.2: Statistical significance of Pearson correlation between similarity based on the catchment metrics and based on all hydrograph properties in this study. The correlation significance based on p-values is denoted using the significance codes. Darker shades of blue with a higher number of ‘*’ symbols indicate a higher significance. Overlapping sub-basin pairs were not included in the analysis.

Metric	Tmax	Qmax	Limb Slope	Skewness	Kurtosis
Slope distribution					
Hypsometric curve	*	***	**	*	
CTI distribution		*			
HAND distribution					
rDUNE distribution					
Long profile		***			
Average slope along flow					
Average CTI along flow					
Average HAND along flow					
Average rDUNE along flow					
Width function	**	*	**	*	*



ungauged basins seems limited, as evidenced by its limited mention in a number of review publications (Wagener et al., 2007; Blöschl et al., 2013). We hope that this result further increases the scholarly interest in the potential of width function within the hydrologic community. The statistical significance of CTI and long profile are of no surprise either. The former is already a well-established metric in hydrology and is actively used in some hydrologic models to distinguish catchment behavior (Beven and Kirkby, 1979; Beven, 1997, 2011). The latter is often used in hydrologic models for computing the time of concentration (SCS, 1986). On the other hand, the use of the hypsometric curve in hydrologic response study is far from prominent. A number of research studies have been showing links between hypsometry and basin fluvial processes (Harlin, 1980, 1984; Howard, 1990; Willgoose and Hancock, 1998; Marani et al., 2001; Luo and Harlin, 2003; Vivoni et al., 2008; Huang and Niemann, 2008a). In particular, Harlin (1980), Harlin (1984), and Luo (2000) demonstrated that statistical moments of the hypsometric curve are related to various basin hydrogeomorphic properties such as the rate of change in the upper, mid, and lower basin slopes, headward development of main stream and its tributaries, and time to hydrograph peak. Luo and Harlin (2003) derived a theoretical travel time that can be calculated from the hypsometric curve based on gravitational potential. Our findings similarly suggest a link between hydrograph properties and hypsometry and further motivate deeper research into the hypsometric curve from the aspect of the hydrologic response.

6.6 Concluding Remarks

With a growing need to improve existing measures to understand and predict hydrologic response in ungauged basins, significant advances have been made in improving hydrologic predictions that are grounded in science and process-based understanding, much remains to be done in achieving robust and reliable predictions

(Blöschl et al., 2013; Hrachowitz et al., 2013). The hydrologic response is a complex interplay between numerous factors, including topography, land use, geology, soil, and climate (Wagener et al., 2007; Gupta et al., 2008; Hrachowitz et al., 2013). In this article, we focused on one aspect, the topography. More specifically, we checked numerous distribution-based topographic metrics in relation to five hydrograph properties. Two key takeaways of this study are highlighted next.

1. The catchment properties can be broadly grouped into three clusters. The first cluster is the width function, which holds key information about the drainage network and connectivity in the watershed. Gupta et al. (1986) recognized that altitude is the missing dimension in linking networks with watershed hydrology. In modeling instantaneous unit response, basins are often conceptually divided into channel networks and hillslope, with the travel time from the hillslope to the stream network assumed as negligible. However, due to the longer time for the sub-surface flow component to reach the network, hillslope contributions cannot be neglected (Mesa and Mifflin, 1986; Gupta et al., 1986). Our study shows that alongside width function, two clusters hold complementary information about the watershed topography. The first cluster includes slope, HAND, CTI, and rDUNE, and the second includes the hypsometric curve and long profile.
2. The width function was found to be statistically correlated with all five hydrograph properties, supporting the prevalent literature that network properties play a key role in shaping basin response (Gupta et al., 1986). The statistical significance of hypsometric curve, long profile, and CTI means that these metrics can be potentially used alongside the width function to address the altitudinal component and take sub-surface processes into account. Since the hypsometric curve and long profile lie in the same cluster, we suggest using only the hypsometric curve due to its better correlation with hydrograph

properties. Attempts have already been made to incorporate CTI to create a rescaled width function that responds to various saturation conditions and varying runoff-contributing portions of the basin (D’Odorico and Rigon, 2003; Rigon et al., 2001). Additionally, various attempts have been made to develop an expression for velocity that can be included in the GIUH model (Al-Wagdany and Rao, 1997; Chen et al., 2019). However, attempts have yet to be made to directly include the hypsometric curve in a GIUH formulation. Bajracharya and Jain (2022) have developed a clustering-based approach to classify watersheds based on hypsometry and width function. Functional formulations of both, width function (Bajracharya and Jain, 2020) and hypsometric curve (Strahler, 1952; Harlin, 1978; Bajracharya and Jain, 2021) have already been developed. Furthermore, Luo and Harlin (2003) have derived a theoretical travel time that can be calculated from the hypsometric curve. These pave the way for the development of an analytical formulation of the GIUH model that goes beyond the moments of the width function and incorporates the hypsometric curve and CTI.

The methodology presented here for the assessment of catchment properties in relation to observed responses can be further extended in various ways. Firstly, we only included the hydrograph properties to keep our scope limited to event scale response. Long-term streamflow characteristics such as the runoff ratio, baseflow index, etc. can be used instead to gauge their relationship with catchment metrics. Secondly, we focused only on distribution metrics in this study. There are a number of single-valued catchment properties and metrics such as the total area, mean elevation, mean slope, drainage density, Horton-Strahler ratios, etc. as well as a suite of metrics related to climate, soil, land use, etc. that hold key information about catchment behavior, and this study can be extended to test the efficacy of these metrics. Of particular interest is the basin area, which is commonly used in

regression equations to predict streamflow characteristics (Asquith and Roussel, 2009; Gotvald et al., 2012; Dudley, 2015). While we do not include basin area in this study, it bears mentioning that area still implicitly plays a role in our analysis. Metrics such as width function and hypsometric curves are scaled by the basin area and the hydrograph properties are scaled by the time base, which itself is a function of area. Any GIUH formulation that takes these catchment metrics into account needs to include an area term to reverse the scaling. Next, these metrics were tested in one watershed in a semi-arid climatic region. For a more conclusive assertion regarding these metrics, these tests need to be replicated in other climatic regions and in watersheds with different landscapes and geomorphology. By demonstrating the potential value of supplementing the width function with the hypsometric curve and CTI, and by providing a methodology for inspecting other metrics, this study takes a few steps forward in determining the building blocks for a comprehensive, process-based GIUH model.

CHAPTER 7

CONCLUSION

This research has sought to contribute to the existing body of literature in the following ways:

1. Several visualization tools and metrics for evaluating the wetland flow dynamics and potential flood attenuation benefits of wetlands were demonstrated. These include popular metrics such as the compound topographic index (CTI), width function, and hypsometry along with some novel tools that focus on storage capacity and connectivity. These tools can be useful for decision-makers to prioritize wetlands for investment at watershed scale.
2. Popular topography-based distribution curves were evaluated and grouped into three clusters. The width function falls in the first cluster, the second cluster consists of the hypsometric curve and long profile, and the third includes slope, HAND, CTI, and rDUNE. These three clusters were shown to carry complementary information about the catchment topography. The width function carries key information about the stream network and connectivity in the watershed. It was found to be statistically correlated with all five hydrograph properties included in the study. This is in line with the prevalent literature that drainage network properties play a key role in determining streamflow response (Gupta et al., 1986). From the second cluster, the hypsometric curve was found to be statistically significant with four of the hydrograph properties, and the long profile with one. Consequently, the hypsometric curve should be a better metric to use in modeling basin response behavior. From the third cluster, CTI was found to be statistically significant

with one of the properties. This suggests that the hypsometric curve and CTI can be used alongside the width function to add an altitudinal component in approaches such as GIUH. Gupta et al. (1986) suggested that altitude is the missing dimension in linking stream networks with drainage basin hydrology.

3. Robust functional forms were developed to represent the width function and hypsometric curve. The mixture of Skew-normal distribution was found to be appropriate for width functions. This approach can capture various shapes and modalities of the width function. Furthermore, the approach provides versatility in capturing the desired level of detail by varying the number of mixture components. A three-parameter function was developed to represent the hypsometric curve. This function was shown to provide improvements over existing functions in capturing diverse shapes of the hypsometric curve. More specifically, it was able to capture the head and toe regions well. These regions are important because have been shown to be linked with various important catchment properties (Willgoose and Hancock, 1998). These functional forms provide multiple advantages. First, they provide an efficient way to capture the response characteristics from large terrain datasets such as the LiDAR-based digital elevation model data and to efficiently store them as a small set of parameter values. In addition to data reduction, this approach is also computationally efficient. Second, these functional forms pave the way for developing a fully analytical GIUH formulation.
4. A hierarchical clustering-based approach was developed that uses width function and hypsometric curve to find analogous basins based on dynamic response characteristics. this physics-informed machine-learning approach is computationally efficient and can be used in large-scale applications, including at national and global scales. Consequently, this enables the possibility of

drawing information from a large, potentially global pool of donor basins with streamflow that can be transferred to ungauged basins. While the width function and hypsometric curve were used to assess hydrological similarity, the approach is flexible enough to allow the use of other similar catchment properties. This use of unsupervised learning and functional data reduction to identify hydrologically similar basins at large scales is consistent with the growing recognition in the hydrological community regarding the use of explainable artificial intelligence techniques that build upon conceptual and machine learning models (Maksymiuk et al., 2020; Althoff et al., 2021).

Taken together, these results take strides in improving current approaches for accessing hydrologic response in ungauged basins. In order to model unit response, basins are often conceptually divided into channel networks and hillslope. Many of these models assume that the travel time from the hillslope to the stream network is negligible. However, due to the longer time for the sub-surface flow component to reach the stream network, this assumption can lead to erroneous results (Mesa and Mifflin, 1986; Gupta et al., 1986). Previous studies have already attempted to incorporate CTI to represent different saturation conditions in watersheds, resulting in a rescaled width function formulation (D’Odorico and Rigon, 2003; Rigon et al., 2001). Additionally, studies have attempted to develop an expression for velocity that can be included in the GIUH model (Al-Wagdany and Rao, 1997; Chen et al., 2019). However, attempts have yet to be made to directly include the hypsometric curve in a GIUH formulation. With robust functional formulations for width function and hypsometric curve now available, an analytical formulation of the GIUH model that goes beyond the moments of the width function and incorporates the hypsometric curve and CTI can now be pursued. While this study sought to lay out the groundwork in this regard, more work is needed to improve our understanding between catchment metrics and hydrological behavior. For instance,

other metrics such as single-valued catchment properties as well as metrics other than catchment-based need to be evaluated. Nonetheless, this study demonstrates the potential for using the width function, hypsometric curve, and CTI as three components of a response model while also providing a methodology for inspecting other metrics. Furthermore, a computationally efficient approach for utilizing these metrics for hydrologic similarity assessment is presented that can be implemented at large scales.

This study opens a multitude of future research avenues. A variety of other metrics, including single-valued topographic metrics and other metrics related to land use, land cover, soil, geology, etc. can be tested regarding their relationships with stream flow response. While these catchment properties do indeed play vital roles in modulating hydrologic behavior, this study limits the scope to only topographic metrics in order to build a parsimonious model that relies only on a widely available dataset. A hydrologic model that maximally extracts topographic controls enable an expansion in several other topics. First, the stream-sediment balance which includes an interplay between sediment load, sediment size, stream discharge, and stream slope (Wilcock et al., 2009) can be further investigated through topographic metrics that dictate stream discharge. Similarly, the affect of changing catchment topographies over the landscape history can influence the evolution of hydrologic regimes which can be studied in the context of landscape memory (Jain et al., 2020). Finally, for a more conclusive assertion regarding the results in this study, these test we performed in Walnut Gulch need to be replicated in other climatic regions and in watersheds with different landscapes and geomorphology. Our results might differ in complex landscapes shaped by tectonic effects and glaciation and where flow is altered by large storages or snow. Furthermore, a sensitivity analysis can be performed to check the robustness of these results based on changes in data quality and GIS analysis. This include

testing the results of various DEM resolutions, stream delineation algorithms, and stream thresholds.

