An improved geography of surface water abundance in lakes and reservoirs

by

Meng Ding

B.S., Harbin Normal University, 2016

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF ARTS

Department of Geography
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Dr. Jida Wang

Copyright

© Meng Ding and Jida Wang, 2019.

Abstract

The reserves of fresh water on the earth are of critical environmental and social importance. The great mass of liquid freshwater, which is readily accessible to human beings, resides in major surface stocks such as lakes, ponds, and reservoirs. Unfortunately, because of the extensive distribution and dynamic nature of global surface water bodies, our knowledge of the accurate extent, quantity, and quality of global lake systems remains surprisingly poor. River flow and storage in lakes/reservoirs are two main surface water resources for human societies. When river flow experiences high stress due to increasing human water demand, water storage in lakes and reservoirs can provide fresh water to mitigate the stress of river flow. Therefore, how river flow and lake water storage are distributed across the continental surface and what is the balance between them are crucially important.

The overarching objective of this thesis is to improve the understanding of the physical abundance of surface water in lakes and reservoirs. This goal was achieved by answering four specific questions: 1. Distribution: Where are freshwater lakes, and where are saline lakes? 2. Abundance: How many freshwater lakes and saline lakes populate across the Earth's surface, and how much area do they cover? 3. Storage: What are the total water storages in freshwater and saline lakes? 4. Storage versus flows: What is the spatial contrast between surface river flow and lake storage?

The question related to the distribution and the abundance of global freshwater and saline lakes was answered by integrating hydrological, spectral, climatological, and literature evidence on various geographic settings. I systematically classified global lakes and reservoirs into freshwater and saline categories from a high-resolution circa-2000 lake dataset that documents perennial water bodies greater than 0.4 hectares (Sheng et al., 2015; Sheng et al., 2016). In order to understand the storage of global lakes/reservoirs, I then estimated the mean depth, volume, and associated uncertainties for each lake/reservoir using geo-statistics, GIS-based terrain analysis, and existing literature. To understand the change of the distribution and the abundance of global lakes during recent 15 years, attributes of each inventoried lake were updated on the recently-produced circa-2015 dataset (Sheng et al., 2015; Sheng et al., 2016) based on the results of freshwater/saline classification and storage estimation for circa 2000. To explore the spatial contrast between lake

storage and river flow, which are two major surface water recourses for human society, the lake storage of circa-2000 and the average annual discharge during 1990–2010 simulated by the PCRaster Global Water Balance (PCR-GLOBWB) hydrological model were compared on the spatial scales of both individual drainage basins and 0.1° geographic grid.

This research improves our understanding of water abundance in global lakes and reservoirs by providing their freshwater/saline classification, number and storage distributions, and the comparison of their storages with river flow. Specifically, results for the first three questions reveal two unprecedentedly detailed inventories of global lake extents and storage with freshwater and saline classification. Our classified global lake inventories conclude 8–9 million lakes, with a total area of 3,002 thousand km^2 and a total volume of 189 (±6) thousand gigatons (Gt) in the circa-2000 inventory, and a total area of 2,976 thousand km^2 and a total volume of 190 (±6) thousand Gt in the circa-2015 inventory. Among them, saline lakes account for less than 2% in number, about a quarter in area and half in volume. The result for the fourth question shows the spatial disparity among surface water resources. Two-scale datasets that contain freshwater lake storage, saline lake storage, and average annual river discharge (1990–2010) were generated to improve our understanding of the spatial distribution among the three factors in different regions. These spatially-explicit contrasts show that high freshwater storage tends to coexist with high flow, except high-latitude (pan-Arctic) regions where lake distributions are extensive and thus water storage largely exceeds the flow. Saline water storage is spatially less or negatively correlated to freshwater storage or discharge, thus implicating arid/semiarid climates and water-stressed regions. In general, river flow and freshwater lakes are the main water sources in humid regions, but saline lakes are dominant in arid/semi-arid regions. The detailed spatial cross-tabulations between surface river flow and storage provide an important guidance for potential water resource management.

