Assessing Natural River Fragmentation Through Waterfalls at a Global Scale

By

Ziqian Han

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of B. Sc. in Geography

Department of Geography
McGill University
Montréal (Québec) Canada

April 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Waterfalls are normally vertically steep falls sitting on a river. Since they create barriers that cut the river network, they cause river and habitat fragmentation which has a series of environmental significance such as promoting the isolation of aquatic populations and facilitating genetic diversification. Waterfalls can also alter the global carbon cycle, benefit fisheries, provide recreation sites, support hydropower, and even modify air characteristics. However, large-scale studies on river fragmentation by waterfalls are constrained by the lack of high-quality dataset with sufficient spatial coverage and completeness. This thesis improves the current version of the global waterfall database HydroFALLS v1a by drawing new waterfall points from various datasets. After data preparation, consolidation, and validation, a national dataset of China contributes 74 new waterfall points that are later co-registered with digital river network of HydroSHEDS and subsequently merged into HydroFALLS v1a to produce the updated version HydroFALLS v1b. Another 53,205 points are collected from two supplemental global dataset, which are ranked to 25 different priorities of workflow based on their proximities to rivers of different sizes. Upon further examination into each rank, the 53,205 points serve as a data pool for the development of future database version HydroFALLS v1.0. The global-scale usage and applicability of HydroFALLS v1b is tested with the relationship between freshwater fish diversity distribution and the presence of waterfalls as a natural habitat fragmenting factor. A variety of levels of correlation in different ecozones reveal large-scale impacts of natural river fragmentation and the feasibility of related studies, which highlights the necessity of building a high-quality global waterfall database.

CHAPTER 1: INTRODUCTION

1.1. Overview and Background

A waterfall normally refers to a steep, rapid or vertical fall on a stream. These abrupt changes in riverbed topography and river networks can lead to natural river fragmentation in multiple aspects, including the longitudinal flow connectivity, aquatic species migration, flow regime, hydrodynamics, sediment transport, and fluvial geomorphology, thus having a variety of ecological and environmental impacts (Hudson, 2013). Numerous local studies have shown that waterfalls have the potential to modify the spatial pattern of gene flow and evolutionary processes, influence the physicochemical properties of riverine water, play a part in quantities and timing in the global carbon cycle, bring benefits or changes to fisheries, provide natural sites for hydroelectricity, serve as scenic attractions or cultural heritages for local people, and even alter air electricity and composition (Rahel, 2007; Natchimuthu et al., 2016; McLachlan et al., 1990; Zhang et al., 2000; Luts et al., 2009; Dibiase et al., 2014; Thé and Nordi, 2006; Uhunmwangho and Okedu, 2009).

However, large-scale waterfall studies are constrained by the lack of high-quality global waterfall datasets. Due to the absence of clear and systematic definitions, classifications, data development, and interdisciplinary studies (Hudson, 2013), current waterfall databases are neither complete and accurate in spatial coverage, locations, and related properties, nor being georeferenced with river networks. Many definitions, criteria, classification systems, and rating scales proposed are inconsistent. Ford (1986) specified a waterfall as a vertically steep fall on a river but did not address certain thresholds for the degree of steepness and flow magnitudes. Mabin (2000) and Dias et al. (2013) proposed the slope thresholds to be 25% and 30% respectively. At the same time, some classifications proposed by scientists and magazines such as National Geographic (2012) based on different criteria including the height, discharge, width, water speed, shape, slope, and possession of plunge pools are mixed in one single system without being comprehensively consolidated for different purposes, making these properties unorganized and unclear for usage by other disciplines. The lack of georeferencing and co-registration between locations of waterfalls and river networks makes it unknown whether the waterfalls are sitting on the river that causes fragmentation or appear at another location nearby.

The current version of a global waterfall database, HydroFALLS v1a (Lehner, 2013), contains some major spatial gaps, particularly in Asia and some other countries that lack national datasets. Meanwhile, the attribute information of a portion of waterfall features recorded is not accurate and complete. Therefore, it is of great interest to develop a unified global waterfall database by building upon HydroFALLS v1a to add more clear, systematic, accurate, and complete waterfall points worldwide.

The digital river network that is part of global hydrographic data framework termed HydroSHEDS (Lehner et al., 2008) was derived from a SRTM Digital Elevation Model (DEM) at a 15 arc-second resolution, which provides high-quality mapping of rivers consistently at a global scale. However, it does not always reflect the exact location of the true river network. Therefore, it is also necessary to co-register the waterfall points with the digital river network to identify the relative location of the waterfall and the river.

1.2. Research Objective and Questions

This thesis aims to develop a systematic, complete, and unified global waterfall database, HydroFALLS, that contains related waterfall parameters, as well as to co-registers the database with the existing digital river network that is part of global hydrographic data framework termed HydroSHEDS (Lehner et al., 2008). To achieve this, a variety of data sources are explored to fill major spatial gaps in HydroFALLS v1a. Points collected from multiple global and regional datasets are further cleaned, merged, consolidated, validated, and co-registered with digital river networks. Depending on properties such as cleanness, completeness, reliability, and level of details of the data sources, different steps are tried and tested to develop a standardized workflow accordingly that allows for the continued mapping of waterfall locations beyond this thesis. Metadata and related attributes of current data points and newly added points are updated with supplemental data sources. In the end, global maps and summary statistics are created regarding data quality and waterfall properties to provide an overview of notable waterfalls and the global pattern of natural river fragmentation, with particular insights into the spatial relationship between waterfalls and land cover types. To investigate the usage and applicability of the data product, as well as understand potential remaining challenges in global studies of natural river fragmentation, genetic diversity of freshwater fish populations separated by major waterfalls is examined as an ecological application of the database.

1.3. Thesis Format

Following this introductory chapter, chapter 2 provides a review of waterfall studies conducted in related fields. I will start by explaining how local studies have shown that waterfalls around the world are participating in a variety of environmental processes. I will introduce the development of waterfall studies and the complexities of defining and classifying waterfalls, as well as how the current system can be problematic. Meanwhile, data availability and the performance of current datasets are assessed, which show the necessity of this thesis. I will introduce the methodology adopted in chapter 3 and results of this project in chapter 4, followed by comprehensive evaluations and discussions about the results in chapter 5. Chapter 6 summarizes and extends the findings and potential implications of this thesis.

CHAPTER 2: RELATED LITERATURE

2.1. Waterfalls’ Roles in Earth Systems

2.1.1. Habitat Fragmentation and Ecological Regulations

Although waterfalls have been understudied systematically on a global scale, numerous local studies have shown that waterfalls are playing important roles in environmental systems. Similarly to dams, waterfalls as natural barriers that can cause habitat fragmentation which controls the evolution of fish species (Rahel, 2007; Cote et al., 2009; Northcote, 2010). In comparison with other types of habitats that allow random intra-range and extra-range dispersal and movement of organisms, riverine freshwater species are regulated by the directional water flow and the hydrographic pattern of river networks, which closely relates aquatic habitats with river connectivity (Dias et al., 2013). With waterfalls cutting the river, watersheds upstream would receive lower immigration of aquatic organisms, reducing the population size of species and making them more likely to become extinct. With low connectivity, different species and sub-populations have a lower chance to genetically interact in watersheds upstream, giving them a longer genetic distance and a smaller genetic variation. Therefore, upstream populations tend to experience genetic divergence (Losos and Parent, 2009), leading to a lower species richness as a result of the homogenized distribution along the genetic gradient (Oberdorff et al., 2011). Meanwhile, different populations of the same species separated by barriers can also go through parallel evolution towards different genetic responses.

These effects have been proven by numerous case studies around the world, especially in large rivers. The spawning migration of endangered species *Xyrauchen texanus* and *Ptychocheilus lucius* in the Colorado River Basin is blocked by a recently formed waterfall (Cathcart et al., 2018); the extinction and speciation rate of fish in the Orinoco River Basin was higher in watersheds isolated by waterfalls (Dias et al., 2013); the genetic distances of *Rhinogobius* sp. (goby) YB and BR in the Iriomote Island in Japan are highly correlated with the waterfall height (Figure 2.1.), suggesting that the waterfalls have been driving parallel evolution of goby at different isolated basins (Kano et al., 2012). These studies cover almost all the major regions in the world and involve a great variety of aquatic species that are important to the ecosystem.

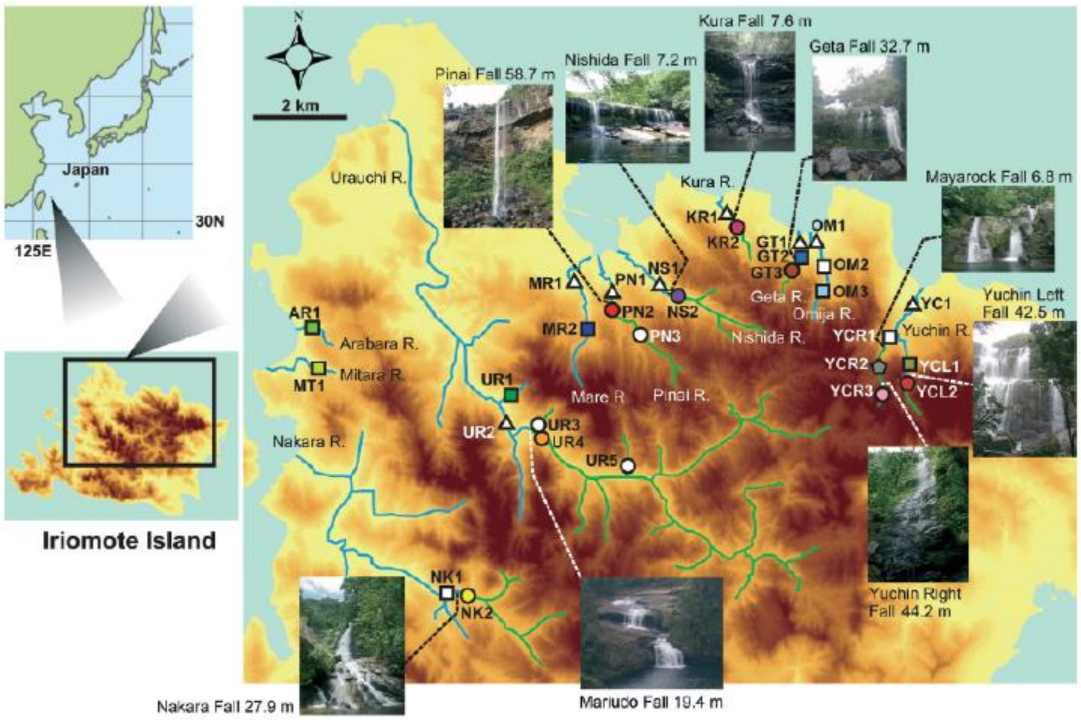

Figure 2.1. Waterfalls and river basins on Iriomote Island, Japan. Rivers are indicated with blue (below waterfall) and green (above waterfall) lines. Distribution of morphological samplings sites are indicated with circles (YB), squares (BR), triangles (neither) and pentagons (both) (Kano et al., 2012).

In addition to genetic interactions, waterfalls can also affect fish species distribution through selective and adaptive pressures on body morphologies and migratory abilities. For example, the genetic response of fish to movement in water current is found to be more positive in above-waterfall populations of *Oncorhynchus mykiss* in the Kokanee River in Canada (Northcote, 1981); the presence of Gastromyzon fish and the absence of Cyprinidae taxa above-waterfall in multiple streams in Brunei were related with their different climbing capabilities controlled by the possession of ventral suckers (Baker et al., 2017). Some waterfalls are accompanied by plunge pools downstream which act as isolated habitat patches with distinct water dynamics and physicochemical properties such as the sediment stock, water temperature, and pH, leading to the formation of special species assemblages or genetic responses such as adaptations to the high-energy environment (Ramsay, 2001).