REFERENCES

- Abrahams, A. D. (1984). Channel networks: a geomorphological perspective. *Water resources research* 20(2), 161–188.
- Abraham, R. J. and L. See (2000). Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological processes* 14(11-12), 2157–2172.
- Achard, F., H. D. Eva, H.-J. Stibig, P. Mayaux, J. Gallego, T. Richards, and J.-P. Malingreau (2002). Determination of deforestation rates of the world’s humid tropical forests. *Science* 297(5583), 999–1002.
- Acreman, M. and J. Holden (2013). How wetlands affect floods. *Wetlands* 33(5), 773–786.
- Al-Wagdany, A. and A. Rao (1997). Estimation of the velocity parameter of the geomorphologic instantaneous unit hydrograph. *Water resources management* 11(1), 1–16.
- Alexander, L., B. Autrey, J. DeMeester, K. Fritz, H. Golden, D. Goodrich, W. Kepner, C. Lane, S. LeDuc, S. Leibowitz, et al. (2015). Connectivity of streams and wetlands to downstream waters: A review and synthesis of the scientific evidence. *US Environmental Protection Agency, Washington DC*.
- Almeida, J., L. Barbosa, A. Pais, and S. Formosinho (2007). Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. *Chemometrics and Intelligent Laboratory Systems* 87(2), 208–217.
- Althoff, D., H. C. Bazame, and J. G. Nascimento (2021). Untangling hybrid hydrological models with explainable artificial intelligence. *H2Open Journal* 4(1), 13–28.
- Aluja, Tomas, M.-A. S. G. (2018). *Principal Component Analysis for Data Science*. Addison-Wesley.
- Archfield, S. A. and R. M. Vogel (2010). Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. *Water Resources Research* 46(10).
- Aryal, S. K., E. M. O’Loughlin, and R. G. Mein (2002). A similarity approach to predict landscape saturation in catchments. *Water Resources Research* 38(10), 26–1.
- Asquith, W. H. and M. C. Roussel (2009). Regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using an L-moment-based, PRESS-minimized, residual-adjusted approach. Technical report, US Geological Survey.

- Assessment, M. E. et al. (2005). *Ecosystems and Human Well-being: Our Human Planet-Summary for Decision-makers*.
- Athira, P., K. Sudheer, R. Cibin, and I. Chaubey (2016). Predictions in ungauged basins: An approach for regionalization of hydrological models considering the probability distribution of model parameters. *Stochastic Environmental Research and Risk Assessment* 30(4), 1131–1149.
- Azzalini, A. (1985). A class of distributions which includes the Normal ones. *Scandinavian Journal of Statistics* 12(2), 171–178.
- Bai, Y., Z. Chen, J. Xie, and C. Li (2016). Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models. *Journal of Hydrology* 532, 193–206.
- Bajracharya, P. and S. Jain (2020). Estimation of watershed width function: A statistical approach using LiDAR data. *Stochastic Environmental Research and Risk Assessment*, 1–15.
- Bajracharya, P. and S. Jain (2021). Characterization of drainage basin hypsometry: A generalized approach. *Geomorphology*, 107645.
- Bajracharya, P. and S. Jain (2022). Hydrologic similarity based on width function and hypsometry: An unsupervised learning approach. *Computers & Geosciences*, 105097.
- Balek, J. (2011). *Hydrology and water resources in tropical Africa*. Elsevier.
- Becker, M., F. Papa, F. Frappart, D. Alsdorf, S. Calmant, J. S. Da Silva, C. Prigent, and F. Seyler (2018). Satellite-based estimates of surface water dynamics in the Congo River Basin. *International Journal of Applied Earth Observation and Geoinformation* 66, 196–209.
- Belknap, D., R. Shipp, and J. Kelley (1986). Depositional setting and quaternary stratigraphy of the sheepscot estuary, maine: a preliminary report. *Géographie physique et Quaternaire* 40(1), 55–69.
- Belknap, D. F., J. T. Kelley, and A. M. Gontz (2002). Evolution of the glaciated shelf and coastline of the northern Gulf of Maine, USA. *Journal of Coastal Research* (36), 37–55.
- Bennett, G., J. Cassin, and N. Carroll (2016). Natural infrastructure investment and implications for the nexus: A global overview. *Ecosystem Services* 17, 293–297.
- Bernard, E. (1945). Le climat écologique de la cuvette centrale congolaise. *Inst. Nat. Etude Agron. Congo Belge, in 4*.
- Betterle, A. and G. Botter (2021). Does catchment nestedness enhance hydrological similarity? *Geophysical Research Letters* 48(13), e2021GL094148.

- Beven, K. (1979). On the generalized kinematic routing method. *Water Resources Research* 15(5), 1238–1242.
- Beven, K. (1997). Topmodel: A critique. *Hydrological processes* 11(9), 1069–1085.
- Beven, K. J. (2011). *Rainfall-runoff modelling: The Primer*. John Wiley & Sons.
- Beven, K. J. and M. J. Kirkby (1979). A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological sciences journal* 24(1), 43–69.
- Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society* 35, 99–109.
- Bian, L., A. M. Melesse, A. S. Leon, V. Verma, and Z. Yin (2021). A deterministic topographic wetland index based on LiDAR-Derived DEM for delineating open-water wetlands. *Water* 13(18), 2487.
- Biron, P. M., G. Choné, T. Buffin-Bélanger, S. Demers, and T. Olsen (2013). Improvement of streams hydro-geomorphological assessment using LiDAR DEMs. *Earth Surface Processes and Landforms* 38(15), 1808–1821.
- Blöschl, G., G. Bloschl, M. Sivapalan, T. Wagener, H. Savenije, and A. Viglione (2013). *Runoff prediction in ungauged basins: synthesis across processes, places and scales*. Cambridge University Press.
- Booij, M. J., D. Deckers, T. H. Rientjes, and M. S. Krol (2007). Regionalization for uncertainty reduction in flows in ungauged basins. *IAHS Publication* 313, 329.
- Brabets, T. P. and M. A. Walvoord (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.
- Brabets, T. P., B. Wang, and R. H. Meade (2000). Environmental and hydrologic overview of the Yukon River Basin, Alaska and Canada. *Water-Resources Investigations Report* 99, 4204.
- Brady, M. (2007). Major land uses. *USDA Economic Research Service*.
- Bras, R. L. (1990). *Hydrology: An Introduction to Hydrologic Science*. Addison Wesley Publishing Company.
- Brauman, K. A., G. C. Daily, T. K. Duarte, and H. A. Mooney (2007). The nature and value of ecosystem services: an overview highlighting hydrologic services. *Annu. Rev. Environ. Resour.* 32, 67–98.
- Bridgham, S. D., J. P. Megonigal, J. K. Keller, N. B. Bliss, and C. Trettin (2006). The carbon balance of north american wetlands. *Wetlands* 26(4), 889–916.

- Brismar, A. (2002). River systems as providers of goods and services: a basis for comparing desired and undesired effects of large dam projects. *Environmental management* 29(5), 598–609.
- Brooks, E., D. J. Allen, and W. R. Darwall (2011). *The status and distribution of freshwater biodiversity in Central Africa*. IUCN.
- Brown, R. D. and R. O. Braaten (1998). Spatial and temporal variability of Canadian monthly snow depths, 1946–1995. *Atmosphere-Ocean* 36(1), 37–54.
- Brugman, M. M., P. Raistrick, and A. Pietroniro (1997). Glacier related impacts of doubling atmospheric carbon dioxide concentrations on British Columbia and Yukon. *Responding to global climate change in British Columbia and Yukon* 1, 6–1.
- Brunner, M. I., R. Furrer, A. E. Sikorska, D. Viviroli, J. Seibert, and A.-C. Favre (2018). Synthetic design hydrographs for ungauged catchments: A comparison of regionalization methods. *Stochastic Environmental Research and Risk Assessment* 32(7), 1993–2023.
- Budyko, M. I., D. H. Miller, and D. H. Miller (1974). *Climate and life*, Volume 508. Academic press New York.
- Burn, D. H. and D. B. Boorman (1993). Estimation of hydrological parameters at ungauged catchments. *Journal of Hydrology* 143(3), 429–454.
- Carey, S. K. and M.-k. Woo (2001). Slope runoff processes and flow generation in a subarctic, subalpine catchment. *Journal of hydrology* 253(1-4), 110–129.
- Carpenter, S. R., R. DeFries, T. Dietz, H. A. Mooney, S. Polasky, W. V. Reid, and R. J. Scholes (2006). Millennium ecosystem assessment: research needs.
- Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.
- Chang, H., G. Johnson, T. Hinkley, and I.-W. Jung (2014). Spatial analysis of annual runoff ratios and their variability across the contiguous US. *Journal of Hydrology* 511, 387–402.
- Chapman, W. L. and J. E. Walsh (1993). Recent variations of sea ice and air temperature in high latitudes. *Bulletin of the American Meteorological Society* 74(1), 33–48.
- Chen, Y., P. Shi, X. Ji, S. Qu, L. Zhao, and F. Dong (2019). New method to calculate the dynamic factor–flow velocity in geomorphologic instantaneous unit hydrograph. *Scientific reports* 9(1), 1–13.
- Cheng, B.-L. M. (1982). *A Study of Geomorphologic Instantaneous Unit Hydrograph*. Ph. D. thesis, Department of Civil Engineering, University of Illinois at Urbana-Champaign, USA, ProQuest Dissertations and Theses.

- Chorley, R. J. and L. Morley (1959). A simplified approximation for the hypsometric integral. *The Journal of Geology* 67(5), 566–571.
- Chow, V. T., D. R. Maidment, and L. W. Mays (1988). *Applied Hydrology*. McGraw-Hill.
- Coates, D. R. et al. (1958). Quantitative geomorphology of small drainage basins of Southern Indiana. *Technical Report (Columbia University. Department of Geology); no. 10*.
- Cohen, M. J., I. F. Creed, L. Alexander, N. B. Basu, A. J. Calhoun, C. Craft, E. D’Amico, E. DeKeyser, L. Fowler, H. E. Golden, et al. (2016). Do geographically isolated wetlands influence landscape functions? *Proceedings of the National Academy of Sciences* 113(8), 1978–1986.
- Cohen, S., G. Willgoose, and G. Hancock (2008). A methodology for calculating the spatial distribution of the area–slope equation and the hypsometric integral within a catchment. *Journal of Geophysical Research: Earth Surface* 113(F3).
- Collier, P., G. Conway, and T. Venables (2008). Climate change and Africa. *Oxford Review of Economic Policy* 24(2), 337–353.
- Coradini, M., M. Fulchignoni, and F. Visicchio (1980). The hypsometric curve of Mars. *The Moon and the Planets* 22(2), 201–210.
- Council, N. R. et al. (2001). *Compensating for wetland losses under the Clean Water Act*. National Academies Press.
- Cragg, M., C. Polek, and S. Polasky (2011). Valuing properties with wetland potential. *Appraisal Journal* 79(2).
- Crave, A. and C. Gascuel-Oudou (1997). The influence of topography on time and space distribution of soil surface water content. *Hydrological processes* 11(2), 203–210.
- Daily, G. C., T. Söderqvist, S. Aniyar, K. Arrow, P. Dasgupta, P. R. Ehrlich, C. Folke, A. Jansson, B.-O. Jansson, N. Kautsky, et al. (2000). The value of nature and the nature of value. *science* 289(5478), 395–396.
- Davidson, N. C. (2014). How much wetland has the world lost? long-term and recent trends in global wetland area. *Marine and Freshwater Research* 65(10), 934–941.
- Dawson, C. W. and R. Wilby (1998). An artificial neural network approach to rainfall-runoff modelling. *Hydrological Sciences Journal* 43(1), 47–66.
- De Groot, R. S., M. A. Wilson, and R. M. Boumans (2002). A typology for the classification, description and valuation of ecosystem functions, goods and services. *Ecological economics* 41(3), 393–408.

- De Wit, M. J., F. Guillocheau, and M. C. De Wit (2015). *Geology and resource potential of the Congo Basin*. Springer Science & Business Media.
- DeFries, R. and C. Rosenzweig (2010). Toward a whole-landscape approach for sustainable land use in the tropics. *Proceedings of the National Academy of Sciences* 107(46), 19627–19632.
- Degetto, M., C. Gregoretti, and M. Bernard (2015). Comparative analysis of the differences between using LiDAR and contour-based DEMs for hydrological modeling of runoff generating debris flows in the Dolomites. *Frontiers in Earth Science* 3.
- Dibike, Y. B., D. Solomatine, and M. B. Abbott (1999). On the encapsulation of numerical-hydraulic models in artificial neural network. *Journal of Hydraulic research* 37(2), 147–161.
- Diebel, M. W., J. T. Maxted, P. J. Nowak, and M. J. Vander Zanden (2008). Landscape planning for agricultural nonpoint source pollution reduction I: a geographical allocation framework. *Environmental management* 42(5), 789–802.
- D’Odorico, P. and R. Rigon (2003). Hillslope and channel contributions to the hydrologic response. *Water resources research* 39(5).
- Dong, P. and Q. Chen (2017). *LiDAR remote sensing and applications*. CRC Press.
- Duan, S., P. Ullrich, and L. Shu (2020). Using convolutional neural networks for streamflow projection in California. *Front. Water* 2: 28.
- Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, M. L. Stiassny, et al. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews* 81(2), 163–182.
- Dudley, R. W. (2015). Regression equations for monthly and annual mean and selected percentile streamflows for ungaged rivers in Maine. Technical report, US Geological Survey.
- Dudov, S. (2017). Modeling of species distribution with the use of topography and remote sensing data on the example of vascular plants of the Tukuringra Ridge low mountain belt (Zeya State Nature Reserve, Amur Oblast). *Biology Bulletin Reviews* 7(3), 246–257.
- Ellouze-Gargouri, E. and Z. Bargaoui (2012). Runoff estimation for an ungauged catchment using geomorphological instantaneous unit hydrograph (GIUH) and copulas. *Water Resources Management* 26(6), 1615–1638.
- Elzhov, T. V., K. M. Mullen, A.-N. Spiess, and B. Bolker (2016). *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*. R package version 1.2-1.