Table of Contents

List of Figures viii

List of Tables x

Acknowledgments xi

Chapter 1 - Introduction 1

1.1 Backgrounds 1

1.1.1 Water scarcity 1

1.1.2 Freshwater resources 1

1.1.3 Surface water bodies 2

1.1.4 Surface river flows and storage 3

1.2 Statement of the Problem 3

1.3 Research Objectives 4

Chapter 2 - Literature review 6

2.1 Global waterbody datasets 6

2.2 Lake volume estimation 11

2.3 Surface river flow 12

Chapter 3 - Methodology 14

3.1 Dataset acquisition 14

3.2 Methods 15

3.2.1 Classifying freshwater and saline lakes 16

3.2.2 Estimating lake water storages 25

3.2.2.1 Reservoirs 26

3.2.2.2 Large natural lakes 28
3.2.2.3 Natural lakes of intermediate sizes 29
3.2.2.4 Lakes smaller than 0.1 km^2 30
3.2.3 Assigning attributes to the updated circa-2015 dataset 30
3.2.4 Lake water storage estimation on the circa-2015 mapping 34
3.2.5 Spatial contrast between surface flow and storage 34
Chapter 4 - Results 36
4.1 Global lake classification (freshwater and saline) 36
4.2 Global lake and reservoir volumes 45
4.3 Comparison between water flow and lake water storage 53
Chapter 5 - Discussion 65
5.1 Comparison with existing studies 65
5.2 Dataset limitations 68
5.2.1 Freshwater/saline classification 68
5.2.2 Lake volume estimations 69
5.3 Future research 70
5.4 Broader impacts 71
Chapter 6 - Conclusions 72
6.1 Assessment of water abundance of global lakes and reservoirs 72
6.1.1 Distribution of freshwater and saline lakes 72
6.1.2 Abundance of freshwater and saline lakes 72
6.1.3 Storages of freshwater and saline lakes 73
6.2 Comparison between water flow and lake water storage 73

References 75

Appendix A - Summary of Codes 82

List of Figures

Figure 2-1 Accuracy comparison between HydroLAKES and the circa-2000 inventory. 10

Figure 2-2 Average annual river discharge during 1990 to 2010. 13

Figure 3-1 Roadmap of the methods applied in this thesis. 16

Figure 3-2 Logic flow of global lake classification (freshwater and saline). 18

Figure 3-3 Spatial distributions of endorheic basins and arid regions. 21

Figure 3-4 Lake Assal used for spectral evidence interpretation. 22

Figure 3-5 One example of a newly constructed reservoir in the circa-2015 lake inventory. 31

Figure 3-6 Spatial distribution of Lake Chad and the Aral Sea on the circa-2000 and the circa-2015 inventories. 31

Figure 3-7 Steps for updating saline/freshwater types of global lakes from the circa-2000 to the circa-2015 inventories. 33

Figure 3-8 HydroSHEDS basins (15 arc second version) with the Yangtze River Basin and Mississippi River Basin highlighted. 35

Figure 4-1 Global spatial distribution of freshwater and saline lakes on the circa-2000 mapping. 38

Figure 4-2 Global spatial distribution of freshwater and saline lakes on the circa-2015 mapping. 39

Figure 4-3 Quantities of freshwater and saline lakes on the circa-2000 and the circa-2015 inventories. 40

Figure 4-4 Relationships between lake area and abundance (count) in each continent. 44

Figure 4-5 Area/Volume percentages for saline and freshwater lakes on two mappings. 47

Figure 4-6 Relationships between lake volume and abundance (count) in each continent. 49

Figure 4-7 Relationships between global lake area/volume and abundance (count). 50

Figure 4-8 Spatial distributions of saline lake storage, freshwater lake storage and annual discharge on 0.1º grid. 57

Figure 4-9 Spatial distributions of saline lake storage, freshwater lake storage and average annual discharge in HydroSHEDS basins. 60

Figure 4-10 Latitudinal and longitudinal distributions of global freshwater lake storage, saline lake storage, and average annual river discharge (1990–2010). 62

Figure 4-11 RGB composite of saline lake storage, freshwater lake storage, and river discharge. 63

List of Tables

Table 2-1 Major global lake datasets. 9

Table 3-1 Detailed information (hydrological, spectral, and reference evidence) for each specific type in the classification of saline and freshwater lakes. 24

Table 3-2 Empirical models for estimating the mean depths of natural lakes larger than 0.1 km^2. 29

Table 3-3 The sequence of lake type (from less to more evidence) assigning procedure from the circa-2000 to the circa-2015 mappings. 33

Table 4-1 Quantities of specific lake types for classification accuracy evaluation. 42

Table 4-2 Comparison of estimated lake volumes (for lakes larger than 0.1 km^2) between circa-2000/2015 inventories and HydroLAKES. 45

Table 4-3 Comparison of estimated reservoir volumes between circa-2000/2015 inventories and GRanD. 45

Table 4-4 Power law distributions for lake abundance (count) under area and volume levels for each continent and the globe. 51

Table 4-5 Summary of global freshwater and saline lake area and volume on both circa-2000 and circa-2015 mappings. 53

Table 4-6 Saline/freshwater storage and discharge in six large river basins. 64

Table 5-1 Comparison between major global lake datasets and our circa-2000 and circa-2015 mappings. 67

Table A-1 List of codes for each analytical step. 82

Acknowledgments

I sincerely appreciate this great opportunity that gives me two years of valuable study time at Kansas State University. During this period, I felt tremendous friendliness and understanding from professors and all fellow students. I also appreciate the department of geography and Dr. Jida Wang for giving me the opportunity to be a Graduate Research Assistant and Graduate Teaching Assistant. As my thesis is gradually completed, my Master's program is coming to an end, but I am so honored to be studying here in the next several years for pursuing my Ph.D. degree.

First, I want to give my most thanks to my supervisor Dr. Jida Wang, who provided me plenty of help in learning, researching, and being a useful person in the society. In his classes and in communications with him, I obtained a large amount of knowledge that was useful for my course work and research. Without his guidance, I think I would still have had many problems and could not finish my thesis research. Moreover, he always offers me the greatest help at the first time when I have any difficulties in either study or life. In a word, I think I am so lucky to be his student and I will make continued efforts working with him together in the coming years.