The alteration of fish species composition interplays with other biotic and abiotic impacts such as the distribution of aquatic invertebrates, periphyton, and aquatic insects, as well as the variation in litter decomposition rate and water properties (Baker et al., 2017).

Since anthropogenic fragmenting factors such as dams also have similar ecological effects, it is important to differentiate waterfalls from dams so that natural factors can be ruled out from anthropogenic impacts. Notably, this has been applied by Grill et al. (2019)'s study which excludes dams below waterfalls from assessing the fragmentation level of upstream reaches, since the waterfalls have been acting as a disconnecting factor.

2.1.2. Physicochemical Alterations of Air and Fluvial Content

Other than habitat fragmentation, waterfalls might be contributing to a considerable amount of emission of Greenhouse Gases (GHG), toxic gases and other organic chemicals through locally generated jet aeration and low pressure (Natchimuthu et al., 2016; McLachlan et al., 1990; Leibowitz et al., 2017; Zhang et al., 2000), which alters the global carbon cycle. The spatiotemporal pattern of riverine GHG emission can also be regulated by waterfalls through variations in discharge, steepness, and height. In karst areas where water calcium ions are often supersaturated, degassing processes soften water by releasing carbon into the air and converting calcium ions to deposits on the riverbed, which lowers the water pH, cements coastal unconsolidated materials, and helps to form speleothems in the riverine system (Usdowski et al., 1991; Zhang et al., 2000; Chen et al., 2004). Along with the permanent modification of air ion spectra through autoionization, fluctuating electric charge rearrangement, collisions, surface protrusion, and Coulomb explosion (Luts et al., 2009), the release of chemicals also alters local air characteristics.

2.1.3. Control of Channel Morphology and Landscape Development

As so-called “knickpoints” in hydraulic geometry and fluvial dynamics, waterfalls may facilitate or inhibit riverbed retreat and landscape adjustment to external forcing depending on the rock strength, sediment supply, water discharge, and riverbed morphology (Haviv et al., 2010). For example, the retreat of the knickzone around Lower Big Tujunga Falls created strath terraces and an inner gorge with oversteepened hillslopes behind (Dibiase et al., 2014). Niagara Falls are also undergoing recession through the detachment of small particles from the waterfall face rock as a result of surface water erosion and frost weathering (Hayakawa and Matsukura, 2010). This would require attention for the relocation of surrounding settlements and urban infrastructures.

Along with bedrock erosion, waterfalls also regulate sediment connectivity in rivers through sediment reservation in plunge pools. Scheingross and Lamb (2016) found that the depth controls jet hydrodynamics and sediment export from the pool as a natural reservoir, while upstream sediment supply conversely regulates the equilibrium of sediment stock in the plunge pool. Adding to the sediment size, waterfall height, and discharge, the waterfall zone alters fluvial sediment transport similarly to dams, although normally at a much smaller magnitude.

2.1.4. Direct Human Uses

Further understanding of fragmentation helps to develop fisheries by optimizing selections of species, sites, timing, and techniques. At the waterfall of Buritizeiro on the Sao Francisco River in Brazil, fishing spots built in the rapids that are regarded as “Channel’s Head” by local fishermen are considered to be a part of Common Property Systems (Thé and Nordi, 2006). The sites are selected to be “locations where fish pass in the waterfall”, fishes such as *curimbata* and *piau* can be caught when they jump or get stranded on the rock during low discharge. In Southern Laos, local people take the advantage of the Khone Falls on the Mekong River where seasonal fish spawning migration occurs (Roberts and Baird, 1995).

At the same time, these abrupt changes in rivers make them potential spots for hydroelectricity due to hydraulic turbulence and elevation difference. Several sites such as Agbokim Waterfalls in Nigeria, Kihansi Waterfall in Tanzania, as well as Sapchari Waterfalls and Sahasradhara waterfall in Bangladesh are assessed to be feasible for hydropower generation, despite some discussions about potential impacts on local ecosystems, ecotourism, and human livelihoods (Uhunmwangho and Okedu, 2009; Zilihona and Nummelin, 2001; Jui et al., 2015; Wazed and Ahmed, 2009).

Other values such as aesthetics and tourism are widely shown by waterfall tourism sites around the world, such as the well-known Niagara Falls, Victoria Falls, and Yellowstone National Park (Hudson, 1998). In many countries, waterfalls are important scenery attractions that have been bringing enormous recreation value and economic income. In some places, waterfalls compose an important part of indigenous cultures, natural heritages, and spiritual symbols. A variety of geographical and cultural contexts such as in China, Jamaica, Europe, North America, Europe, Australia, and New Zealand historically regard waterfalls as “beautiful, sometimes sublime, sometimes picturesque” sites that widely appear in poetry, paintings, and myths (Hudson, 2001, p. 9). However, many of these waterfalls, e.g., Dunn’s River Falls in Jamaica, are undergoing unsustainable exploitation for tourism which has caused issues such as declined flow (Hudson, 1999). This highlights the demand for a balance between tourism development and environmental protection.

2.2. Waterfall Studies, Definitions, and Classifications

For a long period, waterfalls were absent in scientific research (Hudson, 2013). They were firstly studied in various fields including geology, geomorphology, tourism and freshwater ecology, as well as mentioned in tourism guides. Smethurst (2020) raised the need to develop mountain geography, which implicates the potential field of waterfall geography. Recently, waterfalls are receiving an increasing level of scientific attention, but are still lacking systematic and precise investigations on many aspects such as the definition, classification, and interdisciplinary significance (Hudson, 2013).

Ford (1968) claimed a waterfall as a synonym of cataract to be a vertically steep fall sitting on a river, while a cascade means one or a series of falls with a lower magnitude; rapids are less steep but the flow is still distinctively fast for water to appear to be white. Later in 2000, Mabin added that waterfalls are marked by an abrupt steepening in the stream. He also defined the threshold of the horizontal gradient across the lip and the plunge pool to be less than 25% of the height. Different definitions may exist for different purposes. For example, Dias et al. (2013) proposed a slope threshold to be 30% between two neighboring pixels of 30 arc-second (approximately 1 km) resolution to efficiently limit fish migrations.

Since waterfalls have multiple parameters affecting their visual magnitudes, it is also a problem to evaluate the relative size of waterfalls. Beisel (2006) developed a waterfall rating scale that defines a waterfall's size by the base e logarithm of the average water volume present on the fall. This system considers the average discharge, width, and height because the average water volume on the entire fall would depend on both the volume of water flowing through a unit area within a unit of time and the time it takes for water to travel from the top to the bottom of the fall. It also includes the slope and the complexity of the shape by specifying that with the height and average discharge being equal, the more complicated or the less steep ones receive a higher rating as they geometrically have a larger area of the waterfall surface. It then ranks waterfalls from class one to ten with the smallest falls such as Cucumber Falls in the USA in class one and the largest falls such as Niagara Falls in Canada in class ten. This system provides a way to unify different variables measuring the waterfall size to a single all-inclusive standard.

In addition to the size, there are multiple ways to classify waterfalls depending on the purpose. National Geographic (2018) proposed a system that puts waterfalls into block waterfalls, cascades, cataracts, chutes, fan waterfalls, horsetail waterfalls, multistep waterfalls, plunge waterfalls, punchbowl waterfalls and segmented waterfalls based on the morphology and flow shape. A genetic method proposed by Lobeck (1939) classifies waterfalls according to their origins, such as hanging tributary, landslide, and fault.

Some waterfalls also show seasonality in discharge and water state. For example, Yosemite Falls have a large discharge during spring snowmelts but tend to be dry at the end of summer (Plumb, 1993). Some high-latitude waterfalls such as Niagara Falls are frozen in the winter (Hayakawa and Matsukura, 2010), contributing to an even higher scenery value but also a greater complexity of environmental impacts and classification methods.

2.3. Current Database and Remaining Problems

The current version of the waterfall database HydroFALLS v1a (Lehner, 2013) was developed in 2013 as part of a larger research effort to study global river fragmentation and resulted in a first-ever, unpublished draft of a global waterfall database. It contains 4054 waterfall points (Figure 2.2.) mainly collected from global datasets including NGA Geonames, Digital Chart of the World (DMA, 1992), World Waterfall Database (World Waterfall Database, 2018), and Encyclopedia of Hydrology and Lakes (Herschy, 1998). Nine regional datasets drawn from governmental, academic, and private databases covering Norway, Iceland, Mexico, Canada, USA, Australia, Tasmania, Brazil, and New Zealand were taken as supplemental sources for further investigation and correction of individual points provided by global datasets. The prioritization of global datasets when discrepancies existed aimed to avoid biases towards regions or spatial inconsistency on the data standard. Data quality ratings on a scale of 1-5 from “Confident” to “Uncertain” were assigned as a measure of the confidence level of the waterfall’s existence and location, depending on whether it was consistently verified by multiple supplemental data sources and the visual clarity on satellite images (Figure 2.3.). With standardized attributes, the name, country, height, catchment area, discharge, quality rating, description, and coordinates were developed for each waterfall point (Table 2.1.), and co-registration with the HydroSHEDS digital river network (Lehner et al., 2008) was applied.

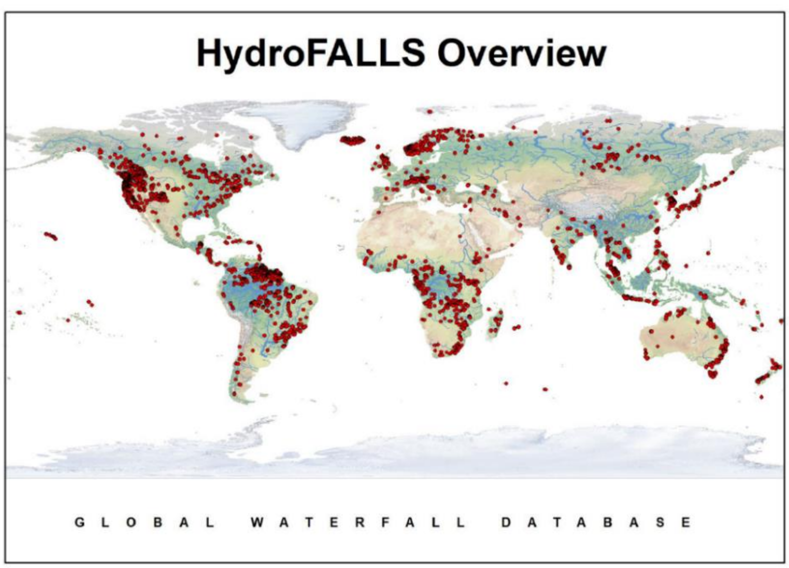

Figure 2.2. Overview map of HydroFALLS v1a (Lehner, 2013). The red dots mark the waterfall locations on the global river network.