- Everitt, B. S., S. Landau, M. Leese, and D. Stahl (2011). Cluster analysis 5th ed.
- Falkenmark, M., L. Gottschalk, J. Lundqvist, and P. Wouters (2004). Towards integrated catchment management: increasing the dialogue between scientists, policy-makers and stakeholders. *International Journal of Water Resources Development* 20(3), 297–309.
- Fan, S.-K. S., H.-K. Huang, and Y.-J. Chang (2013). Robust multivariate control chart for outlier detection using hierarchical cluster tree in SW2. *Quality and Reliability Engineering International* 29(7), 971–985.
- Flight, M. J., R. Paterson, K. Doiron, and S. Polasky (2012). Valuing wetland ecosystem services: A case study of delaware. *National Wetlands Newsletter* 34(5), 16–20.
- Freeman, B. G., M. N. Scholer, V. Ruiz-Gutierrez, and J. W. Fitzpatrick (2018). Climate change causes upslope shifts and mountaintop extirpations in a tropical bird community. *Proceedings of the National Academy of Sciences* 115(47), 11982–11987.
- Gallant, A., E. Binnian, J. Omernik, and M. Shasby (1995). Ecoregions of Alaska. *US Department of Interior: Washington, DC, USA*, 1567.
- Gallen, S. F., K. W. Wegmann, K. L. Frankel, S. Hughes, R. Q. Lewis, N. Lyons, P. Paris, K. Ross, J. B. Bauer, and A. C. Witt (2011). Hillslope response to knickpoint migration in the Southern Appalachians: Implications for the evolution of post-orogenic landscapes. *Earth Surface Processes and Landforms* 36(9), 1254–1267.
- Gao, H., M. Hrachowitz, F. Fenicia, S. Gharari, and H. Savenije (2014). Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China. *Hydrology and Earth System Sciences* 18(5), 1895–1915.
- Ge, S., D. Yang, and D. L. Kane (2013). Yukon River Basin long-term (1977–2006) hydrologic and climatic analysis. *Hydrological Processes* 27(17), 2475–2484.
- Gedan, K. B., M. L. Kirwan, E. Wolanski, E. B. Barbier, and B. R. Silliman (2011). The present and future role of coastal wetland vegetation in protecting shorelines: answering recent challenges to the paradigm. *Climatic change* 106(1), 7–29.
- Gharari, S., M. Hrachowitz, F. Fenicia, and H. Savenije (2011). Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment. *Hydrology and Earth System Sciences* 15(11), 3275–3291.
- Gibbs, H. K., A. S. Ruesch, F. Achard, M. K. Clayton, P. Holmgren, N. Ramankutty, and J. A. Foley (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences* 107(38), 16732–16737.

- Gibbs, J. P. (2000). Wetland loss and biodiversity conservation. *Conservation biology* 14(1), 314–317.
- Gittman, R. K., C. J. Baillie, K. K. Arkema, R. O. Bennett, J. Benoit, S. Blich, J. Brun, A. Chatwin, A. Colden, A. Dausman, et al. (2019). Voluntary restoration: mitigation’s silent partner in the quest to reverse coastal wetland loss in the usa. *Frontiers in Marine Science*, 511.
- Gleason, R. A., B. A. Tangen, M. K. Laubhan, K. E. Kermes, and N. H. Euliss Jr (2007). Estimating water storage capacity of existing and potentially restorable wetland depressions in a subbasin of the red river of the north. *USGS Northern Prairie Wildlife Research Center*, 89.
- Goodrich, D., P. Hellman, S. Moran, J. Garbrecht, D. Marks, D. Bosch, J. Steiner, J. Sadler, M. Romkens, D. Harmel, et al. (2016). The USDA-ARS experimental watershed network-evolution, lessons learned, and moving forward. In *In: Stringer, Christina E.; Krauss, Ken W.; Latimer, James S., eds. 2016. Headwaters to estuaries: advances in watershed science and management-Proceedings of the Fifth Interagency Conference on Research in the Watersheds. March 2-5, 2015, North Charleston, South Carolina. e-General Technical Report SRS-211. Asheville, NC: US Department of Agriculture Forest Service, Southern Research Station. 302 p., Volume 211, pp. 53–53.*
- Gosain, A., S. Rao, and D. Basuray (2006). Climate change impact assessment on hydrology of Indian river basins. *Current Science*, 346–353.
- Gottschalk, L., I. Krasovskaia, E. Leblois, and E. Sauquet (2006). Mapping mean and variance of runoff in a river basin. *Hydrology and Earth System Sciences* 10(4), 469–484.
- Gottschalk, L., E. Leblois, and J. Skøien (2011). Correlation and covariance of runoff revisited. *Journal of Hydrology* 398(1-2), 76–90.
- Gotvald, A. J., N. A. Barth, A. G. Veilleux, and C. Parrett (2012). Methods for determining magnitude and frequency of floods in California, based on data through water year 2006. *US Geological Survey Scientific Investigations Report 5113*, 38.
- Grant, J. and C. Fortezzo (2003a). Basin hypsometry on the Earth, Mars, and the Moon. *Sixth International Conference on Mars*, 3050.
- Grant, J. and C. Fortezzo (2003b). Hypsometric analyses of martian basins. *LPI*, 1123.
- Grimaldi, S., A. Petroselli, G. Alonso, and F. Nardi (2010). Flow time estimation with spatially variable hillslope velocity in ungauged basins. *Advances in Water Resources* 33(10), 1216–1223.

- Grimaldi, S., A. Petroselli, and F. Nardi (2012). A parsimonious geomorphological unit hydrograph for rainfall–runoff modelling in small ungauged basins. *Hydrological Sciences Journal* 57(1), 73–83.
- Group, E. S. W. et al. (1995). A national ecological framework for Canada (report and national map): Ottawa. *Hull, Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch, scale 1(7,500,000)*.
- Guha, S. and V. Jain (2020). Role of inherent geological and climatic characteristics on landscape variability in the tectonically passive Western Ghat, India. *Geomorphology* 350, 106840.
- Guha, S., R. Rastogi, and K. Shim (1998). Cure: An efficient clustering algorithm for large databases. *ACM Sigmod record* 27(2), 73–84.
- Gupta, H., G. J. Chakrapani, K. Selvaraj, and S.-J. Kao (2011). The fluvial geochemistry, contributions of silicate, carbonate and saline–alkaline components to chemical weathering flux and controlling parameters: Narmada River (deccan traps), India. *Geochimica et Cosmochimica Acta* 75(3), 800–824.
- Gupta, H. V., T. Wagener, and Y. Liu (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes: An International Journal* 22(18), 3802–3813.
- Gupta, P. K., S. Panigrahy, and J. S. Parihar (2011). Impact of climate change on runoff of the major river basins of India using global circulation model (HadCM3) projected data. *Journal of the Indian Society of Remote Sensing* 39(3), 337.
- Gupta, V. K. and O. J. Mesa (1988). Runoff generation and hydrologic response via channel network geomorphology—Recent progress and open problems. *Journal of Hydrology* 102(1), 3–28.
- Gupta, V. K. and E. Waymire (1983). On the formulation of an analytical approach to hydrologic response and similarity at the basin scale. *Journal of Hydrology* 65(1), 95–123.
- Gupta, V. K., E. Waymire, and I. Rodríguez-Iturbe (1986). On scales, gravity and network structure in basin runoff. In *Scale problems in hydrology*, pp. 159–184. Springer.
- Gupta, V. K., E. Waymire, and C. T. Wang (1980). A representation of an instantaneous unit hydrograph from geomorphology. *Water Resources Research* 16(5), 855–862.
- Haan, C. and H. Johnson (1966). Rapid determination of hypsometric curves. *Geological Society of America Bulletin* 77(1), 123–126.

- Haddad, K. and A. Rahman (2011). Selection of the best fit flood frequency distribution and parameter estimation procedure: a case study for tasmania in australia. *Stochastic Environmental Research and Risk Assessment* 25(3), 415–428.
- Hailegeorgis, T. T., Y. S. Abdella, K. Alfredsen, and S. Kolberg (2015). Evaluation of regionalization methods for hourly continuous streamflow simulation using distributed models in boreal catchments. *Journal of Hydrologic Engineering* 20(11), 04015028.
- Hallema, D. W. and R. Moussa (2014). A model for distributed GIUH-based flow routing on natural and anthropogenic hillslopes. *Hydrological Processes* 28(18), 4877–4895.
- Halsted, M. (2002). The sheepscot river, atlantic salmon and dams: A historical reflection. *Alna ME: Sheepscot Valley Conservation Association*.
- Hansen, L. (2006). *Better targeting, better outcomes*, Volume 2. USDA Economic Research Serv.
- Harlin, J. M. (1978). Statistical moments of the hypsometric curve and its density function. *Journal of the International Association for Mathematical Geology* 10(1), 59–72.
- Harlin, J. M. (1980). The effect of precipitation variability on drainage basin morphometry. *American Journal of Science* 280(8), 812–825.
- Harlin, J. M. (1984). Watershed morphometry and time to hydrograph peak. *Journal of Hydrology* 67(1-4), 141–154.
- Hebson, C. and E. F. Wood (1982). A derived flood frequency distribution using horton order ratios. *Water Resources Research* 18(5), 1509–1518.
- Hefting, M., J.-C. Clement, D. Dowrick, A.-C. Cosandey, S. Bernal, C. Cimpian, A. Tatur, T. Burt, and G. Pinay (2004). Water table elevation controls on soil nitrogen cycling in riparian wetlands along a european climatic gradient. *Biogeochemistry* 67(1), 113–134.
- Highfield, W. E. and S. D. Brody (2006). Price of permits: Measuring the economic impacts of wetland development on flood damages in florida. *Natural Hazards Review* 7(3), 123–130.
- Hooke, R. L. and D. A. Winski (2014). Unroofing Maine: Relating pressure of crystallization, thermochronological data, tectonics, and topography. *Geomorphology* 210, 36–47.
- Horton, R. E. (1932). Drainage-basin characteristics. *Eos, Transactions of the American Geophysical Union* 13(1), 350–361.

- Horton, R. E. (1945). Erosional development of streams and their drainage basins; Hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin* 56(3), 275–370.
- Horvath, E. K., J. R. Christensen, M. H. Mehaffey, and A. C. Neale (2017). Building a potential wetland restoration indicator for the contiguous United States. *Ecological indicators* 83, 463–473.
- Houser, P., D. Goodrich, and K. Syed (2000). Runoff, precipitation and soil moisture at Walnut Gulch. *Spatial patterns in catchment hydrology: Observations and modelling*, 125–157.
- Howard, A. D. (1990). Role of hypsometry and planform in basin hydrologic response. *Hydrological Processes* 4(4), 373–385.
- Hrachowitz, M., H. Savenije, G. Blöschl, J. McDonnell, M. Sivapalan, J. Pomeroy, B. Arheimer, T. Blume, M. Clark, U. Ehret, et al. (2013). A decade of predictions in ungauged basins (pub)—a review. *Hydrological sciences journal* 58(6), 1198–1255.
- Hsu, K.-l., H. V. Gupta, and S. Sorooshian (1995). Artificial neural network modeling of the rainfall-runoff process. *Water resources research* 31(10), 2517–2530.
- Huang, S., C. Young, M. Feng, K. Heidemann, M. Cushing, D. M. Mushet, and S. Liu (2011). Demonstration of a conceptual model for using lidar to improve the estimation of floodwater mitigation potential of prairie pothole region wetlands. *Journal of Hydrology* 405(3-4), 417–426.
- Huang, X. and J. D. Niemann (2008a). How do streamflow generation mechanisms affect watershed hypsometry? *Earth Surface Processes and Landforms* 33(5), 751–772.
- Huang, X. and J. D. Niemann (2008b). How do streamflow generation mechanisms affect watershed hypsometry? *Earth Surface Processes and Landforms* 33(5), 751–772.
- Imamura, G. (1937). *Past glaciers and the present topography of the Japanese Alps*. Geographical Institute, Tokyo Bunrika Daigaku.
- Jain, A., M. Murty, and P. Flynn (1999). Data clustering: A review. *ACM Computing Surveys* 31(3), 264–323.
- Jain, V., A. S. Sonam, A. Singh, R. Sinha, S. Tandon, et al. (2020). Evolution of modern river systems: an assessment of ‘landscape memory’ in Indian river systems. *Episodes* 43(1), 535–551.