Next, I really want to thank my committee members, Dr. Shawn Hutchinson and Dr. Abigail Langston. I have been benefited substantially from their classes and advice in my research. Their understanding and patience gave me a lot of help during the completion of my thesis.

I am very grateful to my team members, Takuto Urano, Pedro Satori, Blake Walter, and Colin Bailey, who made critical contributions to the data preparation for my thesis research at an earlier stage. I also would like to thank Dr. Yoshihide Wada from the International Institute for Applied Systems Analysis in Austria, who kindly shared the global river discharge simulations from the PCR-GLOBWB model.

At last, thank all my family members and friends. In the past three years, I as an international student overcame many challenges in language and culture when pursuing a degree in an unfamiliar learning environment. With the constant connection to my family and friends in China, I got their encouragements and supports, and have matured in both academics and life.

I will make more contributions in my field to express my thanks to all of you.

Chapter 1 - Introduction

1.1 Backgrounds

1.1.1 Water scarcity

Water scarcity is considered to be one of the major global risks in contemporary human societies. The great threat of water scarcity leads to many social and environmental problems. Accessibility to freshwater resources is affected by several factors at regional or global scales, such as climate change/variability, land use/cover changes, and the increasing population. Water scarcity in some parts of the world with decreasing runoff is being influenced by climate change, including places around the Mediterranean, central and southern America, southern and eastern Asia and southern Africa (Arnell, 2004; Scheffran and Battaglini, 2011; Giorgi, 2006). Furthermore, the recent rapid population growth is also an essential factor that leads to increasing water scarcity. Wada et al. (2011a) estimated that in the recent 40 years, the population affected by water scarcity has increased by ~20% (Veldkamp et al., 2015; Wada et al., 2011a). About two billion people are currently experiencing water scarcity due to the uneven distribution of renewable freshwater resources in time and place (Oki and Kanae, 2006; Vörösmarty et al., 2000). Jackson et al. (2001) noted that in the next 30 years, the Earth's population will likely increase by approximately one-third, but the accessible runoff may only increase by 10% or less. Therefore, it can be predicted that there is a growing population that will be facing water scarcity in the next half century. Consequently, the problem of potential water resources management has received widespread concerns.

1.1.2 Freshwater resources

Fresh water is one of the most vitally important natural resources for all organisms' survival (Oki and Kanae, 2006; Schewe et al., 2014). There are more than one billion cubic kilometers of water on the Earth (Bernhardt and Schlesinger, 2013; Chahine, 1992; Jackson et al., 2001; Schmitt, 1995), but the storage of fresh water only accounts for about 3% of the water mass in the hydrosphere. Furthermore, approximately 83% of the total freshwater resources is stored in ice sheets and glaciers (Kashid and Pardeshi, 2014). Therefore, the quantity of liquid fresh water that is directly accessible to people is extremely scarce, only ~0.5% of the global water storage.

The continuous population growth will increase the demand of freshwater resources. According to the recent news published by the Department of Economic and Social Affairs of the United Nations, it is predicted that the population will reach 11.2 billion by 2100 (Source: https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html). Due to the rapid population growth, economic development and changes in lifestyle, human freshwater withdrawals for industrial, agricultural, and domestic purposes are increasing dramatically (Haddeland et al., 2014; Vörösmarty et al., 2000). Among them, irrigation accounts for ~70% of the global water withdrawal and nearly 90% of the water consumptive loss (i.e., the amount of net water withdrawal not returned to the drainage network) (Rost et al., 2008; Vörösmarty et al., 2000). Therefore, in order to utilize water storage for the irrigation purpose, which accounts for the largest proportion of freshwater consumption, the maximal salinity of irrigable water should be considered as the threshold for classifying freshwater and saline lakes.

In addition to the rapid population growth, people affect water resources with other behaviors. Because different crops and vegetation are found in various seasons, irrigation water consumption has great seasonal variability (Wada et al., 2011b). Additionally, the construction of dams and reservoirs has altered the seasonal distribution of freshwater resources (Davie et al., 2013).

1.1.3 Surface water bodies

There are several well-known components of freshwater resources for human water consumption, including river flow (also termed as streamflow, discharge, or runoff), groundwater, and desalinated water, but the most readily accessible liquid fresh water resides in major surface stocks such as lakes, ponds, and reservoirs. River flow and lake storage are two main components of the surface water system. River flow includes discharge in the river/stream channels accumulated from surface runoff and base flow (Stricker, 1983), and lake storage contains water in two major types of lakes: freshwater lakes and saline lakes. Water in many saline lakes cannot be directly used for human consumption. Consequently, the locations and storages of global freshwater lakes are of immense importance for further understanding the freshwater availability for human society. However, because of the extensive distribution and dynamic nature in the global lake systems (Adrian et al., 2009; Pekel et al., 2016; Smith et al., 2005; Song et al., 2016;

Wang et al., 2012; Wang et al., 2014; Wang et al., 2017; Yamazaki et al., 2015; Yao et al., 2019), our knowledge of their accurate extents, quantity, and quality remains incomplete (Messager et al., 2016; Cael et al., 2017). In order to better understand the global lake water extents, two detailed and quality-assured global lake inventories (circa-2000 and circa-2015) have been produced by Sheng et al. (2015) and Sheng et al. (2016) using archival Landsat images at 30-meter resolution.