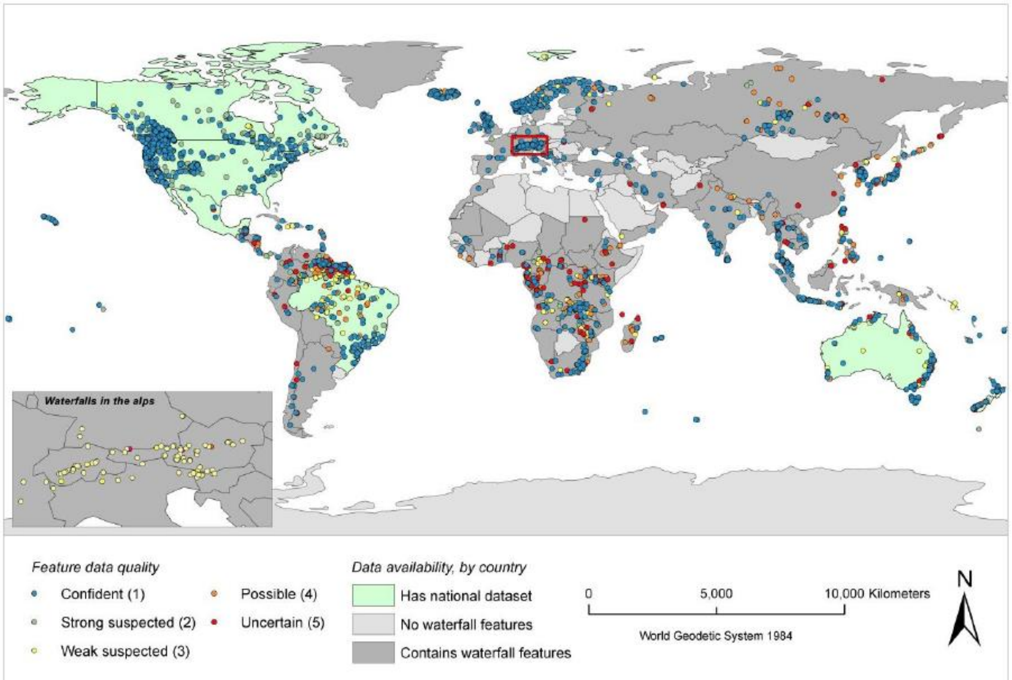

Figure 2.3. Confidence level of waterfall points in HydroFALLS Version v1a (Lehner, 2013)

Table 2.1. Attributes of waterfall points in HydroFALLS v1a (Lehner, 2013).

Attribute	Description
Falls_ID	Unique point identification number (1-4311)
Name	Most common name of the waterfall feature
Alt_Name	The second most common or suspected likely waterfall name, in the presence of a naming conflict. When there is no naming conflict the field is left NULL.
Country	Country which waterfall feature is located
Height_m	Waterfall feature height as specified in source datasets. If not specified it is left NULL.
Catch_KMS	Area of the catchment which waterfall feature is located as specified in source datasets. If not specified it is left NULL
Flow_CMS	Waterfall feature discharge as specified in source datasets. If not specified it is left NULL
Quality	Indicates the user's confidence in the accuracy of the point: 1: (Highly Confident) Waterfall feature is verified by multiple reliable data and supplementary sources and there is no evidence to the contrary 2: (Confident) Waterfall feature is sourced by multiple data sources and uniquely identified. 3: (Suspected) Waterfall feature is sourced by one reliable data source or multiple unreliable data sources. 4: (Possible) Waterfall feature is not uniquely identifiable but there suspicion that waterfall feature exists at location. 5: (Uncertain) Waterfall feature is not expected to exist at location but there is no contradictory evidence.
Comment	Textual descriptions of conflicts, overflow of attribute conflicts in the event that there are more than two suspected names, or more than one height or discharge value.

While the broad variety of data source types used raises the completeness and reliability of the database, it introduces strong variation in data accuracy, spatial coverage, and attribute completeness. National and governmental datasets usually have relatively high reliability, but they only provide data points for a very limited region. Countries without national data are left as gaps that are underrepresented at a global scale. Many national datasets are hard to obtain due to language, regional access authentication, and variation in data type. Existing private global datasets are strongly affected by personal interests, which omit regions that are less accessible and popular. Some waterfalls such as those with large heights but small discharges might have high touristic and aesthetic values due to their spectacular falling distance from high mountains, but they are less important to river fragmentation since they do not affect any large streams or freshwater habitat. These global datasets are also less accurate without verification of information uploaded by the public. Many data sources failed to provide official data or to be updated with new waterfalls being discovered.

These disadvantages of source data result in poor applicability of HydroFALLS v1a for global or regional studies that consider waterfall properties or require georeferenced locations of waterfalls as barriers sitting on particular streams. Some attributes are not available for most of the points. For instance, 209 data points do not have a name registered in the database and a large portion of waterfall points are lacking variables such as height. Even for those with height noted, a large part of them was drawn from unofficial sources. Duplication of data points with different labels from multiple datasets also added difficulties to the differentiation of these features.

In addition, there is still a considerable number of major waterfalls not covered in the database. For instance, China and Russia are left almost completely unaddressed. Meanwhile, many existing waterfall points were suspected to be dams, especially in Europe where numerous dams are present, making it difficult to isolate waterfall points based on online data sources.

CHAPTER 3: METHODOLOGY

To develop HydroFALLS v1b from HydroFALLS v1a, the same principles were adopted but with new and more consistent source data. As shown in the method overview (Figure 3.1.), regional datasets with a relatively manageable number of waterfall points extracted were cleaned up, examined, processed to the final product, and merged into HydroFALLS v1a to produce HydroFALLS v1b. Basic information and map statistics of this new version were studied to get a general idea of the spatial pattern of global waterfalls and data quality. The applicability of this new version was then tested with a global-scale analysis of fish diversity above and below major waterfalls to study the ecological fragmentation effect of waterfalls. To fill the inadequacy of regional datasets, two global datasets with large base numbers of new points were processed and tested for in-depth investigations into individual waterfall points.

All the processing steps were operated with ArcMap 10.7.1 and organized as point layers in geodatabases. The coordinate system of all layers was standardized to GCS_WGS_1984.

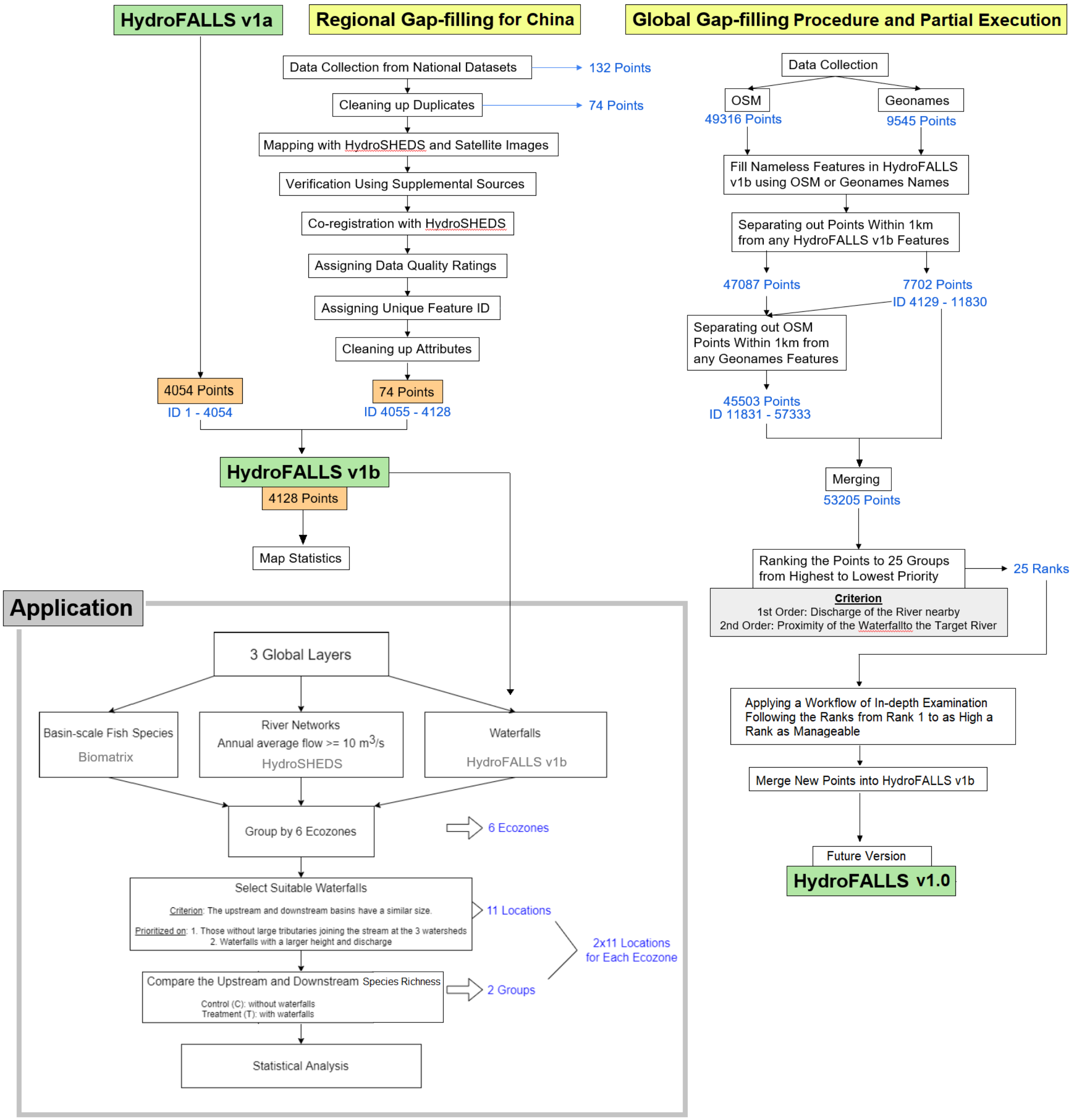

Figure 3.1. Methodology overview.

3.1. Regional Dataset of China

To fill the geographical gap of the current version of HydroFALLS, several regional datasets from governmental, academic, and private sources were explored and tested. However, most of these datasets were not accessible or failed to provide any waterfall points. Eventually, the National Basic Geographic Database of China (https://www.webmap.cn/main.do?method=index) successfully provided waterfall points which were accessed by querying water features with specified Chinese keywords. Extracted points from different provinces were merged.

Among the 132 newly added data points, 58 duplicates were discarded. The 74 points were examined with details: each point was mapped with rivers drawn from HydroSHEDS onto the satellite imagery that is available online as part of ArcMap. Any visible waterfall point within 5 km of the original point location was checked on the relative location with the river channel, the size and shape, and confidence of the feature being a waterfall. The existence and exact location of each point were validated with additional sources including Google Earth and various tourism media such as Tripadvisor (https://www.tripadvisor.ca). If the waterfall was confirmed to locate on the river channel, the original point was snapped to the target location on the river network respecting the topology of river lines (Figure 3.2.) by manually moving the point to the corresponding raster cell. The original relative location of waterfalls and special marking features on the river such as any meandering bends was preserved in the co-registered product so that the mapped river network with waterfall points can realistically reflect the relationship between waterfalls and channels.

Figure 3.2. The original point location (yellow dot) of the waterfall was moved onto the river line (blue line) at the bending vertex (red dot), preserving its original relative location with the actual river channel (left). Another waterfall point at the straight section of a river was moved onto the corresponding position on the river line, as the entire digital river line segment is slightly displaced towards the northwest from the true river segment (right).

Once all the point locations were modified, data quality ratings were assigned based on combined observations from the base map and supplemental validation sources (Table 3.1.). The original location and the adjusted location after co-registration were recorded for each data point. Unique point IDs were added for these newly processed points following the originally existing points in HydroFALLS v1a. By cleaning up the attribute information, HydroFALLS v1b was produced with 74 new points in China.