- Jones, C. N., G. R. Evenson, D. L. McLaughlin, M. K. Vanderhoof, M. W. Lang, G. W. McCarty, H. E. Golden, C. R. Lane, and L. C. Alexander (2018). Estimating restorable wetland water storage at landscape scales. *Hydrological processes* 32(2), 305–313.
- Jones, M. (2009). Kumaraswamy’s distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology* 6(1), 70–81.
- Julie, L., N. Pierre, P. Albert, S. Garba, T.-w. Nadji, B. Alain, H. Michel, et al. (2008). Integrated water resources management in the Congo basin based on the development of Earth observation monitoring systems in the framework of the AMESD programme in Central Africa.
- Jung, Y., H. Park, D.-Z. Du, and B. L. Drake (2003). A decision criterion for the optimal number of clusters in hierarchical clustering. *Journal of Global Optimization* 25(1), 91–111.
- Kadima, E., D. Delvaux, S. Sebagenzi, L. Tack, and S. Kabeya (2011). Structure and geological history of the Congo Basin: An integrated interpretation of gravity, magnetic and reflection seismic data. *Basin Research* 23(5), 499–527.
- Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology* 15(1), 52–60.
- Karandish, F. and J. Šimnek (2016). A comparison of numerical and machine-learning modeling of soil water content with limited input data. *Journal of Hydrology* 543, 892–909.
- Karp, S. and L. B. Stotts (2012). *Fundamentals of electro-optic systems design: communications, LiDAR, and imaging*. Cambridge University Press.
- Kaufman, L. and P. J. Rousseeuw (2009). *Finding groups in data: an introduction to cluster analysis*, Volume 344. John Wiley & Sons.
- Keppel, R. V. and K. Renard (1962). Transmission losses in ephemeral stream beds. *Journal of the Hydraulics Division* 88(3), 59–68.
- Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā: The Indian Journal of Statistics, Series A* 62, 49–66.
- Kessler, A. C. and S. C. Gupta (2015). Drainage impacts on surficial water retention capacity of a prairie pothole watershed. *JAWRA Journal of the American Water Resources Association* 51(4), 1101–1113.
- Kirkby, M. (1975). *Hydrograph Modelling Strategies*.
- Kirkby, M. (1976). Tests of the random network model, and its application to basin hydrology. *Earth Surface Processes* 1(3), 197–212.

- Kirshen, D. M. and R. L. Bras (1983). The linear channel and its effect on the geomorphologic IUH. *Journal of Hydrology* 65(1-3), 175–208.
- Kleidon, A., E. Zehe, U. Ehret, and U. Scherer (2013). Thermodynamics, maximum power, and the dynamics of preferential river flow structures at the continental scale. *Hydrology and Earth System Sciences* 17(1), 225–251.
- Krleža, D., B. Vrdoljak, and M. Brčić (2021). Statistical hierarchical clustering algorithm for outlier detection in evolving data streams. *Machine Learning* 110(1), 139–184.
- Kumar, R., C. Chatterjee, S. Kumar, A. Lohani, R. Singh, and R. Nema (2002). Development of GIS based GIUH model for an ungauged catchment. In *ACE 2002: International Conference on Advances in Civil Engineering*, pp. 8.
- Langbein, W. B. (1947). Topographic characteristics of drainage basins. Technical Report Water Supply Paper 968- C, United States Geological Survey.
- Laraque, A., J. P. Bricquet, A. Pandi, and J. C. Olivry (2009). A review of material transport by the Congo River and its tributaries. *Hydrological Processes* 23(22), 3216–3224.
- Larsen, J. E., M. Sivapalan, N. A. Coles, and P. E. Linnet (1994). Similarity analysis of runoff generation processes in real-world catchments. *Water Resources Research* 30(6), 1641–1652.
- Lashermes, B. and E. Foufoula-Georgiou (2007). Area and width functions of river networks: New results on multifractal properties. *Water Resources Research* 43(9).
- Leturmy, P., F. Lucazeau, and F. Brigaud (2003). Dynamic interactions between the Gulf of Guinea passive margin and the Congo River drainage basin: 1. Morphology and mass balance. *Journal of Geophysical Research: Solid Earth* 108(B8).
- Leung, L. R. and M. S. Wigmosta (1999). Potential climate change impacts on mountain watersheds in the Pacific Northwest. *Journal of the American Water Resources Association* 35(6), 1463–1471.
- Liffner, J. W., G. A. Hewa, and M. C. Peel (2018). The sensitivity of catchment hypsometry and hypsometric properties to DEM resolution and polynomial order. *Geomorphology* 309, 112–120.
- Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory* 37(1), 145–151.
- Linkov, I. and E. Moberg (2011). *Multi-criteria decision analysis: environmental applications and case studies*. CRC Press.

- Lisitsyn, A. P. (1969). Recent sedimentation in the Bering Sea. *National Science Foundation, The Israel Program for Scientific Translations (translated from Russian), Washington, DC*.
- Liu, C., P. Frazier, L. Kumar, C. Macgregor, and N. Blake (2006). Catchment-wide wetland assessment and prioritization using the multi-criteria decision-making method TOPSIS. *Environmental management* 38(2), 316–326.
- Liu, H. and L. Wang (2008). Mapping detention basins and deriving their spatial attributes from airborne lidar data for hydrological applications. *Hydrological Processes: An International Journal* 22(13), 2358–2369.
- Liu, X. and Z. Zhang (2011). Drainage network extraction using LiDAR-derived DEM in volcanic plains. *Area* 43(1), 42–52.
- Liyanagamage, S. P. and G. A. Hewa (2012a). Link between flow regime and the catchment hypsometry: Analysis of South Australian basins. *Journal of Hydrologic Engineering* 17(12), 1287–1295.
- Liyanagamage, S. P. and G. A. Hewa (2012b). Link between flow regime and the catchment hypsometry: Analysis of South Australian basins. *Journal of Hydrologic Engineering* 17(12), 1287–1295.
- Loritz, R., A. Kleidon, C. Jackisch, M. Westhoff, U. Ehret, H. Gupta, and E. Zehe (2019). A topographic index explaining hydrological similarity by accounting for the joint controls of runoff formation. *Hydrology and Earth System Sciences* 23(9), 3807–3821.
- Loukika, K. N., V. R. Keesara, and V. Sridhar (2021). Analysis of land use and land cover using machine learning algorithms on Google Earth engine for Munneru River basin, India. *Sustainability* 13(24), 13758.
- Luo, W. (2000). Quantifying groundwater-sapping landforms with a hypsometric technique. *Journal of Geophysical Research: Planets* 105(E1), 1685–1694.
- Luo, W. (2002). Hypsometric analysis of Margaritifer Sinus and origin of valley networks. *Journal of Geophysical Research: Planets* 107(E10), 1–1.
- Luo, W. and J. M. Harlin (2003). A theoretical travel time based on watershed hypsometry. *JAWRA Journal of the American Water Resources Association* 39(4), 785–792.
- L’vovich, M. I. (1979). *World water resources and their future*. American Geophysical Union.
- Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik (2021). *Cluster: Cluster analysis basics and extensions*. R package version 2.1.1 — For new features, see the ‘Changelog’ file (in the package source).

- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of Indian National Science Academy* 2(1), 49–55.
- Maksymiuk, S., A. Gosiewska, and P. Biecek (2020). Landscape of R packages for explainable Artificial Intelligence. *arXiv preprint arXiv:2009.13248*.
- Malhi, Y. and J. Grace (2000). Tropical forests and atmospheric carbon dioxide. *Trends in Ecology & Evolution* 15(8), 332–337.
- Manabe, S., P. Milly, and R. Wetherald (2004). Simulated long-term changes in river discharge and soil moisture due to global warming/simulations à long terme de changements d'écoulement fluvial et d'humidité du sol causés par le réchauffement global. *Hydrological Sciences Journal* 49(4).
- Marani, M., E. Eltahir, and A. Rinaldo (2001). Geomorphic controls on regional base flow. *Water Resources Research* 37(10), 2619–2630.
- Mathur, Y. (2018). Comparing longitudinal profiles and hypsometric curves for sensitivity analysis of digital elevation model using Tecdem. In *80th EAGE Conference and Exhibition 2018*, Volume 2018, pp. 1–3. European Association of Geoscientists & Engineers.
- McGlynn, B. L. and J. Seibert (2003). Distributed assessment of contributing area and riparian buffering along stream networks. *Water resources research* 39(4).
- McIntyre, N., H. Lee, H. Wheeler, A. Young, and T. Wagener (2005). Ensemble predictions of runoff in ungauged catchments. *Water Resources Research* 41(12).
- McLachlan, G. J., S. X. Lee, and S. I. Rathnayake (2019). *Finite mixture models*, Volume 6.
- McLachlan, G. J. and D. Peel (2004). *Finite mixture models*. John Wiley & Sons.
- McLean, J., L. Sewall, and L. Pugh (2007). Sheepscot river watershed management plan. *Prepared for Time and Tide Resource Conservation and Development Area, Augusta, ME. Available online at <http://www.kcswcd.org/Projects/Sheepscot/WHOLE>* 20.
- McNamara, J. P., D. L. Kane, and L. D. Hinzman (1998). An analysis of streamflow hydrology in the Kuparuk River Basin, Arctic Alaska: A nested watershed approach. *Journal of Hydrology* 206(1-2), 39–57.
- Menges, C. M. and P. A. Pearthree (1989). Late Cenozoic tectonism in Arizona and its impact on regional landscape evolution. *Arizona Geological Society Digest* 17, 649–680.
- Mesa, O. J. and E. R. Miffin (1986). On the relative role of hillslope and network geometry in hydrologic response. In *Scale Problems in Hydrology*, pp. 1–17. Springer.

- Miller, V. C. (1953). Quantitative geomorphic study of drainage basin characteristics in the Clinch Mountain area, Virginia and Tennessee. *Technical report (Columbia University. Department of Geology); no. 3.*
- Milly, P. (1994). Climate, interseasonal storage of soil water, and the annual water balance. *Advances in Water Resources 17*(1-2), 19–24.
- Minns, A. and M. Hall (1996). Artificial neural networks as rainfall-runoff models. *Hydrological sciences journal 41* (3), 399–417.
- Mital, U., D. Dwivedi, J. B. Brown, B. Faybishenko, S. L. Painter, and C. I. Steefel (2020). Sequential imputation of missing spatio-temporal precipitation data using random forests. *Front. Water 2: 20.*
- Mitsch, W. J., B. Bernal, A. M. Nahlik, Ü. Mander, L. Zhang, C. J. Anderson, S. E. Jørgensen, and H. Brix (2013). Wetlands, carbon, and climate change. *Landscape Ecology 28*(4), 583–597.
- Mitsch, W. J., J. W. Day, J. W. Gilliam, P. M. Groffman, D. L. Hey, G. W. Randall, and N. Wang (2001). Reducing nitrogen loading to the gulf of mexico from the mississippi river basin: Strategies to counter a persistent ecological problem: Ecotechnology—the use of natural ecosystems to solve environmental problems—should be a part of efforts to shrink the zone of hypoxia in the gulf of mexico. *BioScience 51*(5), 373–388.
- Mitsch, W. J. and J. G. Gosselink (2015). *Wetlands*. John Wiley & Sons.
- Moglen, G. E. and R. L. Bras (1995). The importance of spatially heterogeneous erosivity and the cumulative area distribution within a basin evolution model. *Geomorphology 12*(3), 173–185.
- Molinario, G., M. Hansen, and P. V. Potapov (2015). Forest cover dynamics of shifting cultivation in the Democratic Republic of Congo: A remote sensing-based assessment for 2000–2010. *Environmental Research Letters 10*(9), 094009.
- Mondal, A., D. Khare, S. Kundu, P. K. Meena, P. Mishra, and R. Shukla (2015). Impact of climate change on future soil erosion in different slope, land use, and soil-type conditions in a part of the Narmada River Basin, India. *Journal of Hydrologic Engineering 20*(6), C5014003.
- Montgomery, D. R., G. Balco, and S. D. Willett (2001). Climate, tectonics, and the morphology of the Andes. *Geology 29*(7), 579–582.
- Mooney, H. A., P. R. Ehrlich, and G. Daily (1997). Ecosystem services: a fragmentary history. *Nature's services: Societal dependence on natural ecosystems*, 11–19.
- Moré, J. J. (1978). The Levenberg-Marquardt algorithm: Implementation and theory. In *Numerical analysis*, pp. 105–116. Springer.