1.1.4 Surface river flows and storage

Many recent studies (e.g., Herschy, 2014; Lin et al., 2019; Milly et al., 2005; Nijssen et al., 1997) focused on measuring and simulating river flow on a global scale. Those research can provide the estimation of river flow to help find out the influences of climate change and human activities on river flow (Devia et al., 2015). Nevertheless, both river flow and lake water storage can contribute to domestic, industrial, and agricultural uses. In other words, when one of the two sources (i.e., surface flow and storage) is scarce, the other may supplement additional fresh water. Water flow in streams and rivers is considered to be the main renewable/sustainable water resources (Oki and Kanae, 2006), but rivers may experience severe water stress in some cases, such as arid regions or regions with high population densities. Therefore, lake water storage can be regarded as an important component in mitigating river water stress. How river flow and lake water storage are distributed in different places and what is the spatial contrast between them are thus worth additional research.

1.2 Statement of the Problem

Freshwater resources are finite and vital for human activities and food production (Wada et al., 2011b). Currently, climate change and human water managements have generated critical threats to global freshwater resources and are predicted to bring more negative impacts on water resources in the future, such as earlier snow/glacier melt seasons, reduced flow, and increased consumptive loss (Wada et al., 2016; Barnett et al., 2005). Freshwater lakes are an important constituent of the global water cycle and are one of the most commonly accessible water resources to human societies (Pekel et al., 2016). However, current studies on the physical abundance of lake water storage at the global scale are surprisingly inadequate. Existing studies have estimated the

volumes and depths of global lakes without a thorough distinction between the freshwater and saline types. Therefore, the total volume of freshwater lakes and its spatial variation are is still poorly known. In order to better understand the abundance of surface freshwater storage in global lakes and reservoirs, a spatially-explicit differentiation among different waterbody types should be first given. Therefore, classifying global lakes into freshwater and saline types is a prerequisite before estimating the abundance of surface freshwater storages/volumes.

Human surface water withdrawals stem from both water flow (such as in rivers and streams) and water storage (such as in freshwater lakes and groundwater), but the abundance of both surface river flow and storage varies across space. As previously described, river flow is usually considered as a renewable water resource, whereas lake storage can provide water resources when river flow experiences severe water stress (Wang et al., 2010). Lake water storage, especially freshwater lake storage, is a critical component of the hydrological cycle, because it can be easily assessed and used by human. Comparing lake water storage (freshwater storage and saline storage) and river flow within a specific unit area can clearly reveal which of them is more abundant and how they can potentially supplement each other in the provision and management of water resources.

1.3 Research Objectives

The overarching objective of my thesis is to improve the understanding of the physical abundance of surface water in lakes and reservoirs. Based on the background introduction and the previous explanation on the research problems, I proposed the following four specific questions. The detailed solutions and results for each question will be presented in the chapters of methodology and results, respectively.

(1) Distribution: Where are freshwater lakes, and where are saline lakes?

(2) Abundance: How many freshwater lakes and saline lakes populate across the Earth's surface, and how much area do they cover?

(3) Storage: What are the total water storages in freshwater and saline lakes?

(4) Storage vs. flows: What is the spatial contrast between surface river flow and lake storage?

Chapter 2 - Literature review

This literature review is organized into three subsections, including the summary of the advantages and limitations of the existing datasets about global waterbodies, the methods for global lake storage estimation, and the availability of river discharge data which will be needed for comparing river flow and lake water storage.

2.1 Global waterbody datasets

The development of a precise and high-resolution global waterbody inventory has been one of the critical issues that hydrologists are working on in the recent years. Fortunately, due to the advances of remote sensing and geographic information systems on measuring global-scale terrestrial environments, several different datasets on global waterbodies have been produced successively. For example, the Mullard Space Science Laboratory (MSSL) Global Lakes Database (MGLD) contains the information of locations and types for 1415 inland natural and artificial waterbodies based on three primary categories, including closed lakes, open lakes, and reservoirs. (Birkett and Mason, 1995). The Shuttle Radar Topography Mission (SRTM) Water Body Data (SWBD) (SWBD, 2005; Slater et al., 2006) was developed from the SRTM Digital Elevation Model (DEM) and constitutes waterbodies with a total area of 2.76 million km^2. However, the data for the mapped waterbodies were only collected from 56°S to 60°N, limited by the spatial coverage of the SRTM. In order to fill in this incomplete coverage, Carroll et al. (2009) created a comprehensive global water mask at 250-m resolution by combining SWBD and MODIS imagery.