Table 3.1. Data quality rating assignment

Quality Rating	Description
1 (Confident)	Waterfall feature is verified by multiple reliable data and supplementary sources and there is no evidence to the contrary
2 (Strongly Suspected)	Waterfall feature is sourced by one reliable data source and verified by multiple unreliable data sources
3 (Weakly Suspected)	Waterfall feature is sourced by one reliable data source, but not verified by other data sources.
4 (Possible)	Waterfall feature is not uniquely identifiable but there is suspicion that waterfall feature exists at location
5 (Uncertain)	Waterfall feature is not expected to exist at location but there is no contradictory evidence

3.2. Global Datasets

Due to the difficulty of accessing individual national databases with varying data schemes, formats, completeness, and languages, governmental sources failed to contribute sufficient data points that cover the missing areas. Alternatively, 49,316 new data points were drawn from OpenStreetMap (OSM) (https://www.openstreetmap.org) with queries built from specified feature tag. Another 9545 new data points were extracted from the Geonames database (http://www.geonames.org/). In comparison with OSM which is characterized by unauthenticated features reported by users, Geonames is more systematic and accurate with relatively clean and ordered data points, thus serving as a good supplemental dataset for comparison and correction.

Step 1 aims to fill the missing name in HydroFALLS v1a with OSM or Geonames information. Points drawn from OSM and Geonames within 1 km of any nameless points in HydroFALLS v1b were paired up with these HydroFALLS v1b points based on proximity. Each pair of points were examined and verified using supplemental sources to make sure that they represent the same waterfall. If multiple names from OSM and Geonames existed or names with different languages were available, the feature name was formatted using a primary name with priorities on the English name or the name drawn from Geonames after verification. Another column containing alternative names was created to keep the full information, which was normally filled by secondary names.

In step 2, the 49,316 waterfall points obtained from OSM were queried with HydroFALLS v1b using selection by distance to clean up features that are already present in HydroFALLS v1b. 2229 points lying within 1km from any waterfalls in HydroFALLS v1b were separated since we can assume that they represent the same feature being mispositioned or marked differently due to their close proximity. Same procedures were applied to the 9545 points collected from Geonames, with 7702 points left more than 1km away from any points in HydroFALLS v1b. Since Geonames has higher accuracy than OSM, Geonames was prioritized if the same feature appear in both datasets. To clean up these duplicating features from OSM but keep them in Geoneames, the OSM dataset was then further filtered so that those within 1km from the

Geonames points were removed, leaving 45,503 points in OSM. Unique feature IDs were added to the 45,503 OSM points and the 7702 Geonames points, which produced 53,205 new waterfall points from the two global datasets.

In step 3, these points were then merged to a single layer to be mapped and co-processed with HydroFALLS v1b, with all the attribute information from both OSM and Geonames preserved. Since a standardized workflow of investigation from points with the highest to the lowest likelihoods of sitting on large rivers was aimed to be developed, these points were put into 25 ranks with different priorities based on the proximity to rivers with different discharges (Table 3.2.). Waterfalls closer to larger rivers are prioritized. The ranking scheme aims to prepare the data pool for future works following the workflow designed. To avoid repetitively counting the same set of points into multiple ranks (e.g. the point is located within 5 km of a large river and 1 km of a small river at the same time), points that have been grouped are separated from the pool before querying for the next rank. The thresholds of minimum river discharge and maximum distance to the targeting river of interest were set to be 1 m^3/sec and 5 km since waterfalls on minor tributaries or very far away from the river channel are not important in terms of river fragmentation at a global scale. These left-over points beyond the minimum or maximum thresholds were put into the lowest rank r25.

Table 3.2. Ranking scheme applied to the 53,205 points from OSM and Geonames.

Rank	River Discharge (cm^3/yr)	Distance to Rivers (km)
r1	> 1000	<1
r2		1-3
r3		3-5
r4	[500, 1000)	<1
r5		1-3
r6		3-5
r7	[100, 500)	<1
r8		1-3
r9		3-5
r10	[50, 1000)	<1

r11		1-3
r12		3-5
r13	[10, 50)	<1
r14		1-3
r15		3-5
r16	[5, 10)	<1
r17		1-3
r18		3-5
r19	[1, 5)	<1
r20		1-3
r21		3-5
r22	< 1	<1
r23		1-3
r24		3-5
r25	Points left over	

The highest rank, r1, were further examined following the same method as the regional database to estimate the general reliability and value of these new points. Specifically, the points were paired with the closest point from HydroFALLS v1b and checked whether the two points in each pair have a similar name. The names were considered to be the same as long as they have similar special words, e.g., *Vermilion Falls* in English vs. *Vermilion Chutes* in French, as well as *Mocona, Saltos del* vs. *Gran Salto de Macona*, despite the format or the spelling might not be exactly the same.

3.3. Application: Natural River Fragmentation and Fish Diversity

To test the applicability of the database in scientific studies, a global-scale analysis of freshwater fish diversity above and below waterfalls in large rivers was investigated to reveal potential fragmentation effects of waterfalls on fish evolution. By comparing the Freshwater Fish Species Richness (FFS) as a diversity index that refers to the number of freshwater fish species found, the geographic relationship between the genetic content of fish populations and waterfall locations can be evaluated.

Since other factors affecting fish diversity such as water properties, net primary productivity (NPP), habitat size, and longitudinal position of the watershed at the river network all add to the complexities of the fish biodiversity (Currie, 1991; Ricklefs, 1987), the method was designed to minimize the effects of these disturbing factors. Three global spatial datasets including BioMatrix, HydroSHEDS, and HydroFALLS v1b were overlaid. The basin-scale freshwater fish species listing BioMatrix (IUCN, 2021) includes comprehensively assessed taxonomic freshwater groups at a basin scale mapped to basins provided by HydroSHEDS. Only large rivers with an annual average discharge that is no less than 10 m^3/s were selected for analysis to maximize the potential FFS detected so that a more obvious comparison can be obtained.

These datasets were then grouped by 6 ecozones (Nearctic, Palearctic, Neotropic, Afrotropic, Indo-Malaya, Australasia) adopted from Udvardy (1975)'s division system to take rough controls on the commonality of evolutionary history and biotic characteristics, which also to some degree coincide with the NPP. To minimize the impact of habitat size, the watersheds directly upstream and downstream of the waterfall must have a visually similar size to be selected. While the confluence of small tributaries was unavoidable due to the nature of watershed division, those without visible tributaries (large ones with an annual discharge no less than 10 m^3/s) joining the target stream in three consecutive watersheds were preferred, as this normally leaves only watersheds with a similar discharge upstream and downstream by restricting the confluence of large tributaries into the main stream. Furthermore, waterfalls with a larger height and discharge were prioritized. Based on the data availability and result after initial clean-ups specified by the criteria, most of the waterfall points were ruled out from further analysis, leaving only approximately 10 target points in each ecozone. In the end, 11 suitable waterfalls were selected for each of them as treatment groups due to comprehensive consideration for all ecozones.

To further filter out the disturbing impact of the longitudinal position and the order of the watershed along the river, a control group with the same sample size of 11 but with no waterfall present in the middle watershed was also selected for each ecozone for further comparison (Figure 3.3.). Thus, 2 groups each with 11 locations were analyzed for each ecozone.

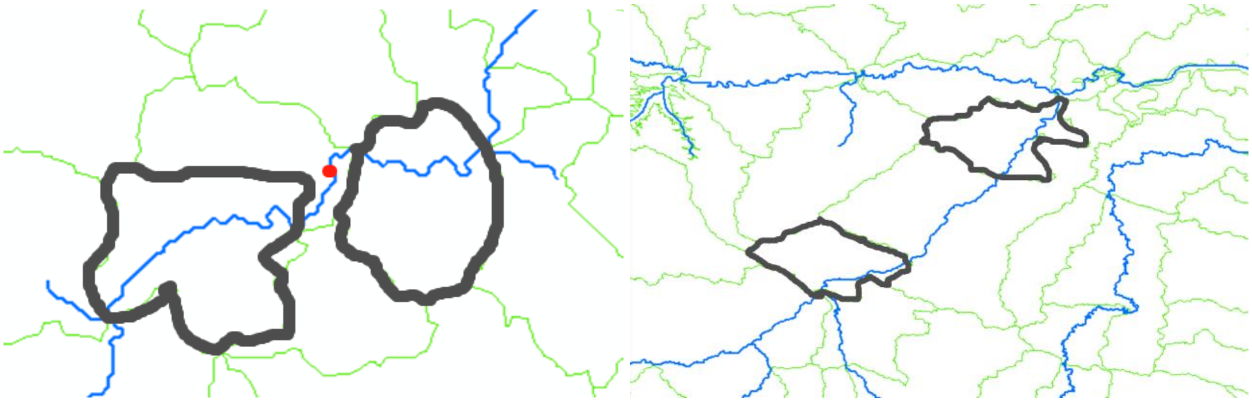

Figure 3.3. Example of treatment sites (left) with the waterfall (red dot) on the river (blue line) crossing watersheds (green polygon), and control sites (right) without waterfalls in the middle watershed. The upstream and downstream watersheds (outlined in black) were selected to have a visually similar size, the middle watershed must have no large tributaries joining in the main stream.

For each location, the FFS in the two watersheds directly upstream and downstream of the middle watershed where one or multiple waterfall(s) sit were recorded as specified in BioMatrix. The difference between downstream and upstream FFS was calculated. Lastly, the statistical significance of the overall difference between the control group and the treatment group for each ecozone was analyzed using the two-sample t-test.

CHAPTER 4: RESULT

4.1. Regional Gap-filling in China

The 74 waterfalls added for China (Figure 4.1.) were merged into the global database, creating the new version HydroFALLS v1b containing 4128 points in total. Most of the waterfalls are in southern China falling within the Yangtze River Basin, Pearl River Basin (PRB), a small part of the Lancang River (Upper Mekong River) Basin, with Huangguoshu Fall being the highest waterfall in China. Six waterfalls sit on the Yellow River Basin (YRB), including Hukou Fall which has the largest discharge among waterfalls in China. Less than 10 waterfall points were found in the Brahmaputra River Basin in the Tibetan Plateau, including high mountain regions.