- Moussa, R. (2008). What controls the width function shape, and can it be used for channel network comparison and regionalization? *Water Resources Research* 44(8).
- Müller, M. and S. Thompson (2015). TopREML: A topological restricted maximum likelihood approach to regionalize trended runoff signatures in stream networks. *Hydrology and Earth System Sciences* 19(6), 2925–2942.
- Murray, B. R., G. C. Hose, D. Eamus, and D. Licari (2006). Valuation of groundwater-dependent ecosystems: a functional methodology incorporating ecosystem services. *Australian Journal of Botany* 54(2), 221–229.
- Mushi, C., P. Ndomba, M. Trigg, R. Tshimanga, and F. Mtalo (2019). Assessment of basin-scale soil erosion within the Congo River Basin: A review. *Catena* 178, 64–76.
- Naden, P. S. (1992). Spatial variability in flood estimation for large catchments: The exploitation of channel network structure. *Hydrological Sciences Journal* 37(1), 53–71.
- Nazari-Sharabian, M., M. Taheriyoun, and M. Karakouzian (2020). Sensitivity analysis of the dem resolution and effective parameters of runoff yield in the SWAT model: a case study. *Journal of Water Supply: Research and Technology-Aqua* 69(1), 39–54.
- Nicholson, B., G. Hancock, S. Cohen, G. Willgoose, and O. Rey-Lescure (2013). An assessment of the fluvial geomorphology of subcatchments in Parana Valles, Mars. *Geomorphology* 183, 96–109.
- Nijssen, B., G. M. O'Donnell, A. F. Hamlet, and D. P. Lettenmaier (2001). Hydrologic sensitivity of global rivers to climate change. *Climatic Change* 50(1-2), 143–175.
- NOAA (2020). What is LIDAR?
<https://www.oceanservice.noaa.gov/facts/lidar.html>. 2020-04-09.
- Nobre, A. D., L. A. Cuartas, M. Hodnett, C. D. Rennó, G. Rodrigues, A. Silveira, and S. Saleska (2011). Height Above the Nearest Drainage—a hydrologically relevant new terrain model. *Journal of Hydrology* 404(1-2), 13–29.
- Office, M. S. P. (2001). Casco Bay watershed wetland characterization. Technical report.
- Oja, H. (2010). *Multivariate nonparametric methods with R: An approach based on spatial signs and ranks*. Springer Science & Business Media.
- Oppel, H. and B. Mewes (2020). On the automation of flood event separation from continuous time series. *Frontiers in Water* 2, 18.

- Osberg, P., A. Hussey, and G. Boone (1985). Bedrock geologic map of maine: Maine geological survey. *Department of Conservation 1* (500,000).
- Osterkamp, W. (2008). Geology, soils, and geomorphology of the Walnut Gulch Experimental Watershed, Tombstone, Arizona. *Journal of the Arizona-Nevada Academy of Science 40*(2), 136–154.
- Oudin, L., A. Kay, V. Andréassian, and C. Perrin (2010). Are seemingly physically similar catchments truly hydrologically similar? *Water Resources Research 46*(11).
- Palmer, M. A. (2009). Reforming watershed restoration: science in need of application and applications in need of science. *Estuaries and coasts 32*(1), 1–17.
- Pandey, B. K. and D. Khare (2018). Identification of trend in long term precipitation and reference evapotranspiration over Narmada River basin (India). *Global and Planetary Change 161*, 172–182.
- Patil, S. and M. Stieglitz (2011). Hydrologic similarity among catchments under variable flow conditions. *Hydrology and Earth System Sciences 15*(3), 989–997.
- Patil, S. and M. Stieglitz (2012). Controls on hydrologic similarity: Role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrology and Earth System Sciences 16*(2), 551–562.
- Pedraza, A., J. V. Pérez-Peña, J. Galindo-Zaldívar, J. M. Azañón, and A. Azor (2009). Testing the sensitivity of geomorphic indices in areas of low-rate active folding (eastern Betic Cordillera, Spain). *Geomorphology 105*(3-4), 218–231.
- Péguy, C.-P. (1942). Principes de morphométrie alpine. *Revue de géographie alpine 30*(3), 453–486.
- Peñacoba-Antona, L., M. Gómez-Delgado, and A. Esteve-Núñez (2021). Multi-criteria evaluation and sensitivity analysis for the optimal location of constructed wetlands (METland) at oceanic and Mediterranean Areas. *International Journal of Environmental Research and Public Health 18*(10), 5415.
- Pérez-Peña, J., J. Azañón, G. Booth-Rea, A. Azor, and J. Delgado (2009). Differentiating geology and tectonics using a spatial autocorrelation technique for the hypsometric integral. *Journal of Geophysical Research: Earth Surface 114*(F2).
- Pérez-Sánchez, J., J. Senent-Aparicio, F. Segura-Méndez, D. Pulido-Velazquez, and R. Srinivasan (2019). Evaluating hydrological models for deriving water resources in peninsular Spain. *Sustainability 11*(10).
- Peters-Lidard, C. D., M. Clark, L. Samaniego, N. E. Verhoest, T. Van Emmerik, R. Uijlenhoet, K. Achieng, T. E. Franz, and R. Woods (2017). Scaling, similarity, and the fourth paradigm for hydrology. *Hydrology and Earth System Sciences 21*(7), 3701–3713.

- Pike, R. J. and S. E. Wilson (1971, 04). Elevation–relief ratio, hypsometric integral, and geomorphic area–altitude analysis. *GSA Bulletin* 82(4), 1079–1084.
- Poppenga, S. K., B. B. Worstell, J. M. Stoker, and S. K. Greenlee (2010). Using selective drainage methods to extract continuous surface flow from 1-meter lidar-derived digital elevation data. Technical report, U. S. Geological Survey.
- Prates, M. O., C. R. B. Cabral, and V. H. Lachos (2013). mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software* 54(12), 1–20.
- Preston, E. M. and B. L. Bedford (1988). Evaluating cumulative effects on wetland functions: a conceptual overview and generic framework. *Environmental Management* 12(5), 565–583.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rampi, L. P., J. F. Knight, and C. F. Lenhart (2014). Comparison of flow direction algorithms in the application of the CTI for mapping wetlands in Minnesota. *Wetlands* 34(3), 513–525.
- Rao, A. R. and W. Tirtotjondro (1995). Computation of unit hydrographs by a Bayesian method. *Journal of Hydrology* 164(1-4), 325–344.
- Ray, L. K. and N. K. Goel (2019). Flood frequency analysis of Narmada River basin in India under nonstationary condition. *Journal of Hydrologic Engineering* 24(8), 05019018.
- Razavi, T. and P. Coulibaly (2013). Streamflow prediction in ungauged basins: Review of regionalization methods. *Journal of Hydrologic Engineering* 18(8), 958–975.
- Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh (2000). Conservation equations governing hillslope responses: Exploring the physical basis of water balance. *Water Resources Research* 36(7), 1845–1863.
- Reichl, J. P. C., A. W. Western, N. R. McIntyre, and F. H. S. Chiew (2009). Optimization of a similarity measure for estimating ungauged streamflow. *Water Resources Research* 45(10).
- Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, et al. (2019). Deep learning and process understanding for data-driven earth system science. *Nature* 566(7743), 195–204.
- Renard, K., L. Lane, J. Simanton, W. Emmerich, J. Stone, M. Wertz, D. Goodrich, and D. Yakowitz (1993). Agricultural impacts in an arid environment: Walnut Gulch studies. *Hydrological Science and Technology* 9(1-4), 145–190.

- Rennó, C. D., A. D. Nobre, L. A. Cuartas, J. V. Soares, M. G. Hodnett, and J. Tomasella (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment* 112(9), 3469–3481.
- Rigon, R., M. Bancheri, G. Formetta, and A. de Lavenne (2016). The geomorphological unit hydrograph from a historical-critical perspective. *Earth Surface Processes and Landforms* 41(1), 27–37.
- Rigon, R., A. Cozzini, and S. Pisoni (2001). Getting the rescaled width function and the derived WGIUH.
- Rinaldo, A. and I. Rodriguez-Iturbe (1996). Geomorphological theory of the hydrological response. *Hydrological processes* 10(6), 803–829.
- Rodríguez-Iturbe, I. and A. Rinaldo (2001). *Fractal river basins: Chance and self-organization*. Cambridge University Press.
- Rodriguez-Iturbe, I. and J. B. Valdes (1979). The geomorphologic structure of hydrologic response. *Water Resources Research* 15(6), 1409–1420.
- Rodríguez-Iturbe, I. and J. B. Valdés (1979). The geomorphologic structure of hydrologic response. *Water Resources Research* 15(6), 1409–1420.
- Rokach, L. and O. Maimon (2005). Clustering methods. In *Data mining and knowledge discovery handbook*, pp. 321–352. Springer.
- Rolander, N., A. Ceci, and M. Berdugo (2003). A framework for mcdm method selection. *Georgia Institute of Technology, Tech. Rep., 5th December*.
- Rosell, F., O. Bozser, P. Collen, and H. Parker (2005). Ecological impact of beavers castor fiber and castor canadensis and their ability to modify ecosystems. *Mammal review* 35(3-4), 248–276.
- Rosenblatt, P., P. Pinet, and E. Thouvenot (1994). Comparative hypsometric analysis of Earth and Venus. *Geophysical Research Letters* 21(6), 465–468.
- Rossi, R. J. (2018). *Mathematical statistics: An introduction to likelihood based inference*. John Wiley & Sons.
- Sahoo, B. and P. Saritha (2015). Estimating floods from an ungauged river basin using GIUH-based Nash model. In *ISFRAM 2014*, pp. 123–133. Springer.
- Sahoo, R. and V. Jain (2018). Sensitivity of drainage morphometry based hydrological response (GIUH) of a river basin to the spatial resolution of DEM data. *Computers & Geosciences* 111, 78–86.
- Sahu, R. K., J. Müller, J. Park, C. Varadharajan, B. Arora, B. Faybishenko, and D. Agarwal (2020). Impact of input feature selection on groundwater level prediction from a multi-layer perceptron neural network. *Frontiers in Water* 2, 46.

- Sankarasubramanian, A. and R. M. Vogel (2002). Annual hydroclimatology of the united states. *Water Resources Research* 38(6), 19–1.
- Sarkar, A. and P. Patel (2011). Topographic analysis of the Dulung River Basin. *The Indian Journal of Spatial Science* 2(1), 19.
- Sawicz, K., T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences* 15(9), 2895–2911.
- Schumm, S. A. (1956, 05). Evolution of drainage systems and slopes in Badlands at Perth Amboy, New Jersey. *GSA Bulletin* 67(5), 597–646.
- Schwarz, B. (2010). Mapping the world in 3D. *Nature Photonics* 4(7), 429–430.
- SCS, U. (1986). Urban hydrology for small watersheds, technical release no. 55 (TR-55). *US Department of Agriculture, US Government Printing Office, Washington, DC*.
- Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär (2006). Land–atmosphere coupling and climate change in europe. *Nature* 443(7108), 205–209.
- Seong, K.-W. and J. H. Sung (2021). Derivation of s-curve from oscillatory hydrograph using digital filter. *Water* 13(11), 1456.
- Shaw, D. A., G. Vanderkamp, F. M. Conly, A. Pietroniro, and L. Martz (2012). The fill–spill hydrology of prairie wetland complexes during drought and deluge. *Hydrological Processes* 26(20), 3147–3156.
- Sherman, L. K. (1932). Streamflow from rainfall by the unit-graph method. *Eng. News Record* 108, 501–505.
- Shook, K. R. and J. W. Pomeroy (2011). Memory effects of depression storage in northern prairie hydrology. *Hydrological Processes* 25(25), 3890–3898.
- Shulski, M. and G. Wendler (2007). *The Climate of Alaska*. University of Alaska Press.
- Singh, K. P. (1976). Unit hydrographs—a comparative study. *JAWRA Journal of the American Water Resources Association* 12(2), 381–392.
- Singh, O., A. Sarangi, and M. C. Sharma (2008). Hypsometric integral estimation methods and its relevance on erosion status of North-Western lesser Himalayan watersheds. *Water Resources Management* 22(11), 1545–1560.
- Sit, M., B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology* 82(12), 2635–2670.