Several major and recent waterbody datasets with detailed information (including spatial resolution, lake/waterbody number, area, and total volume) are compared in **Table 2-1** (note that the dataset list may not be exhaustive). The GLObal WAter BOdies (GLOWABO) dataset with a resolution of 0.002 km^2 or 0.2 hectares (**Table 2-1**) was generated by using 30-m-resolution Landsat ETM+ imagery (Verpoorter et al., 2014). Although this dataset has the finest spatial resolution in comparison with others in the table, ~90 million lakes out of a total number of 117 million are in the smallest size level (0.002 to 0.01 km^2). However, the total area of these small lakes only accounts for 0.27% of the non-glaciated land surfaces. The volumes of these small lakes were estimated rather than measured, so the estimated results may have high inaccuracies due to

the great number of small lakes. Other datasets have relatively lower spatial resolutions. For example, 1.5 million waterbodies larger than 0.1 km^2 were mapped in the Global Lakes and Wetlands Dataset (GLWD) (Lehner and Döll, 2004). Waterbodies in GLWD were classified into several types including wetlands, lakes, rivers, swamps, and reservoirs. The Global Inland surface Waterbody dataset (GIW) (Feng et al., 2016) mapped global waterbodies from 30-m resolution Landsat images (pixel size ~0.001 km^2), covering a total area of ~3.65 million km^2. The limitations of this inventory are that some of the included waterbodies are ephemeral or intermittent, and some waterbodies were not mapped completely because of cloud contamination. Moreover, there is no explicit distinction between rivers and lakes/reservoirs. One of the major breakthroughs of the recently-published HydroLAKES dataset (Messager et al., 2016) is the estimation of global lake water storage. In this dataset, Messager et al. (2016) integrated a large number of documentations and novel geostatistical models to estimate the depth and volume for each inventoried lake/reservoir larger than 0.1 km^2 (refer to Section 2.2 for more details). Due to the comprehensiveness of HydroLAKES dataset and the differentiation between natural lakes and reservoirs, the result of the storage estimation may have higher accuracy than other previous estimations. One major limitation of HydroLAKES is the relatively low resolution so that a great number of lakes smaller than 0.1 km^2 were not mapped. These datasets will be also compared to the products of my research in the discussion chapter.

Although these global waterbody datasets with different focuses have been released, none of them explicitly separated freshwater lakes from those that are saline. Among these existing datasets, Sheng et al. (2015) and Sheng et al. (2016) have produced two high-resolution global waterbody masks from archival Landsat images. Both of them will be applied as the major datasets in this thesis. These two datasets present the accurate lake locations and extents in both two periods (circa 2000 and circa 2015). Both of them contain millions of perennial waterbodies larger than 0.004 km^2 or 0.4 hectares (equivalent to about 4 Landsat TM/ETM+/OLI pixels), and improved our understanding of the spatial distribution of global lakes and reservoirs. Unfortunately, both circa-2000 and circa-2015 inventories do not have any information about water quality and lake storage, such as the classification of freshwater and saline lakes. Abbott et al. (2019) implicated that due to the absence of the classification of freshwater and saline lakes, the freshwater available for human use is overrepresented in some water diagrams. Therefore, clearly understanding the

locations and the storages of freshwater and saline lakes has profound implications for water management and freshwater allocation.

Table 2-1 Major global lake datasets.

Dataset	*Spatial resolution*	*Total number (million)*	*Total area (million km^2)*	*Classification of freshwater and saline lakes*	*Total volume (thousand km^3)*	*Reference*
Global Lakes and Wetlands Dataset (GLWD)	0.1 km^2	1.5	2.7	No	---	Lehner and Döll (2004)
Merge of SRTM Water Body Dataset v2.0 (SWBD 2005) and GTOPO30 Lakes	~0.1 km^2	0.70	2.76	No	---	Slater et al. (2006)
GlObal WAter BOdies Database (GLOWABO)	0.002 km^2	117	5.0	No	199 (±3) (for > 0.01 km^2)	Cael et al. (2017); Verpoorter et al. (2014)
Global Inland Surface Waterbody Dataset (GIW)	~0.001 km^2	---	3.65	No	---	Feng et al. (2016)
HydroLAKES	0.1 km^2	1.42	2.67	No	182.9	Messager et al. (2016)

Note: The dataset list may not be exhaustive, and some numbers are approximations.

As previously mentioned, the recently published HydroLAKES has a relatively coarse resolution but includes the first global-scale lake volume estimation. In order to compare the mapping accuracy between HydroLAKES and the circa-2000 dataset (Sheng et al. 2016), the mapping result of small lakes should be considered. Due to the scraping action of glaciers and permafrost processes, Canada has the highest lake density in over the world, so a location was randomly selected (**Figure 2-1**) from central Canada (about 64°35'N and 116°37'W) as an example to show the limitation of HydroLAKES. Apparently, there are many small lakes that were omitted in HydroLAKES but exist in the circa-2000 lake mapping. Therefore, the circa-2000 and circa-2015 mappings are determined to be used as the primary global lake datasets for this thesis.