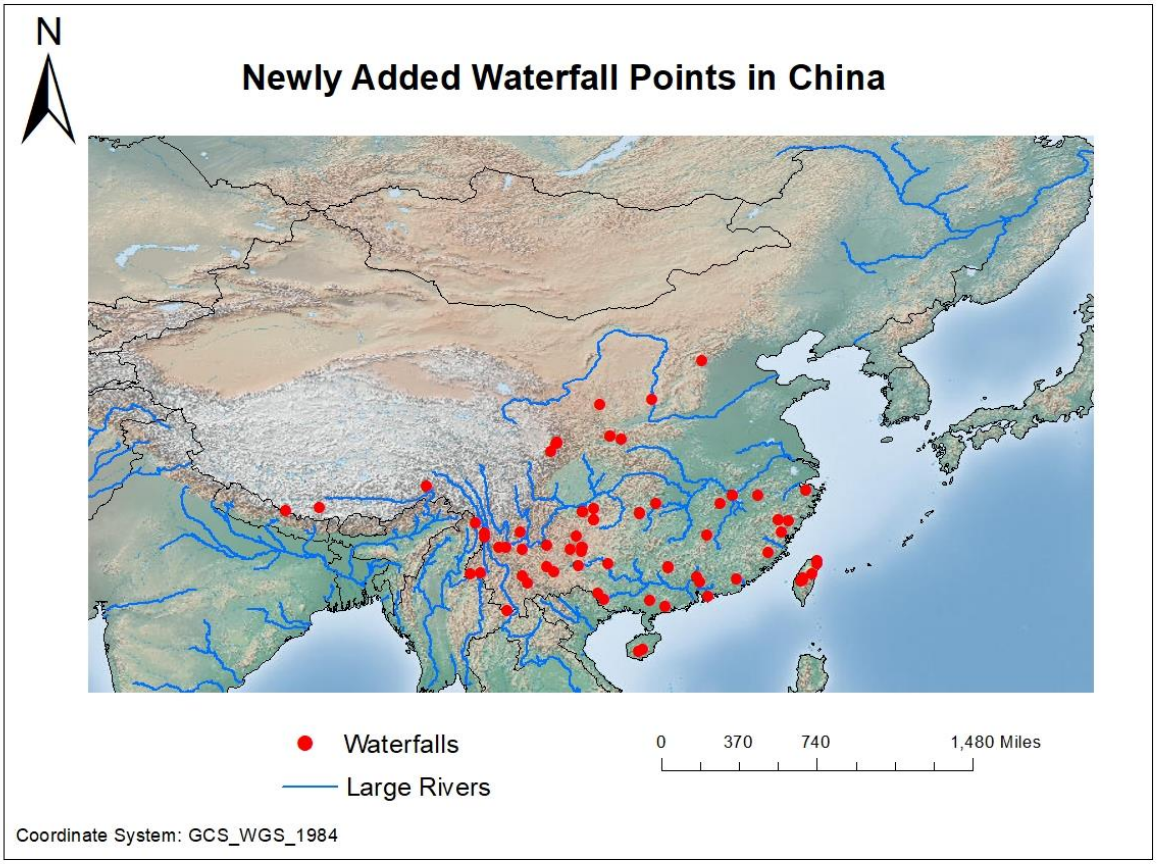

Figure 4.1. Overview map of the 74 newly added waterfall points in China, co-registered with HydroSHEDS.

With over half of the points receiving the highest quality rating and none of the points falling in the poorest two quality ratings (Table 4.1.), the overall accuracy of this dataset is relatively high. However, 33% of the points are not identified on satellite images or other sources. Many of the features are blocked by ground features such as forests or failed to be spotted due to shades or unclear scenes. The other 16% are verified by unreliable data sources such as tourism websites and personal blogs.

Table 4.1. Data quality rating assigned to the 74 newly added waterfall points in China.

Quality Rating	Number of Points	Percentage (%)
1 (Confident)	38	51
2 (Strongly Suspected)	12	16
3 (Weakly Suspected)	24	33
4 (Possible)	0	0
5 (Uncertain)	0	0

4.2. Global Gap-filling

Spatial overlaps between HydroFALLS v1b points and the other two global datasets are concentrated in regions where HydroFALLS v1b points are present (Figure 4.2.). This suggests that OSM and Geonames do coincide with the majority of HydroFALLS v1b, confirming their validity. At the same time, they provide more extensive coverage of waterfall points, which serve as good supplemental data pools for new points after further examination and clean-up. The major regions missing or insufficient in HydroFALLS v1b but present in Geonames or OSM are the west coast of North and South America, Mexico, Europe, the Middle East, as well as West and Central Asia.

At a global scale, the distribution of waterfalls roughly follows the ground relief, such as transitions from high mountains to plains. Major regions with high waterfall density include the Rocky Mountains, Colorado Plateau, and Laurentian Plateau in North America, Andes Mountains, Brazilian Highlands, and Amazon Basin in South America, Great Rift Valley and

Congo Basin in Africa, entire Europe, as well as some coastal areas such as Japan, New Zealand, and Southeast Asia.

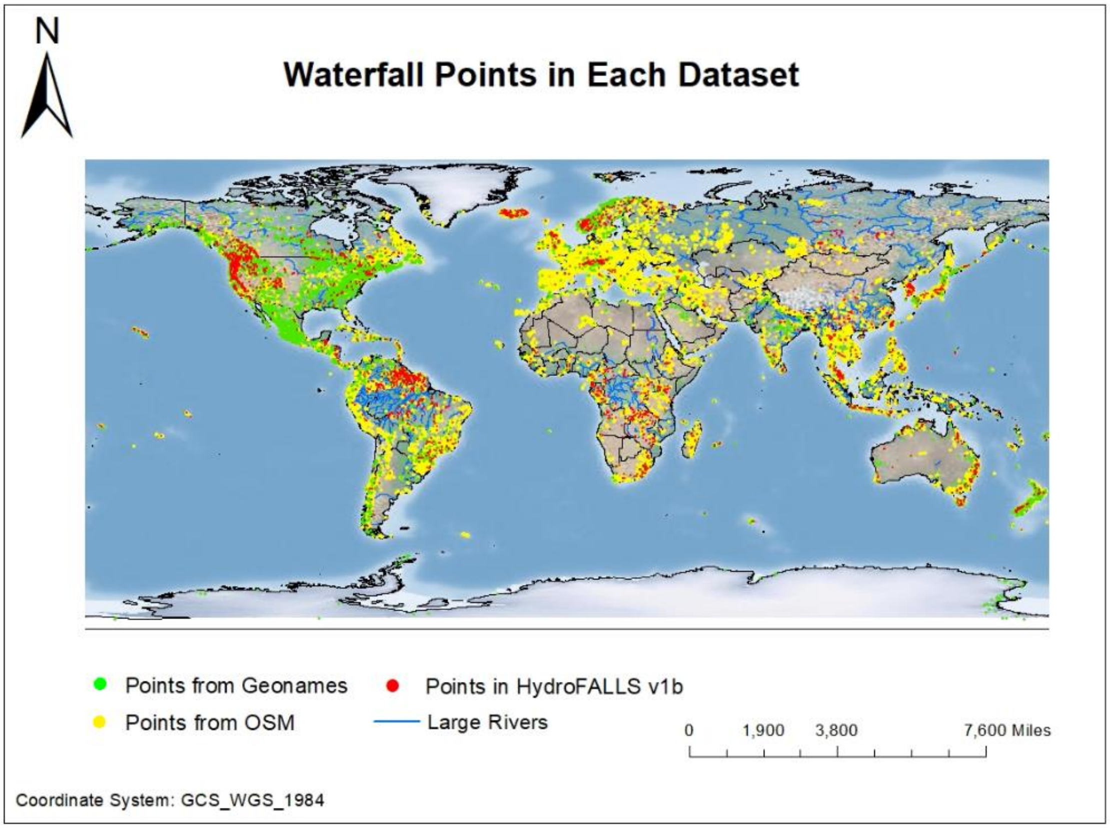

Figure 4.2. Overview Map of waterfall points in OSM, Geonames, and HydroFALLS v1b.

The 53,205 new points extracted from OSM and Geonames after clean-ups were ranked to 25 priority classes (Table 4.2.) based on the river discharge and distance to the target river. Although the number of points does not show a monotonically increasing trend with lower prioritization, the general pattern shows that more waterfalls are found near small rivers and further away from the target rivers.

Table 4.2. Priority ranking of newly collected points from OSM and Geonames.

Rank	Number of Points	River Discharge (cm^3/yr)	Distance to Rivers (km)
r1	240	> 1000	<1
r2	218		1-3

r3	153		3-5
r4	172	[500, 1000)	<1
r5	135		1-3
r6	93		3-5
r7	1146	[100, 500)	<1
r8	765		1-3
r9	565		3-5
r10	959	[50, 1000)	<1
r11	671		1-3
r12	590		3-5
r13	3434	[10, 50)	<1
r14	2423		1-3
r15	2160		3-5
r16	1782	[5, 10)	<1
r17	1653		1-3
r18	2002		3-5
r19	7250	[1, 5)	<1
r20	5873		1-3
r21	5331		3-5
r22	7678	< 1	<1
r23	5512		1-3
r24	1284		3-5
r25	1116	Points left over	

Among the 240 points in the highest rank r1, 21 points have the same name as the closest point found in HydroFALLS v1b (Figure 4.3.). Considering their proximity and name consistency, the two points in each pair are considered to represent the same feature. Since the point locations in HydroFALLS v1b have already been snapped to the river network of HydroSHEDS, these 21 points from OSM or Geonames only provide supplemental attribute information such as alternative names and height. Among the other 219 points with different names from the closest point in HydroFALLS v1b, 27 points do not represent any waterfalls as multiple supplemental sources confirmed that no waterfalls exist nearby. 76 waterfall points are not seen at the location.

One possibility is that the waterfall is seasonal, which is not present as a water feature at the time when the satellite image was taken. Other cases include being too small to be discovered, getting blocked by ground features such as forests, or simply non-existing as the previous 27 points.
For the remaining 116 points, waterfall-like features were confirmed nearby. Of those, 18 of them being confirmed to be actual waterfalls whereas the other 98 points were supported by partial or related visual confirmation. Several cases include (1) Only part of the waterfall system is seen. Features normally appearing together with waterfalls such as plunge pools are signs of waterfall presence, but they do not necessarily claim a waterfall nearby; (2) White water splash is seen, but the elevation discontinuity is not clear; (3) It is hard to differentiate whether the structure is natural (e.g. rocky cliff) or man-made (e.g. concrete wall), which confuses waterfalls with dams; (4) The satellite image is not clear to give the distinct shape of the waterfall.

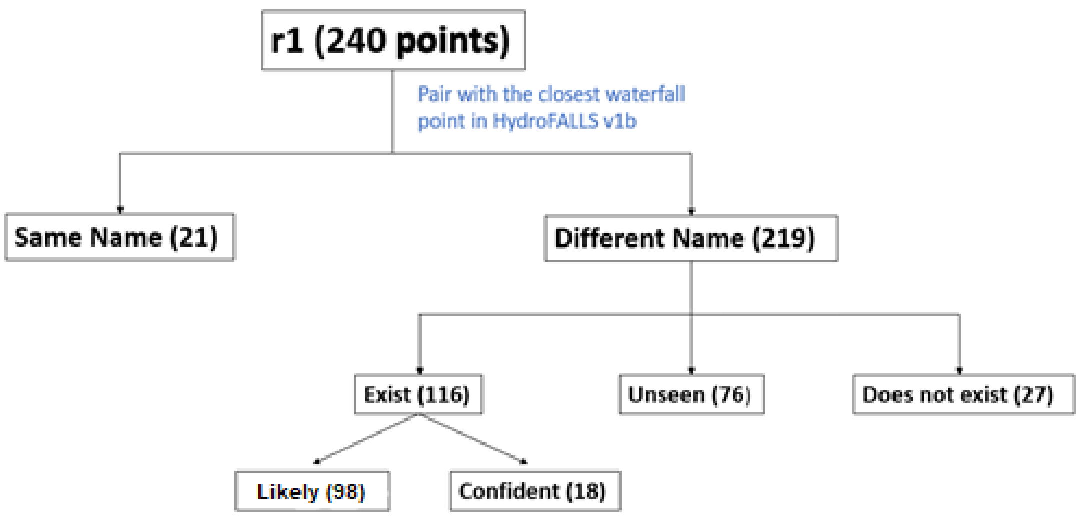

Figure 4.3. Data quality of waterfall points in rank 1.

4.3. Natural River Fragmentation and Fish Diversity

For all ecozones, the mean difference between downstream and upstream FFS is larger in the treatment group than in the control group (Table 4.3.). For Nearctic, Palearctic and Australasia, this is further supported by the general pattern of range and distribution of values in the two groups (Figure 4.4.). In comparison, the general pattern in Neotropic, Afrotropic and Indo-Malaya is less pronounced. This is confirmed by the low statistical significance (<90%) of the difference between the control and the treatment group in these three ecozones, in comparison to the result for Nearctic, Palearctic and Australasia with a significance larger than 95% or between 90% and 95%. The main spread of values in Neotropic shows a similar pattern except for a high upper outlier in the control group. The result for Afrotropic has a large variance with differences in both the control and the treatment group spread out over a larger range. The control group for Indo-Malaya also has a larger variance than the treatment group.