- Sivapalan, M. (2003). Process complexity at hillslope scale, process simplicity at watershed scale: Is there a connection? In *EGS-AGU-EUG Joint Assembly*, pp. 7973.
- Sivapalan, M. (2006). Pattern, process and function: elements of a unified theory of hydrology at the catchment scale. *Encyclopedia of hydrological sciences*.
- Sivapalan, M., K. Beven, and E. F. Wood (1987). On hydrologic similarity: 2. a scaled model of storm runoff production. *Water Resources Research* 23(12), 2266–2278.
- Sivapalan, M., K. Takeuchi, S. Franks, V. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J. McDonnell, E. Mendiondo, P. O’Connell, et al. (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.
- Skøien, J. O. and G. Blöschl (2007). Spatiotemporal topological kriging of runoff time series. *Water Resources Research* 43(9).
- Skøien, J. O., R. Merz, and G. Blöschl (2006). Top-kriging-geostatistics on stream networks. *Hydrology and Earth System Sciences* 10(2), 277–287.
- Smith, G. W. and L. E. Hunter (1989). Late Wisconsinan deglaciation of coastal maine. *Studies in Maine Geology* 6, 13–32.
- Solomatine, D. P. and A. Ostfeld (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of hydroinformatics* 10(1), 3–22.
- Spence, C. (2006). Hydrological processes and streamflow in a lake dominated watercourse. *Hydrological Processes* 20(17), 3665–3681.
- Sreedevi, P., K. Subrahmanyam, and S. Ahmed (2005). The significance of morphometric analysis for obtaining groundwater potential zones in a structurally controlled terrain. *Environmental Geology* 47(3), 412–420.
- Sridhar, V., M. M. Billah, and J. W. Hildreth (2018). Coupled surface and groundwater hydrological modeling in a changing climate. *Groundwater* 56(4), 618–635.
- Ssegane, H., E. Tollner, Y. Mohamoud, T. Rasmussen, and J. Dowd (2012). Advances in variable selection methods II: Effect of variable selection method on classification of hydrologically similar watersheds in three Mid-Atlantic ecoregions. *Journal of Hydrology* 438, 26–38.
- State of Maine (1983). Title 12: Conservation, Part 2: Forests, parks, lakes and rivers, Chapter 200: Maine’s rivers, §403. Special protection for outstanding rivers.

- State of Maine (2018). Maine Office of GIS.
<http://https://www.maine.gov/megis/catalog/>. Accessed: 2019-05-05.
- Stone, J. J., M. H. Nichols, D. C. Goodrich, and J. Buono (2008). Long-term runoff database, Walnut Gulch Experimental Watershed, Arizona, United States. *Water resources research* 44(5).
- Strahler, A. N. (1952). Hypsometric (area–altitude) analysis of erosional topography. *Geological Society of America Bulletin* 63(11), 1117–1142.
- Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Eos, Transactions American Geophysical Union* 38(6), 913–920.
- Strahler, A. N. (1964). Part II. Quantitative geomorphology of drainage basins and channel networks. *Handbook of Applied Hydrology: McGraw-Hill, New York*, 4–39.
- Subramanian, V., V. Ittekkot, D. Unger, and N. Madhavan (2006). Silicate weathering in south Asian tropical river basins. *The Silicon Cycle: Human Perturbations and Impacts on Aquatic Systems* 66, 3–13.
- Sun, A. and G. Tang (2020). Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets. *Front. Water* 2: 536743.
- Surkan, A. J. (1969). Synthetic hydrographs: Effects of network geometry. *Water Resources Research* 5(1), 112–128.
- Swain, J., R. Jha, and K. Patra (2015). Stream flow prediction in a typical ungauged catchment using GIUH approach. *Aquatic Procedia* 4, 993–1000.
- Tardif, S., A. St-Hilaire, R. Roy, M. Bernier, and S. Payette (2009). Statistical properties of hydrographs in minerotrophic fens and small lakes in mid-latitude québec, canada. *Canadian Water Resources Journal* 34(4), 365–380.
- Tegegne, G., D. K. Park, and Y.-O. Kim (2017). Comparison of hydrological models for the assessment of water resources in a data-scarce region, the Upper Blue Nile River Basin. *Journal of Hydrology: Regional Studies* 14, 49–66.
- Temmerman, S., P. Meire, T. J. Bouma, P. M. Herman, T. Ysebaert, and H. J. De Vriend (2013). Ecosystem-based coastal defence in the face of global change. *Nature* 504(7478), 79–83.
- Tetzlaff, D., J. Seibert, and C. Soulsby (2009). Inter-catchment comparison to assess the influence of topography and soils on catchment transit times in a geomorphic province; the Cairngorm mountains, Scotland. *Hydrological Processes* 23(13), 1874–1886.
- Thomopoulos, N. T. (2017). Statistical Distributions. *Applications and Parameter Estimates*. Cham, Switzerland: Springer International Publishing.

- Thorslund, J., J. Jarsjo, F. Jaramillo, J. W. Jawitz, S. Manzoni, N. B. Basu, S. R. Chalov, M. J. Cohen, I. F. Creed, R. Goldenberg, et al. (2017). Wetlands as large-scale nature-based solutions: Status and challenges for research, engineering and management. *Ecological Engineering* 108, 489–497.
- Tibshirani, R., G. Walther, and T. Hastie (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2), 411–423.
- Troutman, B. M. and M. R. Karlinger (1984). On the expected width function for topologically random channel networks. *Journal of Applied Probability* 21(4), 836–849.
- Troutman, B. M. and M. R. Karlinger (1988). Asymptotic Rayleigh instantaneous unit hydrograph. *Stochastic Hydrology and Hydraulics* 2, 73–78.
- Troutman, B. M. and M. R. Karlinger (1989). Predictors of the peak width for networks with exponential links. *Stochastic Hydrology and Hydraulics* 3, 1–16.
- Tsai, F. T.-C. (2010). Bayesian model averaging assessment on groundwater management under model structure uncertainty. *Stochastic Environmental Research and Risk Assessment* 24(6), 845–861.
- Tshimanga, R. M. (2012). *Hydrological uncertainty analysis and scenario-based streamflow modelling for the Congo River basin*. Ph. D. thesis, Rhodes University.
- Tsybakov, A. B. (2008). *Introduction to nonparametric estimation*. Springer.
- Tung, Y.-K., K.-C. Yeh, and J.-C. Yang (1997). Regionalization of unit hydrograph parameters: 1. Comparison of regression analysis techniques. *Stochastic Hydrology and Hydraulics* 11, 145–171.
- USGS (1996). USGS EROS Archive – Digital Elevation – Global 30 Arc-Second Elevation (GTOPO30). <https://doi.org/10.5066/F7DF6PQS>. Accessed: 2019-06-02.
- Ussyshkin, V. and L. Theriault (2011). Airborne LiDAR: Advances in discrete return technology for 3D vegetation mapping. *Remote Sensing* 3(3), 416–434.
- Valdes, J. B., Y. Fiallo, and I. Rodríguez-Iturbe (1979). A rainfall–runoff analysis of the geomorphologic IUH. *Water Resources Research* 15(6), 1421–1434.
- Valéry, A., V. Andréassian, and C. Perrin (2014). ‘As simple as possible but not simpler’: What is useful in a temperature-based snow-accounting routine? Part 2 - Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. *Journal of Hydrology* 517, 1176–1187.
- Vanderwaal, J. A. and H. Ssegane (2013). Do polynomials adequately describe the hypsometry of Monadnock phase watersheds? *JAWRA Journal of the American Water Resources Association* 49(6), 1485–1495.

- Veneziano, D., G. E. Moglen, P. Furcolo, and V. Iacobellis (2000). Stochastic model of the width function. *Water Resources Research* 36(4), 1143–1157.
- Verhegghen, A., P. Mayaux, C. De Wasseige, and P. Defourny (2012). Mapping Congo basin vegetation types from 300 m and 1 km multi-sensor time series for carbon stocks and forest areas estimation. *Biogeosciences* 9(12), 5061.
- Verhoeven, J. T., B. Arheimer, C. Yin, and M. M. Hefting (2006). Regional and global concerns over wetlands and water quality. *Trends in ecology & evolution* 21(2), 96–103.
- Veyssieres, M. and R. E. Plant (1998). Identification of vegetation state and transition domains in California’s hardwood rangelands. *University of California* 101.
- Vivoni, E. R., F. Di Benedetto, S. Grimaldi, and E. A. Eltahir (2008). Hypsometric control on surface and subsurface runoff. *Water Resources Research* 44(12).
- Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007). Catchment classification and hydrologic similarity. *Geography Compass* 1(4), 901–931.
- Wagener, T. and H. S. Wheater (2006). Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. *Journal of hydrology* 320(1), 132–154.
- Walcott, R. C. and M. Summerfield (2008). Scale dependence of hypsometric integrals: An analysis of southeast African basins. *Geomorphology* 96(1-2), 174–186.
- Walvoord, M. A. and R. G. Striegl (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(12).
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(301), 236–244.
- Watson, K. B., T. Ricketts, G. Galford, S. Polasky, and J. O’Niel-Dunne (2016). Quantifying flood mitigation services: The economic value of otter creek wetlands and floodplains to middlebury, vt. *Ecological Economics* 130, 16–24.
- Wayand, N. E., J. D. Lundquist, and M. P. Clark (2015). Modeling the influence of hypsometry, vegetation, and storm energy on snowmelt contributions to basins during rain-on-snow floods. *Water Resources Research* 51(10), 8551–8569.
- Whitehurst, D., B. Friedman, K. Kochersberger, V. Sridhar, and J. Weeks (2021). Drone-based community assessment, planning, and disaster risk management for sustainable development. *Remote Sensing* 13(9), 1739.

- Wilcock, P., J. Pitlick, and Y. Cui (2009). Sediment transport primer: estimating bed-material transport in gravel-bed rivers. *Gen. Tech. Rep. RMRS-GTR-226*. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 78 p. 226.
- Willgoose, G. (2018). *Principles of soilscape and landscape evolution*. Cambridge University Press.
- Willgoose, G. and G. Hancock (1998). Revisiting the hypsometric curve as an indicator of form and process in transport-limited catchment. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Group* 23(7), 611–623.
- Wood, W. F. and J. B. Snell (1960). A quantitative system for classifying landforms. Technical report, Quartermaster Research and Engineering Command, US Army Command, Natick MA.
- Woods, R. (2003). The relative roles of climate, soil, vegetation and topography in determining seasonal and long-term catchment dynamics. *Advances in Water Resources* 26(3), 295–309.
- Woods, R. (2006). Global similarity indices for mean and seasonal hydrology of ungauged basins. In *Presentation at USA PUB Workshop*, Volume 16.
- Woodward, R. T. and Y.-S. Wui (2001). The economic value of wetland services: a meta-analysis. *Ecological economics* 37(2), 257–270.
- Wooldridge, S. A. and J. D. Kalma (2001). Regional-scale hydrological modelling using multiple-parameter landscape zones and a quasi-distributed water balance model. *Hydrology and Earth System Sciences* 5(1).
- Wu, Q. and C. R. Lane (2016). Delineation and quantification of wetland depressions in the prairie pothole region of north dakota. *Wetlands* 36(2), 215–227.
- Wu, Q. and C. R. Lane (2017). Delineating wetland catchments and modeling hydrologic connectivity using LiDAR data and aerial imagery. *Hydrology and Earth System Sciences* 21(7), 3579–3595.
- Xhardé, R., B. Long, and D. Forbes (2006). Accuracy and limitations of airborne LiDAR surveys in coastal environments. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pp. 2412–2415. IEEE.
- Xiong, S., M. E. Johansson, F. M. Hughes, A. Hayes, K. S. Richards, and C. Nilsson (2003). Interactive effects of soil moisture, vegetation canopy, plant litter and seed addition on plant diversity in a wetland community. *Journal of Ecology* 91(6), 976–986.

- Yadav, M., T. Wagener, and H. Gupta (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in water resources* 30(8), 1756–1774.
- Yang, P., D. P. Ames, A. Fonseca, D. Anderson, R. Shrestha, N. F. Glenn, and Y. Cao (2014). What is the effect of LiDAR-derived DEM resolution on large-scale watershed model results? *Environmental Modelling & Software* 58, 48–57.
- Yang, Z. and D. Han (2006). Derivation of unit hydrograph using a transfer function approach. *Water Resources Research* 42(1).
- Young, G., J. Krolak, and J. Phillippe (1986). Evaluation of alternative hydrograph methods for hydraulic design. *Hydraulics and Hydrology*, 28.
- Zhou, S., Z. Xu, and F. Liu (2016). Method for determining the optimal number of clusters based on agglomerative hierarchical clustering. *IEEE transactions on neural networks and learning systems* 28(12), 3007–3017.
- Zhu, X. and D. Guo (2014). Mapping large spatial flow data with hierarchical clustering. *Transactions in GIS* 18(3), 421–435.