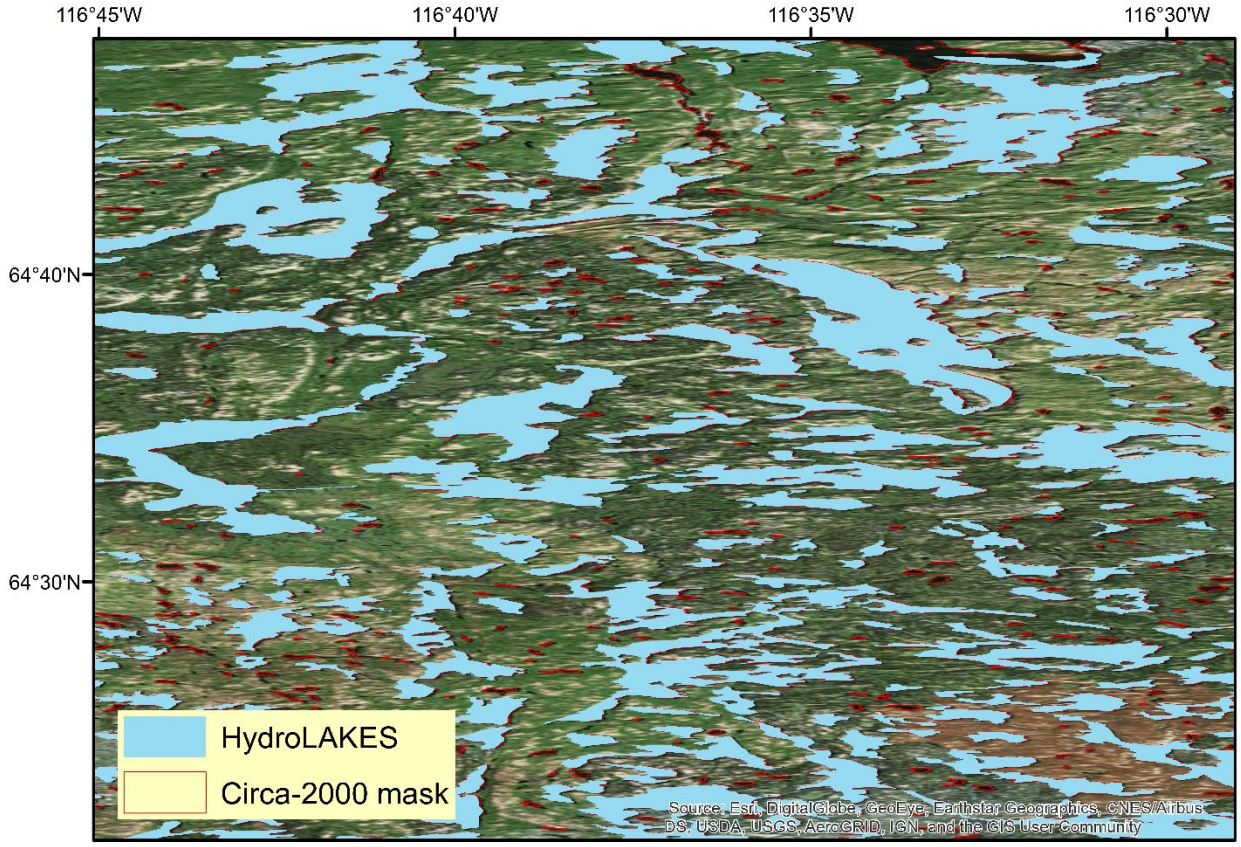


Figure 2-1 Accuracy comparison between HydroLAKES and the circa-2000 inventory. The selected location is a small part of Canada. Red (hollow) and blue (filled) polygons depict waterbody extents as from the circa-2000 global lake dataset and HydroLAKES, respectively.

2.2 Lake volume estimation

Estimating lake/reservoir depths and volumes is crucial to understanding the physical abundance of surface water storage. Most studies on the estimation of lake water storages focused on regional scales (Hollister and Milstead, 2010; Kameyama et al., 2004; Sima and Tajrishy, 2013). Some of them applied the lake area and the topographic characteristics in the vicinity of the lake to predict the lake depth and the volume, and some of them measured the lake depth and lake volume in a variety of ways. However, detailed bathymetric measurements for individual lakes on a global scale is practically prohibited. Therefore, it is necessary to develop an efficient model for lake storage estimation. Two recent studies, Cael et al. (2017) and Messager et al. (2016), made important progress in estimating global lake volumes through novel approaches. The results of these two articles reveal the scarcity of global terrestrial water storage.

Specifically, Cael et al. (2017) assumed the statistical relationship between vertical and horizontal scales on self-affine surfaces is characterized by the Hurst exponent H (Dodds and Rothman, 2000). Cael et al. (2017) developed a statistical volume-area function based on the Hurst coefficient of the Earth's surface to estimate the mean depth and volume of global lakes. The estimated mean depth is 41.8 m and the global storage is ~199 thousand km^3. The applied global lake dataset is GLOWABO (Verpoorter et al., 2014) which mapped 117 million lakes larger than 0.002 km^2 (see details in **Table 2-1**). Based on the previous comparison between GLOWABO and other existing global waterbody datasets, the result of global volume estimation in Cael et al. (2017) highlighted the influence of small lakes that were neglected in other mappings. As one of the limitations, their method was based on a mathematical scaling between only lake area and depth. The derived function may be used to infer the aggregate lake volume on a collective scale, but likely unsuitable for determining the volume of any individual lake.