The occurrence of negative difference values representing a higher FFS upstream is more frequent in the control group for each ecozone, suggesting that sites in the control groups are more affected by natural variabilities and complex factors. The mean difference between downstream and upstream FFS occurs to be positive for both treatment groups and control groups for all ecozones, with an exception of the control group for Afrotropic.

Table 4.3. The mean difference, standard deviation, and the confidence level of the difference between the control group (C) and the treatment group (T).

	Mean Difference		**Standard Deviation**		**Confidence Level of the Difference Between C and T**
	C	**T**	**C**	**T**	
Nearctic	0.55	3.36	3.00	7.06	90-95%
Palearctic	1.56	3.55	1.71	2.39	> 95%
Neotropic	1.09	3.18	3.42	4.22	< 90%
Afrotropic	-1.91	1.73	10.26	10.45	< 90%
Indo-Malaya	2.82	5.45	2.72	3.37	< 90%
Australasia	1.73	8.91	4.94	7.34	> 95%

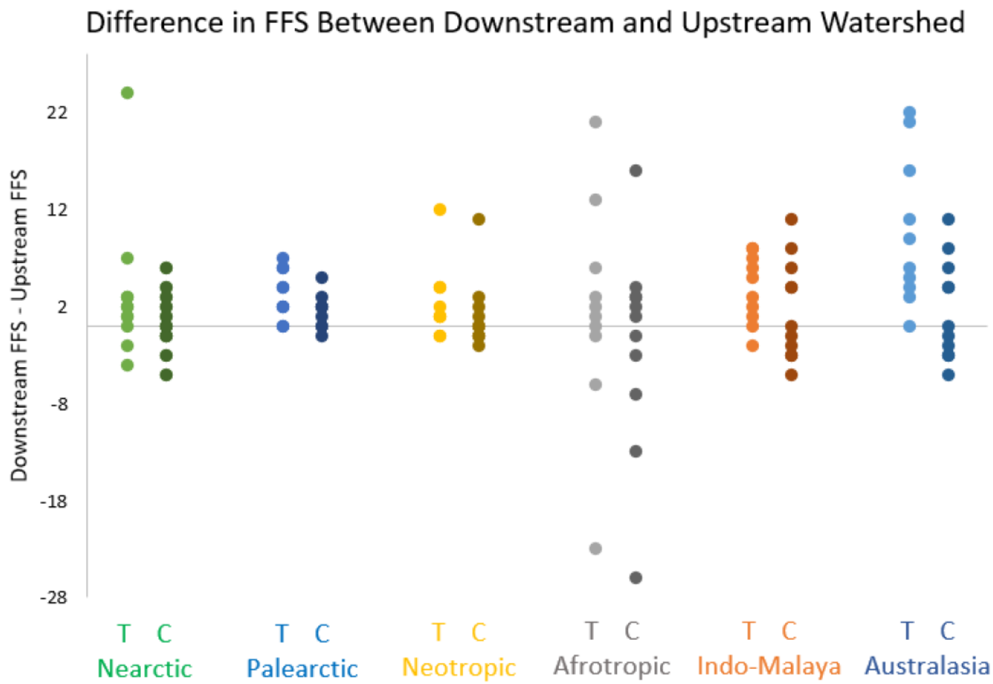

Figure 4.4. Difference in FFS between the downstream and upstream watershed (Downstream FFS - Upstream FFS) for the treatment group (T) and the control group (C) for each ecozone. A positive value means that the FFS in the downstream watershed is higher than that upstream.

CHAPTER 5: DISCUSSION

5.1. Waterfalls in China

The distribution of waterfalls in China well matches with major river basins, with the majority of them falling in the southern part, particularly the southwest region. The Yellow River originates from the Tibet Plateau, flowing through the Loess Plateau to the North China Plain (Figure 5.1.). Faults and mountain ranges in the YRB are mostly in the east-west direction align with the flow direction. High-order streams of the Yellow River rarely cross vast regions with highly fluctuated topography. In comparison, southern rivers flow through a series of north-south mountain ranges. The Hengduan Mountain Range at the transition from Tibet Plateau to Sichuan Basin and Yunnan-Kweichow Plateau further east cuts the surface into topographically isolated pieces. A few small-ranged fault systems at Yangtze Plain and southeastern hill regions also create waterfalls at the Yangtze River Basin and the PRB.

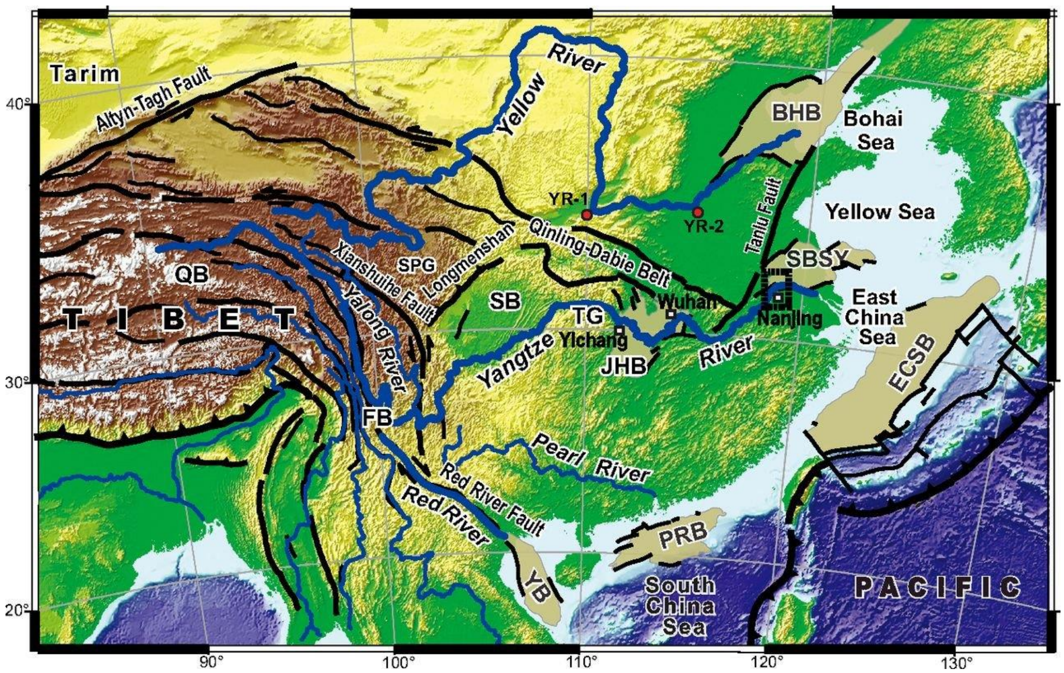

Figure 5.1. Topographic map of China with major rivers (blue line) and faults (Zheng et al., 2013).

Adding to the overall high data quality rating (Table 3.1.), the consistency between the waterfalls collected and the topographic pattern of China confirms the reliability of the data source. However, this dataset only catches 74 major waterfalls at a national scale. It is highly likely that numerous smaller waterfalls exist that are not included in the data. Therefore, this dataset gives relatively reliable, high-quality, but incomplete data. Nevertheless, these major waterfalls have higher environmental significance than smaller ones excluded from this dataset, so it successfully fills the major gap of HydroFALLS v1a in China.

5.2. Waterfalls from OSM and Geonames

At a global scale, the distribution of waterfalls in HydroFALLS v1b generally matches with points in OSM and Geonames, but it has a much smaller coverage and point density than the two supplemental datasets, particularly OSM which is contributed and updated extensively by users all around the world. Major geographical gaps in HydroFALLS v1b that are present in OSM or Geonames reflect regions that are either underrepresented in global datasets or lack national datasets. Regional completeness in global datasets is affected by a variety of factors such as the accessibility of the regional language, interest of local groups to waterfall points, personal preferences of dataset builders, and access to geographical resources (e.g. internet, satellite images, tourism development).

As suggested by the broad spatial coverage and a large number of waterfall points collected, OSM and Geonames serve as data pools with a high level of detail. They contain waterfalls of various sizes sitting on water features from small gullies to major river channels. However, completeness coexists with mess, duplication, suspicion, and wrong information that raise challenges on data cleaning and processing. Since waterfalls on larger rivers are prioritized, the ranking scheme effectively extracts those of higher interest by isolating a small portion within a few high-order ranks and leaving the other to low-order ranks which are contributing less important information to the database. Of the 240 points in r1, 18 are confidently identified waterfalls sitting on large rivers, which can be added to later versions of HydroFALLS. The 184 unseen and suspect points also have the potentials to be confirmed as waterfalls as long as further data sources or updated satellite images are available. In-depth examinations on these points would be necessary to accurately spot, locate, and verify the features. Since HydroFALLS v1b already covers major waterfalls on large rivers, it is expected that lower ranks with a smaller targeting river but higher proximity to the river might have higher potentials to contribute new points, despite these waterfalls being less important.

Of waterfalls that exist in both HydroFALLS v1b and at least one of the two supplemental datasets, most points in HydroFALLS v1b are verified by nearby data points in the other two datasets via similar names and descriptions. From this point of view, the two supplemental datasets not only contribute extra data points but also serve for validation purpose.

5.3. Natural River Fragmentation and Fish Diversity

The relatively clear pattern of larger differences between downstream and upstream FFS obtained from the treatment groups for Nearctic, Palearctic and Australasia indicate that waterfalls in these regions are potentially contributing to a larger gradient in fish biodiversity and genetic compositions of local populations above and below waterfalls. For Neotropic, Afrotropic and Indo-Malaya, the pattern does not statistically support the same argument. Although the mean difference is always higher for treatment groups, the result involves obvious complexities, outliers and variabilities. Although the division of world ecozones is not based on the climate or physical characteristics of the regions, the result shows that the three ecozones with an unclear pattern roughly coincide with regions that have more tropical characteristics, such as South and Southeast Asia, Africa, and South America. This might be explained by the complicated and actively interacting ecological processes in the tropics and other highly productive regions (Brown, 2013). This spatial pattern of FFS has further implications on fisheries and human livelihoods along the river. Processes such as the seasonal migratory behaviours underpinning the variation in fish species and population size above and below waterfalls might help fishers to optimize the timing, practices, as well as the choice of fishing sites and species (Ikpi et al., 2009; Thé and Nordi, 2006; Sant'Anna et al., 2014; Roberts and Baird, 1995). Moreover, this pattern helps to recognize that in addition to human activities that have been disconnecting fluvial processes such as dam construction, water degradation and water use (Grill et al., 2019), natural factors such as waterfalls are also affecting the aquatic ecosystems. This is important in differentiating and isolating anthropogenic fragmentations so that natural factors can be ruled out.