APPENDIX A
SUPPLEMENTAL MATERIALS FOR CHAPTER 3

A.1 Standard Normal Distribution

Probability Density Function (*pdf*),

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in (-\infty, \infty) \quad (\text{A.1})$$

Cumulative Density Function (*cdf*),

$$\Phi(x^*) = \int_{-\infty}^{x^*} \phi(x) dx = \int_{-\infty}^{x^*} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad (\text{A.2})$$

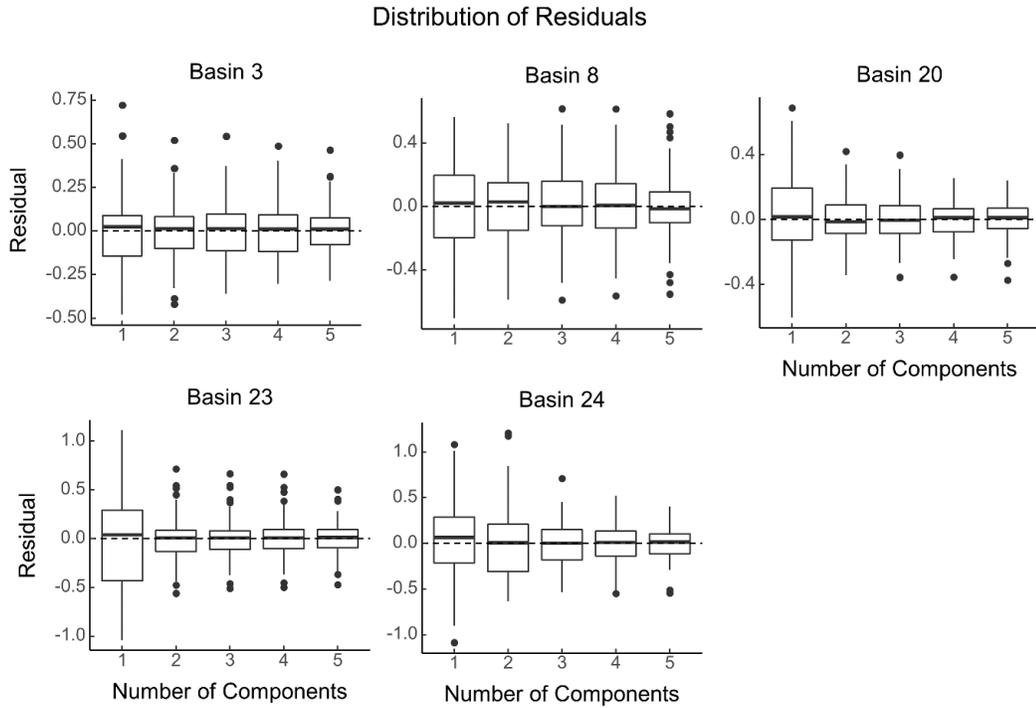


Figure A.1: Distribution of residuals for selected width functions. Here, residual has been defined as the difference between the actual width function and the modeled width function. A positive value indicates that the model value underestimates the actual value. The dots represent the outliers whose residual values are beyond 1.5 times the interquartile range.

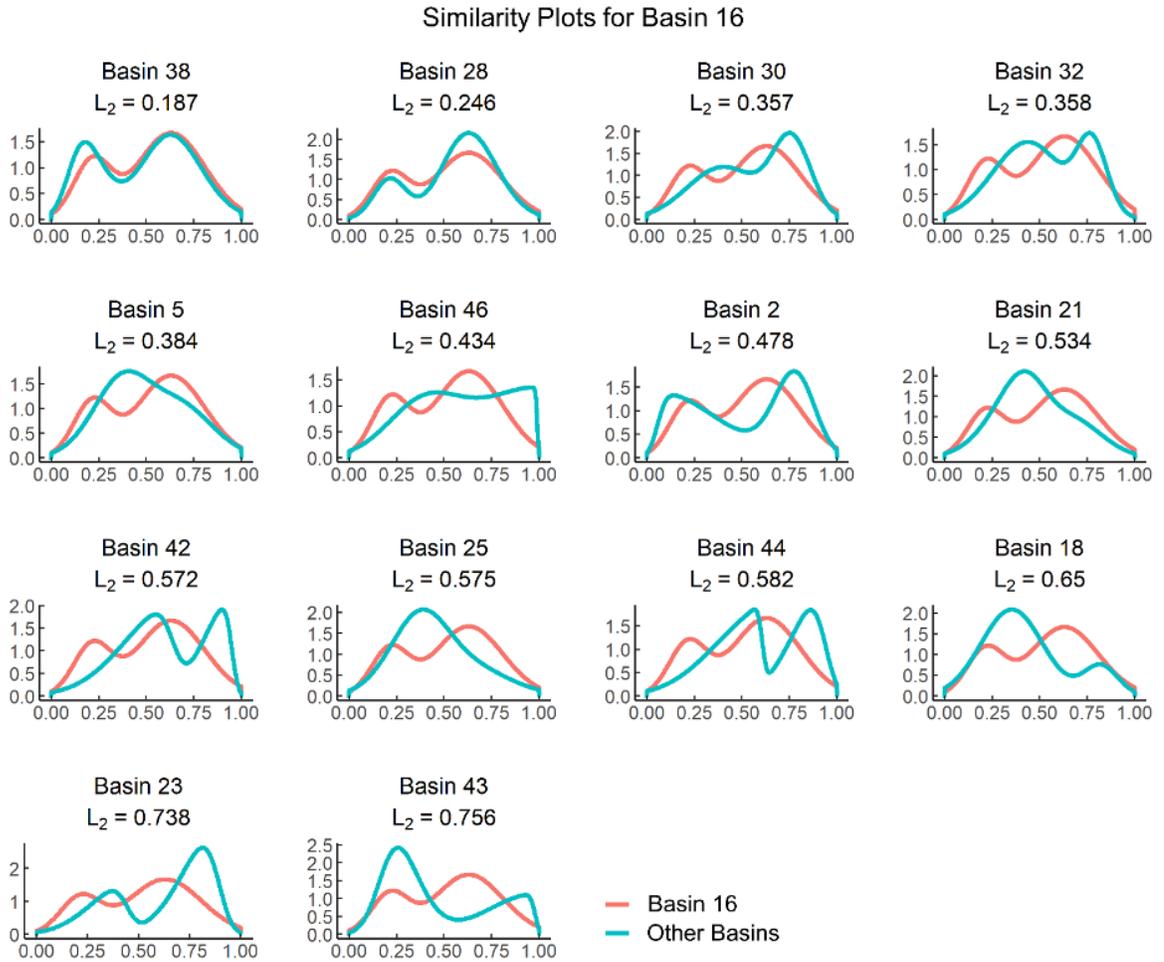


Figure A.2: Comparison of Basin 16 with all other basins with between 20% maximum hydrological distance and less than 40% overlap. The figures are ordered from the highest similarity to the lowest based on the L_2 divergence measure.

APPENDIX B

SUPPLEMENTAL MATERIALS FOR CHAPTER 4

B.1 Parameter Estimation

Parameter estimation for all functions, including the Generalized Hypsometric function, was done by non-linear least-squares fitting (NLSF) in R programming language using the *nlsLM* function from the "minpack.lm" package (Elzhov et al., 2016). NLSF is a form of regression where the fitting equation is non-linear in its parameters. Unlike linear regression, NLSF solutions rely on iterative procedures to find the best fit. The *nlsLM* function uses the Levenberg-Marquardt algorithm (LMA) for finding a local minimizer for NLSF problems (Moré, 1978).

For a set of n empirical area-elevation data points, (x_i, y_i) , where y_i represents a set of relative elevations and x_i represents the set of corresponding relative areas above the elevations, and a hypsometric function $y = f(x, \beta)$, with parameters, β , the sum of square deviations to be minimized is defined by:

$$\phi(\beta) = \sum_{i=1}^n [y_i - f(x_i, \beta)] \quad (\text{B.1})$$

The LMA algorithm adaptively varies parameter updates using a damping parameter to solve Eq. (B.1),

$$(J^T J + \lambda I)\delta = J^T (y - \hat{y}) \quad (\text{B.2})$$

where J is the Jacobian matrix, λ is the damping parameter, I is the identity matrix, δ is the parameter update matrix for each iteration, and $\hat{y} = f(x, \beta)$. Small values of λ result in a Gauss-Newton update, whereas large values result in a gradient descent update.

A sample code for parameter estimation of the Generalized Hypsometric function (Eq. (4.8) in Table 4.1) in R is provided next. The elevation data can be extracted from the DEM, with each value on the gridded DEM representing a single elevation point. The data can then be normalized using the total elevation range of the basin to obtain the relative elevation, y . The relative area above any elevation can be obtained by computing the proportion of data points larger than said elevation.

```
> # Given, x = relative area above elevation, y = relative elevation
> library(minpack.lm)
> nls_fit <- nlsLM(formula = y ~ ((r*(1-x^m)/((1-r)*x^(m+r)))^z),
                   data = data.frame(x, y),
                   start = list(r = 0.5, z = 0.5, m = 0.5),
                   lower = c(r = 0, z = 0, m = 0),
                   upper = c(r = 1, z = Inf, m = Inf),
                   control = nls.lm.control(maxiter = 1000))
```

B.2 Discordance Index

The discordance index (DI) was defined as the absolute area between two hypsometric curves to facilitate the comparison of different hypsometric functions. The DI ranges between zero and one. A value of zero indicates identical hypsometric functions whereas larger values indicate deviations. The DI between two hypsometric functions, $f_1(x)$ and $f_2(x)$ is defined by:

$$DI = \int_0^1 |f_1(x) - f_2(x)| dx \quad (\text{B.3})$$

B.3 Constrained Formulations

Scaled hypsometric curves have two boundary conditions (BC): they must start in the upper left-hand corner ($x = 0, y = 1$), and end at the lower right-hand corner

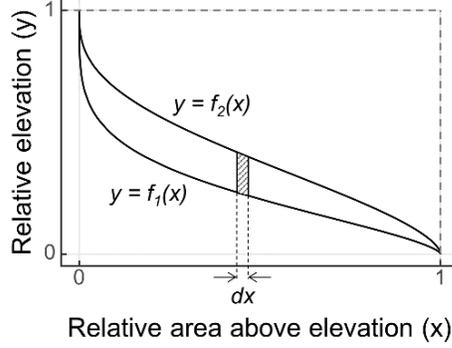


Figure B.1: Discordance index, defined as an integral of the absolute value of the difference between two hypsometric functions.

($x = 1, y = 0$). The Polynomial (Eq. (4.4)), Sigmoidal (Eq. (4.5)), and Double Exponential (Eq. (4.7)) functions do not satisfy these boundary conditions.

However, these functions can be modified to satisfy them by applying constraints to their parameters (Table 4.1). The modified Polynomial, Sigmoidal, and Double Exponential functions are shown in Eqs. (B.4), (B.5), and (B.6), respectively:

$$y = 1 + a_1x + a_2x^2 - (1 + a_1 + a_2)x^3 \quad (\text{B.4})$$

$$y = \frac{b(1 - x^d)}{b + x^d} \quad (\text{B.5})$$

$$y = (1 - a_3)e^{a_2x} + a_3e^{\left[\ln\left(-\frac{(1-a_3)}{b}e^{a_2}\right)\right]x} \quad (\text{B.6})$$

B.4 Parameter Interpretation Based on Hypsometric Integral

Here we check the relationship between the parameter values of the Generalized Hypsometric fits and HI values for the study regions. We found no clear and significant correlation between the parameter values and HI (Fig. B.2 and Table B.1). r values had a positive correlation with HI, whereas z values had a negative correlation. The strongest correlation was for z values in the YRB.

Table B.1: Pearson correlation coefficients between the parameter values of Generalized Hypsometric fits and the hypsometric integral (HI) values.

Watershed	r	z	m
Congo	0.409**	-0.021	0.211**
Narmada	0.446**	-0.286*	-0.097
Yukon	0.309**	-0.498**	-0.263**

* $p < 0.05$, ** $p < 0.01$

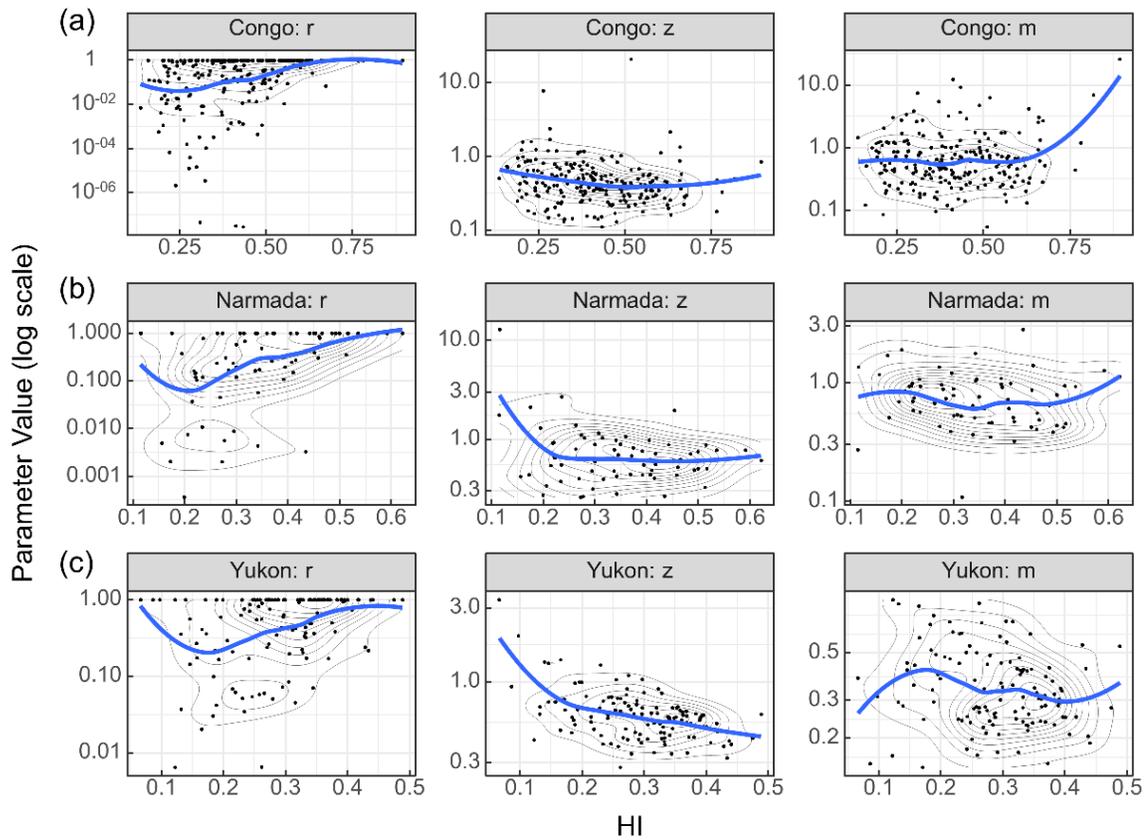


Figure B.2: Relationship between the three parameters of the Generalized Hypsometric functional fits and hypsometric integral (HI) values for each sub-basin. Along with the point plots, contours have been added to show the point density, and a Locally Weighted Scatterplot Smoothing (LOWESS) line has been shown to indicate any relationship trends.