Alternatively, Messager et al. (2016) estimated the mean depths and volumes of the 1.42 million global lakes and reservoirs (larger than 0.1 km^2) in HydroLAKES (**Table 2-1**), by considering the variability of not only lake areas but also the terrain in the vicinity of the lakes. They also estimated the mean and median hydraulic residence time for each inventoried lake by using the discharge through the lake simulated from a global high-resolution hydrological model. The result shows that the total volume of global lakes in HydroLAKES is 187.9 million km^3, which is enough to inundate the world's landmass by a depth of 1.26 m. To achieve this result, Messager

et al. (2016) classified global lakes (1.42 million in total) into three types. The volume information of 266 large natural lakes (Type 1) was directly acquired from existing literature and documentations, and the volumes for 6,797 major reservoirs (Type 2) was obtained from the Global Reservoir and Dam (GRanD) database (Lehner et al., 2011). The volume of each of the remaining lakes larger than 0.1 km^2 (Type 3) was determined empirically by a series of geostatistical models that considers both lake size and the surrounding topography. Among them, model 5 (see Section 3.2.2.3 for details) was selected as the best one for volume calculation. This model consists of four equations, each fitted to estimate volumes for the lakes within a specific size range (i.e., 0.1–1 km^2, 1–10 km^2, 10–100 km^2, and 100–500 km^2, respectively). However, the limitation of this inventory is that it only includes the waterbodies larger than 0.1 km^2, but the circa-2000 and circa-2015 inventories mapped another ~6 million lakes smaller than 0.1 km^2 (particularly in climate-sensitive Arctic regions).

The two studies described above are remarkable progress in the field of hydrology because of their novel methods for lake storage estimation. Nevertheless, neither of the two studies made an explicit separation between freshwater and saline lakes. Without such a separation, we are still unclear of how much fresh (or saline) water the global lakes and reservoirs store.

2.3 Surface river flow

River discharges measured in situ are often unavailable because operational gauging stations are limited in time and space, and the measurements are not always internationally shared. Therefore, global hydrological models have been commonly applied to understanding and assessing global river flows. Over the last few decades, many different hydrological models have been developed for simulating river flows. They include three general types: empirical models, conceptual methods, and physically-based models (Devia et al., 2015). Each global hydrological model was simulated and developed for different objectives and applications. For example, Interactions between Soil, Biosphere, and Atmosphere – Total Runoff Integrating Pathways model (ISBA-TRIP) was produced in 2010 to measure the terrestrial water storage, and it is also applicable to measure the seasonal and annual change of the water storage (Sood and Smakhtin, 2015). Another existing river discharge dataset was simulated by using the PCRaster Global Water Balance (PCR-GLOBWB) hydrological and water resources model (Van et al., 2011; Wada et al.,

2011a, 2011b, 2014, 2016). This hydrological model was generated for multiple purposes, and has been applied in water stress assessment, estimation of future global runoff changes, and the analysis on the influence of human abstraction on surface water and groundwater (Sood and Smakhtin, 2015). Compared with other global hydrological models, PCR-GLOBWB model has advantages in spatial resolution (0.1º to 0.5º) and time resolution (daily), and it is suitable for runoff analysis due to its fine temporal resolution (Beek et al., 2011).

The existing river discharge datasets (0.1º × 0.1º) simulated by using the PCR-GLOBWB hydrological model contains a long-period daily river discharge from 1990 to 2010. The value in each grid cell calculates the aggregate water flow in any main streams and tributaries (surface runoff, baseflow, runoff from the interflow, and non-infiltrate melt water) located within this pixel on a daily level (Beek et al., 2011). **Figure 2-2** shows the annual river discharge (in the unit of cubic meters per second) that is averaged from two decades of data (1990–2010). This dataset will be helpful to compare the spatial distribution between river flow and lake water storage around circa 2000.