However, these natural factors are not limited to waterfalls. Complicated environmental processes as well as interplaying biotic and abiotic conditions are adding large complexities and uncertainties to the study, which might explain the ambiguous result for Neotropic, Afrotropic and Indo-Malaya. The entire Lower Congo River Basin is found to be fragmented by the special flow regime, hydrological processes, and water properties as a result of the complicated climate and topography (McKaye and Gray, 1984). The reaches with high-gradient riverbeds, high annual discharge, and dense distribution of waterfalls of different sizes are rich in rapids that create small shelters in relatively still areas, isolating fish species and promoting speciation; the presence of interior drainages since Miocene and large lakes in the Cuvette Centrale trap considerable aquatic populations, forming isolated inland water habitats; the evaporative cooling generated by strong turbulence and the confluences of numerous tributaries give downstream reaches a lower water temperature, which is also responsible for the reverse trend of lower downstream FFS. This large-scale isolation is further supported by the discovery of 34 species endemic to near-rapid habitats along the main stream of the Lower Congo (Roberts and Stewart, 1976), as well as the high occurrence of negative differences in both the treatment group and the control group for Afrotropic. The large variance and the presence of large downstream and upstream differences are caused by the high FFS inherent to the Afrotropic, which is consistent with the rich biodiversity and productive land covers in this ecozone.

A characteristic of Indo-Malaya is the mixed distribution of oceans and islands with long coastlines (Loucks et al., 2008), which has potentially strengthened the biogeographical island effect and promoted gene exchanges between freshwater and marine water. Although the first 3 ecozones show a pattern potentially caused by waterfalls, there are plenty of regional complexities as well. For example, the headwater systems of many rivers in Australasia are only a few watersheds inland from the river mouth, which tends to give these aquatic habitats a steep longitudinal gradient of FFS as a result of the absence of aquatic species in the inland desert followed by a sudden increase in FFS as moving downstream.

In addition to natural complexities, anthropogenic disconnectivity such as damming, fishing, and sediment mining are already playing a role in the current spatial pattern of freshwater fish species, which makes it difficult to achieve the aim of distinguishing natural and anthropogenic fragmentation. Studies focusing on either topic are strongly influenced by the other, this disturbance is particularly strong in regions with highly disconnected rivers, such as China, India, the US, and some parts of Europe (Grill et al., 2019). The spatial unevenness of the degree of disturbance is also contributing to the variation in the powerfulness of the pattern shown for different ecozones.

5.4. Challenges and Limitations

To filter out potential new features from a large population of data points extracted, the distance threshold was set to 1km. Although a data point falling within 1km from an existing one are likely to represent the same feature as its joint point, there is still a small portion of cases in which it indeed introduces a new waterfall. Nevertheless, the differentiation would not be scientifically significant as large-scale or even global fragmentation studies mainly focus on scales that are large enough to omit processes that occur within 1km.

In addition, many waterfall points are not clearly spotted on satellite images, which adds difficulties to co-registration with HydroSHEDS. Locations of waterfalls blocked by clouds or ground features are not able to be identified on clearly visible water channels. Some waterfalls joining from mountain tributaries into the central river are hanging on the cliff and easy to be confused with the cut-section of rocks. Due to the low-quality DEM used in HydroSHEDS for northern high-latitude regions, the digital river network in these areas is misplaced from the actual location of rivers and distorted in topology, making it harder to co-locate waterfalls onto HydroSHEDS.

Due to the lack of systematic and reliable waterfall metadata, related attributes such as height, width, and seasonality are partly or completely missing for a large portion of the waterfalls collected. The lack of an appropriate classification scheme also makes it hard to specify the type of the waterfall, leading to inconsistent criteria of defining waterfalls at a global scale. Without specific waterfall information, their impact is more difficult to be quantified and correlated with other environmental consequences.

These problems lead to obvious constraints on the application of the database. To study natural river fragmentation and FFS pattern, a size of eleven points in each group is insufficient in the context of thousands of waterfalls in the world. Although the sample size is limited by the accuracy of watershed delineation of the fish species data, it is worth noting that a more complete global waterfall database would effectively enlarge the pool of waterfall points that satisfy the criteria specified for watershed-scale FFS comparison. If attribute information is available, regression analysis between waterfall parameters (e.g. height) and the FFS difference would also be achieved to investigate the quantitative relationship between factors of fragmentation and FFS distribution.

Nevertheless, other problems of this application study in addition to data constraints are also causing inaccuracies and biases. The simple approach taking species richness as the single index evaluating biodiversity is insufficient to tell the full story, such as the migratory behaviors, morphological adjustments, and spatial uniqueness of certain species potentially created by waterfall fragmentation. Other studies conducting laboratory genetic and taxonomic analysis on fish samples, taking multiple speciation metrics such as the endemic FFS as well as the ratio of the number of genera with two or more endemic species to the number of genera with one or more endemic species, or comparing the occurrences of individual species are helpful in revealing detailed mechanisms of speciation processes influenced by waterfalls (Deiner et al., 2006; Haugen et al., 2008; Decru et al., 2015; Phillips et al., 2009). In fact, these metrics might be less disturbed by natural complexities of the surrounding environment, as they focus specifically on certain aspects of speciation which accurately reflect fragmentation effects.

Meanwhile, this study fails to quantify the impacts of common disturbing factors as statistical predictors. The habitat size, longitudinal position of the watershed, and NPP are estimated and roughly controlled to be similar between pairs of comparison, but their exact influences are not modelled as dependent variables into the approach. This also loses control on the comparison between control and treatment groups because it is not guaranteed that the overall level of discharge, local NPP and habitat size of the sample locations selected for control and treatment groups as well as for different ecozones are always similar. Many waterfalls in the Palearctic, Indo-Malaya and Australasia sit on low-discharge streams, which might produce biased results considering other ecozones rich in high-discharge waterfalls. The analysis scale is also relatively small with high-order watershed division, which tends to include local variabilities.

5.5. Future Works

To extract valuable waterfall points from OSM and Geonames, each rank needs to be further examined, compared, and verified with supplemental data sources. It is estimated that a few hundred new points would be collected from the 25 ranks, despite requiring a large amount of work. National datasets with high reliability would also be a main focus. Many countries already built their governmental datasets but it requires further exploration to access each dataset and collect waterfall points. Attributes of each waterfall need to be filled and verified using other data sources to improve the completeness of quantitative parameters. In the end, differentiation between natural and anthropogenic river fragmentation requires in-depth investigation of each waterfall, co-registration with global dam databases, and consolidation of the two fragmentation types.

The application study highlights the feasibility of global studies on natural habitat fragmentation by waterfalls, as shown by the clearly higher difference spotted from treatment groups for Nearctic, Palearctic and Australasia as well as the weak but visible pattern in Afrotropic, Neotropic and Indo-Malaya. This result is caused by various degrees of disturbances by natural complexities regulating freshwater fish biodiversity in different ecozones, regional uniqueness, and intrinsic limitations of the approach taken. Considering these regional variations, it is not practical and meaningful to generalize the fragmentation effect to the entire world as a single study region, but this phenomenon can definitely be globalized by systematic and large-scale studies to assess its commonality across larger extents with similar environmental characteristics, such as evaluations grouped by continents, ecozones or entire major river basins, to add onto random case studies targeting at individual waterfalls or river reaches. Comparisons and contrasts among different study regions can also be made to investigate the spatial similarities and variations of different regions composing the globe.

CHAPTER 6. CONCLUSION

This thesis explored potential approaches of building a global waterfall database by adding new points from various national and global datasets, as well as tested the usage of the database developed by studying the relationship of basin-scale FFS difference and waterfall distribution. Through data preparation, consolidation, validation, and co-registration, 74 new points were collected from a national dataset of China and located on HydroSHEDS, which filled a major spatial gap in HydroFALLS v1a. 53,205 points extracted from two global datasets, OSM and Geonames, were ranked to 25 classes with different priorities. The highest rank produced 18 new waterfall points that can be added to later versions of HydroFALLS, which gave a rough estimation of a few hundred points in total that can potentially be drawn from the 25 classes in reasonably short time.

Nevertheless, the updated database, HydroFALLS v1b, still has remaining problems. As a global database, different regions differ tremendously in data completeness, accuracy, and coverage. Attributes are strongly incomplete with a large portion of related waterfall parameters left empty. Together with the lack of consistent classification schemes of waterfalls, there is a spatial unevenness in waterfall type and seasonality, thus there are different levels of environmental significance that are hard to be quantitatively analyzed. These limitations add constraints to applications of the database via spatial variation in data availability and lack of waterfall attributes as factors of the fragmentation level.

In addition to data limitation, global-scale natural river fragmentation studies are also disturbed by spatial variations in natural conditions, making it hard to produce unified and consistent fragmentation processes at a global scale. However, large-scale studies grouped by regions with similar natural and ecological conditions are feasible and meaningful, which would require further improvement and updates of global datasets. In this sense, the standardized workflow developed in this thesis following the data point ranking scheme provides important approach for consistent works in the future.

REFERENCES

Baker, K., Chadwick, M. A., Wahab, R. A., & Kahar, R. (2016). Benthic community structure and ecosystem functions in above- and below-waterfall pools in Borneo. *Hydrobiologia*, *787*(1), 307–322. https://doi.org/10.1007/s10750-016-2975-4

Beisel, R. H. (2006). *International Waterfall Classification System*. Outskirts Press.

Brown, J. H. (2013). Why are there so many species in the tropics? *Journal of Biogeography*, *41*(1), 8–22. https://doi.org/10.1111/jbi.12228

Cathcart, C. N., Pennock, C. A., Cheek, C. A., McKinstry, M. C., MacKinnon, P. D., Conner, M. M., & Gido, K. B. (2018). Waterfall formation at a Desert River-reservoir delta isolates endangered fishes. *River Research and Applications, 34*(8), 948–956. https://doi.org/10.1002/rra.3341

Chen, J., Zhang, D. D., Wang, S., Xiao, T., & Huang, R. (2004). Factors controlling tufa deposition in natural waters at Waterfall Sites. *Sedimentary Geology*, *166*(3-4), 353–366. https://doi.org/10.1016/j.sedgeo.2004.02.003

Cote, D., Kehler, D. G., Bourne, C., & Wiersma, Y. F. (2008). A new measure of longitudinal connectivity for Stream Networks. *Landscape Ecology, 24*(1), 101–113. https://doi.org/10.1007/s10980-008-9283-y

Currie, D. J. (1991). Energy and large-scale patterns of animal- and plant-species richness. *The American Naturalist*, *137*(1), 27–49. https://doi.org/10.1086/285144

Decru, E., Moelants, T., De Gelas, K., Vreven, E., Verheyen, E., & Snoeks, J. (2015). Taxonomic challenges in freshwater fishes: A mismatch between morphology and DNA barcoding in fish of the north-eastern part of the Congo Basin. *Molecular Ecology Resources*, *16*(1), 342–352. https://doi.org/10.1111/1755-0998.12445

Defense Mapping Agency (DMA), 1992. *Digital Chart of the World.* Defense Mapping Agency, Fairfax, Virginia. (Four CD-ROMs.)

Deiner, K., Garza, J. C., Coey, R., & Girman, D. J. (2006). Population structure and genetic diversity of trout (Oncorhynchus mykiss) above and below natural and man-made barriers in the Russian River, California. *Conservation Genetics*, *8*(2), 437–454. https://doi.org/10.1007/s10592-006-9183-0

Dias, M. S., Cornu, J.-F., Oberdorff, T., Lasso, C. A., & Tedesco, P. A. (2012). Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography, 36*(6), 683–689. https://doi.org/10.1111/j.1600-0587.2012.07724.x

DiBiase, R. A., Whipple, K. X., Lamb, M. P., & Heimsath, A. M. (2014). The role of waterfalls and Knickzones in controlling the style and pace of landscape adjustment in the western San Gabriel Mountains, California. *Geological Society of America Bulletin*, *127*(3-4), 539–559. https://doi.org/10.1130/b31113.1

Ford D.C., 1968. Waterfalls. In: R.W.Fairbridge (ed.), *The Encyclopedia of Geomorphology*. Reinhold, 1219–20.