B.5 Comparison of Strahler and Generalized Hypsometric Fits

The RMSE values are lower for Generalized Hypsometric fits compared to Strahler fits for every sub-basin as expected. Fig. B.3 shows a comparison of fits for a variety of shapes and also highlights some of the deficiencies of these formulations. The Generalized Hypsometric function provides an improved fit for CRB sub-basin 100 and 106, NRB 8 and 60, and YRB 1, 16, and 81. However, in cases where the hypsometric curve exhibits more non-traditional shapes such as a significant flat region as in CRB sub-basin 101, and multimodality in distribution as in CRB 106, and NRB 8 and 17, both these fits fail to capture the entirety of the shapes.

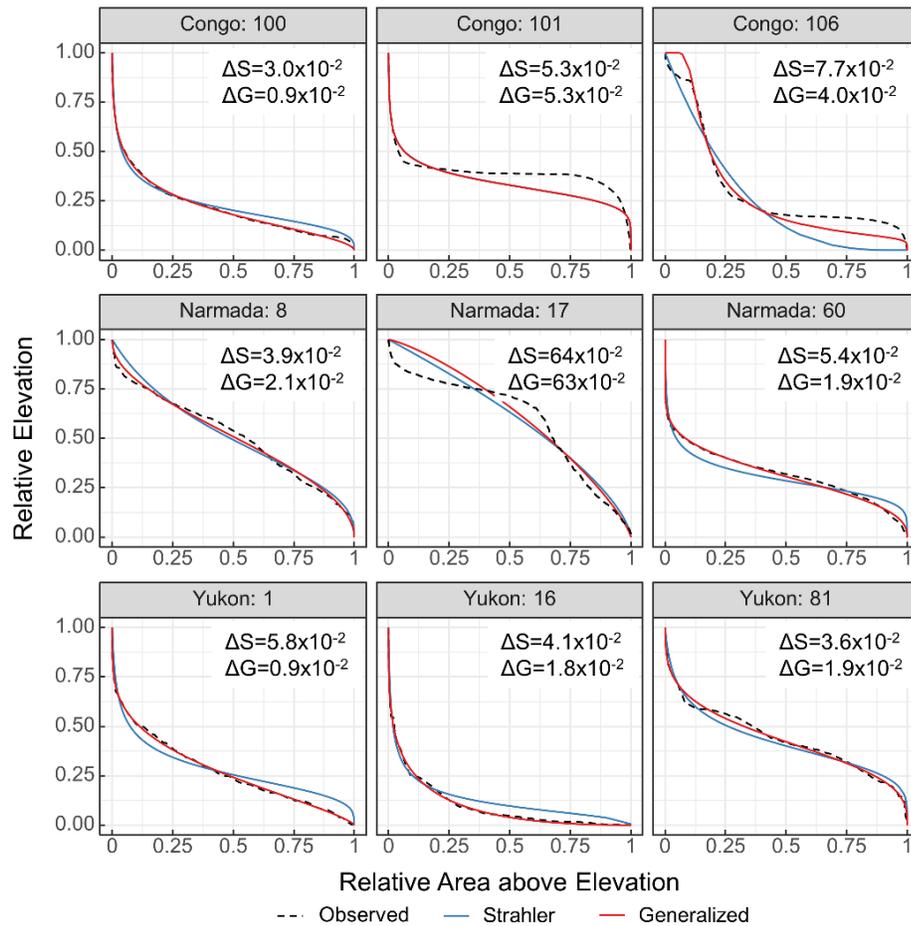


Figure B.3: Comparison of Strahler and Generalized Hypsometric function fits for a sample of nine sub-basins (three from each study region). ΔS and ΔG indicate the root mean square errors (RMSE) for the Strahler and Generalized Hypsometric function fits respectively.

B.6 RMSE

RMSE represents the standard deviation of the residuals of a model (Eq. (B.7)). It is a measure of differences between the observations and the model predicted values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}} \quad (\text{B.7})$$

where x_i represents the observations, \hat{x}_i represents the predicted values, and N represents the number of observations.

APPENDIX C
SUPPLEMENTAL MATERIALS FOR CHAPTER 5

C.1 Original Width Function Clusters

Figure C.1 shows the width function clusters before the removal of outliers. Clusters 1, 5, and 6 have higher peaks in the right SN component while cluster 3 has a higher peak in the left SN component, potentially indicative of different location of peak flows in hydrographs. Furthermore, the high slopes on right sides of the curves for clusters 2 and 6 could be indicative of more rapidly falling recession limbs of hydrographs.

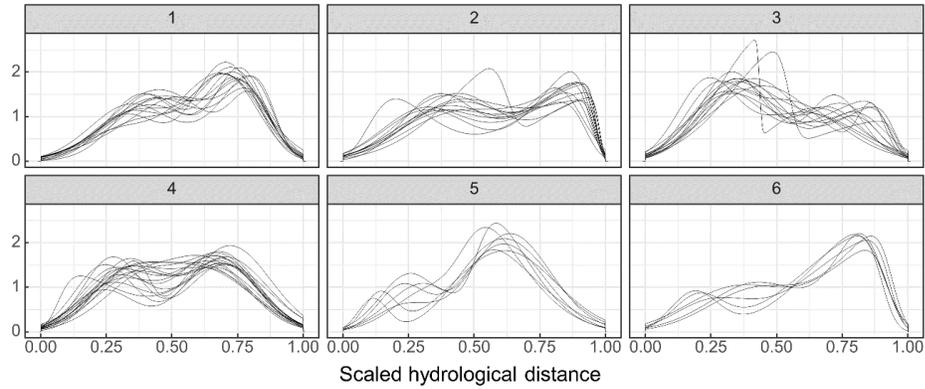


Figure C.1: Width functions in each cluster.

APPENDIX D

SUPPLEMENTAL MATERIALS FOR CHAPTER 6

D.1 Including Overlapping Sub-basin Pairs

We repeated the correlation analysis by including the nested sub-basins to see the consistency in the results. We still believe that the results for non-overlapping sub-basins provide a more accurate reflection of the relationship between the catchment metrics and the hydrograph properties due to the shared information might bias (Gottschalk et al., 2006; Skøien et al., 2006; Skøien and Blöschl, 2007; Gottschalk et al., 2011; Müller and Thompson, 2015). However, Betterle and Botter (2021) suggests that nested catchments show a pronounced decrease in streamflow correlation compared to non-nested catchments when inter-catchment distances increases.

Figure D.1 shows the relation between similarity based on width function and similarity based on the hydrograph properties. For four of the properties, the relationship still holds true for the nested sub-basin pairs. The linear regression lines for overlapping and non-overlapping sub-basins are parallel in two of these cases—for skewness and kurtosis, while they are aligned in the same direction in the cases of time to peak and limb slope. However, the relationship fails to hold in the case of maximum discharge. The relative sizes of the sub-basin pair seem to have no influence over the correlation. The results for all metrics are summarized in Table D.1. The width function is no longer statistically correlated with maximum discharge when overlapping sub-basins are included in the analysis. However, the width function and hypsometric are still the two most strongly correlated metrics with the hydrograph properties. The results for the long profile remain unchanged. The CTI distribution is no longer statistically correlated with the maximum

discharge. Instead, the average CTI along the flow path is now correlated with limb slope and kurtosis. While the results are slightly different here, the overall importance of width function, hypsometric curve, long profile, and rDUNE are still evident.

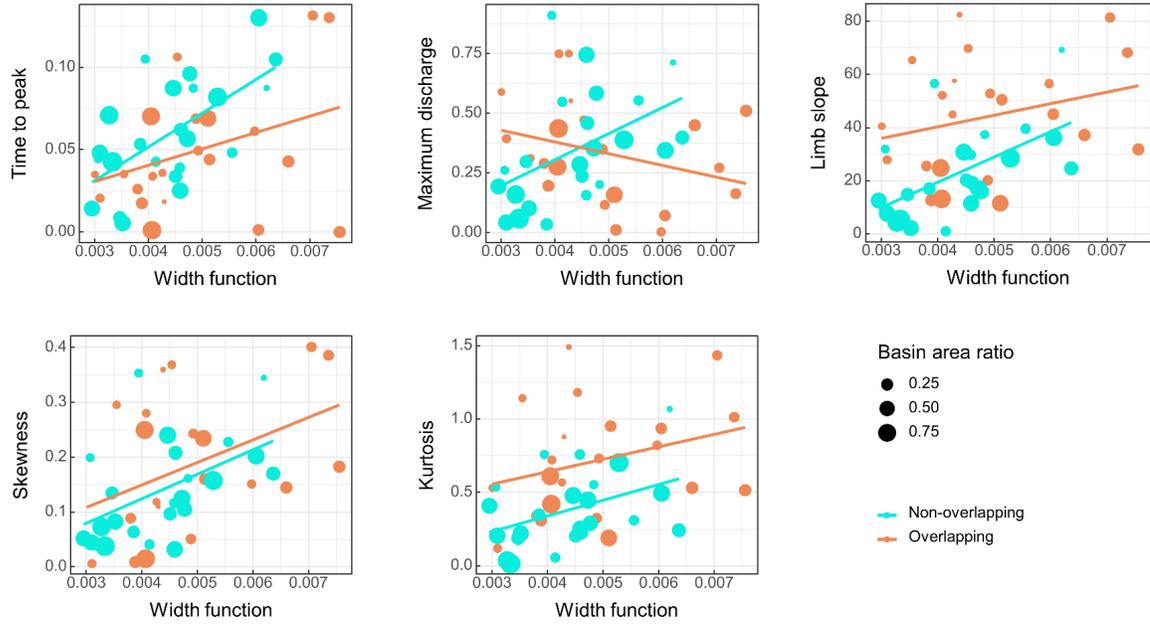


Figure D.1: Comparison of similarity based on width function and similarity based on hydrograph properties. Pairwise DIs for width function are plotted on the x-axes and the arithmetic difference between hydrograph properties are plotted on the y-axes. Overlapping and non-overlapping sub-basin pairs are distinguished by color and linear regression lines are shown separately for them. The size of the point indicates the ratio of basin areas.

Table D.1: Statistical significance of Pearson correlation between similarity based on the catchment metrics and based on all hydrograph properties in this study. The correlation significance based on p-values is denoted using the significance codes. Darker shades of blue with a higher number of ‘*’ symbols indicate a higher significance. Overlapping sub-basin pairs were included in the analysis.

Metric	Tmax	Qmax	Limb Slope	Skewness	Kurtosis
Slope distribution					
Hypsometric curve		***	***	**	**
CTI distribution					
HAND distribution					
rDUNE distribution					
Long profile		***			
Average slope along flow					
Average CTI along flow			*		*
Average HAND along flow					
Average rDUNE along flow					
Width function	**		**	***	**



BIOGRAPHY OF THE AUTHOR

Prashanta Bajracharya got his Bachelor's degree in civil engineering from the Tribhuvan University, Nepal in 2014. He came to Maine in 2017 to pursue his master's in civil engineering at the University of Maine. He is interested in data analysis to plan for a sustainable future. In his free time, Prashanta enjoys playing tennis, hiking, skiing, reading, and playing board games.

Prashanta Bajracharya is a candidate for the Doctor of Philosophy degree in Civil Engineering from The University of Maine in August 2022.