GeoNames Database. NGA (National Geospatial-Intelligence Agency). (2013). Retrieved December 13, 2021, from https://geonames.nga.mil/gns/html/.

GeoNames. (n.d.). Retrieved March 28, 2022, from http://www.geonames.org/

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M. E., Meng, J., Mulligan, M., … Zarfl, C. (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215–221. https://doi.org/10.1038/s41586-019-1111-9

Haugen, T. O., Aass, P., Stenseth, N. C., & Vøllestad, L. A. (2008). Changes in selection and evolutionary responses in migratory brown trout following the construction of a fish ladder. *Evolutionary Applications*, *1*(2), 319–335. https://doi.org/10.1111/j.1752-4571.2008.00031.x

Haviv, I., Enzel, Y., Whipple, K. X., Zilberman, E., Matmon, A., Stone, J., & Fifield, K. L. (2010). Evolution of vertical knickpoints (waterfalls) with resistant caprock: Insights from numerical modeling. *Journal of Geophysical Research*, *115*(F3). https://doi.org/10.1029/2008jf001187

Hayakawa, Y. S., & Matsukura, Y. (2010). Stability Analysis of Waterfall Cliff Face at Niagara Falls: An implication to erosional mechanism of Waterfall. *Engineering Geology*, *116*(1-2), 178–183. https://doi.org/10.1016/j.enggeo.2010.08.004

Herschy, R. W. (1998). Waterfalls. In *Encyclopedia of Hydrology and Lakes*. *Encyclopedia of Earth Science.* Springer, 695-698.

Hudson, B. J. (1998). Waterfalls Resources for Tourism. *Annals of Tourism Research*, *25*(4), 958–973. https://doi.org/10.1016/s0160-7383(98)00043-7

Hudson, B. J. (1999). Fall of beauty: The story of a Jamaican waterfall — a tragedy in three acts. *Tourism Geographies*, *1*(3), 343–357. https://doi.org/10.1080/14616689908721326

Hudson, B. J. (2001). *Waterfalls of jamaica: Sublime and beautiful objects*. University of the West Indies Press.

Hudson, B. J. (2013). Waterfalls, science and aesthetics. *Journal of Cultural Geography*, *30*(3), 356–379. https://doi.org/10.1080/08873631.2013.828482

Ikpi, G. U., Jenyo-Oni, A., Ojo, S. O., & Akinyemi, A. (2009). Sport fisheries potentials of agbokim waterfalls, Cross River State, Nigeria. *Journal of Agriculture, Forestry and the Social Sciences*, *5*(2). https://doi.org/10.4314/joafss.v5i2.46722

Jui, F. S., Alam, S., Alam, M. D., & Chowdhury, S. (2015). A feasibility study of mini hydroelectric power plant at Sahasradhara Waterfall, Sitakunda, Bangladesh. *2015 International Conference on Advances in Electrical Engineering (ICAEE)*. https://doi.org/10.1109/icaee.2015.7506801

Kano, Y., Nishida, S., & Nakajima, J. (2012). Waterfalls drive parallel evolution in a freshwater goby. *Ecology and Evolution*, *2*(8), 1805–1817. https://doi.org/10.1002/ece3.295

Lehner, B., Verdin, K., Jarvis, A. (2008): New global hydrography derived from spaceborne elevation data. Eos, Transactions, AGU, 89(10): 93-94.

Lehner, B. (2013): Global Waterfall Database, Technical Documentation, Version 1.0. Available at www.hydrosheds.org

Leibowitz, Z. W., Brito, L. A., De Lima, P. V., Eskinazi-Sant’Anna, E. M., & Barros, N. O. (2017). Significant changes in water PCO2 caused by turbulence from waterfalls. *Limnologica*, *62*, 1–4. https://doi.org/10.1016/j.limno.2016.09.008

Lobeck, A. K. (1939). *Geomorphology*. McGraw-Hill.

Losos, J. B., & Parent, C. E. (2009). The speciation-area relationship. *The Theory of Island Biogeography Revisited*, 415–438. https://doi.org/10.1515/9781400831920.415

Loucks, C., Ricketts, T. H., Naidoo, R., Lamoreux, J., & Hoekstra, J. (2008). Explaining the global pattern of protected area coverage: Relative importance of vertebrate biodiversity, human activities and agricultural suitability. *Journal of Biogeography*, *35*(8), 1337–1348. https://doi.org/10.1111/j.1365-2699.2008.01899.x

Luts, A., Parts, T.-E., Laakso, L., Hirsikko, A., Grönholm, T., & Kulmala, M. (2009). Some air electricity phenomena caused by waterfalls: Correlative study of the Spectra. *Atmospheric Research*, *91*(2-4), 229–237. https://doi.org/10.1016/j.atmosres.2008.02.019

Mabin, M. C. G. (2000). In search of Australia's highest waterfalls. *Australian Geographical Studies*, *38*(1), 85–90. https://doi.org/10.1111/1467-8470.00102

McKaye, K. R., & Gray, W. N. (1984). Extrinsic barriers to gene flow in rock-dwelling cichlids of Lake Malawi: Macrohabitat heterogeneity and reef colonization.

McLachlan, M., MacKay, D., & Jones, P. H. (1990). A conceptual model of organic chemical volatilization at waterfalls. *Environmental Science & Technology*, *24*(2), 252–257. https://doi.org/10.1021/es00072a015

Natchimuthu, S., Wallin, M. B., Klemedtsson, L., & Bastviken, D. (2017). Spatio-temporal patterns of stream methane and carbon dioxide emissions in a hemiboreal catchment in southwest Sweden. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep39729

National Geographic Society. (2012, October 9). *Waterfall*. National Geographic Society. Retrieved December 13, 2021, from https://www.nationalgeographic.org/encyclopedia/waterfall/.

Northcote, T. G. (1981). Juvenile current response, growth and maturity of above and below waterfall stocks of rainbow trout, Salmo gairdneri. *Journal of Fish Biology*, *18*(6), 741–751. https://doi.org/10.1111/j.1095-8649.1981.tb03815.x

Northcote, T. G. (2010). Controls for trout and char migratory/resident behaviour mainly in stream systems above and below waterfalls/barriers: A Multidecadal and Broad Geographical Review. *Ecology of Freshwater Fish*, *19*(4), 487–509. https://doi.org/10.1111/j.1600-0633.2010.00435.x

Oberdorff, T., Tedesco, P. A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S., & Dürr, H. H. (2011). Global and regional patterns in riverine fish species richness: A Review. *International Journal of Ecology*, *2011*, 1–12. https://doi.org/10.1155/2011/967631

OpenStreetMap. (n.d.). Retrieved March 28, 2022, from https://www.openstreetmap.org/#map=19/49.03163/19.24420

Phillips, R. D., Storey, A. W., & Johnson, M. S. (2009). Genetic structure of *Melanotaenia australisat* local and regional scales in the East Kimberley, Western Australia. *Journal of Fish Biology*, *74*(2), 437–451. https://doi.org/10.1111/j.1095-8649.2008.02099.x

Plumb, G. A. (1993). A scale for comparing the visual magnitude of waterfalls. *Earth-Science Reviews*, *34*(4), 261–270. https://doi.org/10.1016/0012-8252(93)90059-g

Rahel, F. J. (2007). Biogeographic barriers, connectivity and homogenization of freshwater faunas: It's a Small World after all. *Freshwater Biology*, *52*(4), 696–710. https://doi.org/10.1111/j.1365-2427.2006.01708.x

Ramsay, P.M. (ed.) (2001) The Ecology of Volcán Chiles: high-altitude ecosystems on the Ecuador-Colombia border. Plymouth: Pebble & Shell. Pp. 193-199.

Ricklefs, R. E. (1987). Community diversity: Relative roles of local and regional processes. *Science*, *235*(4785), 167–171. https://doi.org/10.1126/science.235.4785.167

Roberts, T.R., & Baird, I. G. (1995). Traditional Fisheries and Fish Ecology on the Mekong River at the Khone Waterfalls in Southern Laos. *Natural History Bulletin of the Siam Society*. *43*, 219-262.

Roberts, T. & Stewart, D. (1976). An ecological and systematic survey of fishes in the rapids of the Zaire or Congo River. *Bulletin of the Museum of Comparative Zoology*, 147, 239-317

Sant'Anna, I. R., Doria, C. R., & Freitas, C. E. (2014). Pre-impoundment stock assessment of two Pimelodidae species caught by small-scale fisheries in the Madeira River (Amazon Basin - Brazil). *Fisheries Management and Ecology*, *21*(4), 322–329. https://doi.org/10.1111/fme.12082

Scheingross, J. S., & Lamb, M. P. (2016). Sediment transport through self-adjusting, bedrock-walled waterfall plunge pools. *Journal of Geophysical Research: Earth Surface*, *121*(5), 939–963. https://doi.org/10.1002/2015jf003620

Smethurst, D. (2000). Mountain geography. *Geographical Review*, *90*(1), 35. https://doi.org/10.2307/216174

The IUCN Red List of Threatened Species. IUCN Red List of Threatened Species. (n.d.). Retrieved December 2, 2021, from https://www.iucnredlist.org/resources/spatial-data-download.

Thé, A. P. G., & Nordi, N. (2006). Common Property Resource System in a Fishery of the São Francisco River, Minas Gerais, Brazil. *Human Ecology Review*, *13*(1), 1–10. http://www.jstor.org/stable/24707553

Udvardy, M. D.F. (1975). A classification of the biogeographical provinces of the world. *IUCN Occasional Paper*. International Union for Conservation of Nature and natural resources (IUCN) (18)

Uhunmwangho, R., & Okedu, E. K. (2009). Macro-and Micro-Hydropower: An Option for Socioeconomic Development. Case Study–Agbokim Waterfalls, Cross River State, Nigeria. *Pac J Sci Technol*, *10*(2), 529-34.

Usdowski E., Michaelis J., Böttcher M. E., and Hoefs J. (1991) Factors for the oxygen isotope equilibrium fractionation between aqueous and gaseous CO2, carbonic acid, bicarbonate, carbonate and water (19°C). *Z. Phys. Chem. 170*, 237–249.

Wazed, M. A., & Ahmed, S. (2009). A feasibility study of MICOR-hydroelectric power generation at Sapchari Waterfall, Khagrachari, Bangladesh. *Journal of Applied Sciences*, *9*(2), 372–376. https://doi.org/10.3923/jas.2009.372.376

Webmap.cn - 全国地理信息资源目录服务系统 [National Geographic Information Resource Catalog of China]. (n.d.). Retrieved March 28, 2022, from https://www.webmap.cn/

World waterfall database. World Waterfall Database. (2018). Retrieved December 13, 2021, from https://www.worldwaterfalldatabase.com

Zhang, D. D., Peart, M., Zhang, Y. J., Zhu, A., & Cheng, X. (2000). Natural water softening processes by waterfall effects in karst areas. *Desalination*, *129*(3), 247–259. https://doi.org/10.1016/s0011-9164(00)00065-5

Zheng, H., Clift, P. D., Wang, P., Tada, R., Jia, J., He, M., & Jourdan, F. (2013). Pre-miocene birth of the Yangtze River. *Proceedings of the National Academy of Sciences*, *110*(19), 7556–7561. https://doi.org/10.1073/pnas.1216241110

Zilihona, I.J., Nummelin, M. (2001) Coleopteran diversity and abundance in different habitats near Kihansi waterfall, in the Udzungwa Mountains, Tanzania. *Biodiversity and Conservation*, *10*, 769–777. https://doi.org/10.1023/A:1016623119865