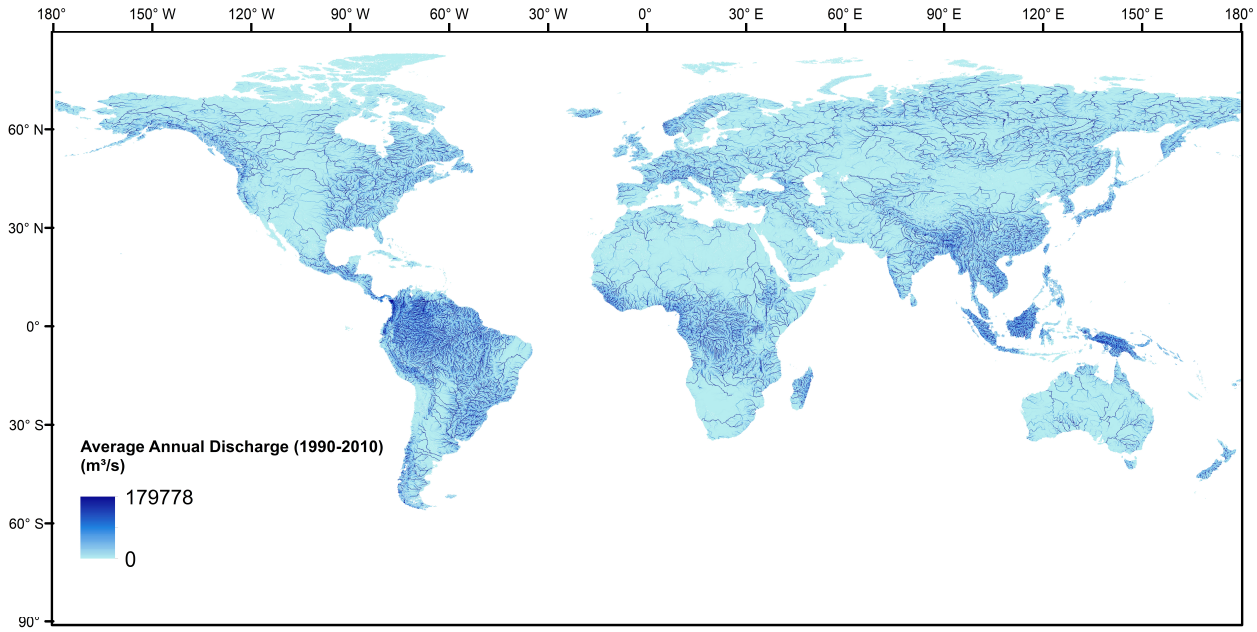


Figure 2-2 Average annual river discharge during 1990 to 2010.
Data was derived from daily river discharges (in 0.1º × 0.1º grid) simulated by the PCRaster Global Water Balance (PCR-GLOBWB) hydrological model.

Chapter 3 - Methodology

The contributions of my thesis have both scientific and technical values. For scientific contributions, my research aims to improve our understanding of the physical abundance of surface water storage in lakes and reservoirs. I achieved this goal by answering the four specific questions as described in the Introduction. For technical contributions, the comprehensive separation of global freshwater and saline lakes, particularly on the basis of a high-resolution global lake product, is to my knowledge the first of its kind. To enable this separation, I implemented a novel and integrated method that synergizes multiple evidence sources (hydrological, spectral, and literature/documentation). This technical contribution is a necessary and intermediate step to answer my overarching scientific questions.

In this chapter, the datasets and materials used are given in Section 3.1 dataset acquisition. Then in Section 3.2 methods, five subsections are provided to illustrate the detailed methodology for answering each of the research questions. To answer the first two questions, Section 3.2.1 explains how lakes in the global mappings were categorized into freshwater and saline types. To answer the third question, Section 3.2.2 articulates the methods of estimating lake/reservoir depths and volumes. Sections 3.2.3 and 3.2.4 describe the procedure of updating the lake attributes of the circa-2015 inventory, followed by Section 3.2.5 that describes the method of contrasting surface water flow and storage (to answer the fourth question).

3.1 Dataset acquisition

This thesis utilizes the following six datasets:

(1) Two recently completed global inventories, which document open water bodies larger than 0.4 hectares, mapped with quality assurance from thousands of circa-2000 and circa-2015 Landsat images by Sheng et al. (2015) and Sheng et al. (2016). They are hereafter referred to as circa-2000 and circa-2015 global lake inventories or mappings. The circa-2000 and circa-2015 global lake mappings contain ~8.08 million and ~9.09 million lakes, respectively.

(2) 15-arc-second HydroSHEDS dataset, which was developed from the SRTM DEM. It depicts the detailed hydrographic information, including river networks, drainage directions, and watershed boundaries (Lehner et al., 2008). The version I used here has a resolution of 15 arc seconds.

(3) HydroLAKES dataset, which contains the geometric extents, mean depths, and volumes of ~1.42 million global lakes and reservoirs larger than 0.1 km^2, generated by Messager et al. (2016) (also refer to Section 2.1).

(4) Global Reservoir and Dam (GRanD) database (Lehner et al., 2011), which inventories the geometric extents, areas, mean depths, storage capacities, and other attributes of 6,862 man-made reservoirs. The GRanD database was applied as a major reference for estimating reservoir volumes in the HydroLAKES database.

(5) Daily river discharges during 1990–2010 (at 0.1° × 0.1° resolution), simulated by the PCR-GLOBWB hydrological model (Van et al., 2011; Wada et al., 2011a, 2011b, 2014, 2016).

(6) A large number of other literature. For example, for the lake classification, the book *Saline Lake Ecosystems of the World* (Hammer, 1986) and many other references (e.g., Herdendorf, 1990) were used to retrieve or validate the locations, salinities, and alkalinities of many saline lakes worldwide. The volumes of large natural lakes were estimated from the documentation in the HydroLAKES dataset which was collected from a large number of references (see more details in Messager et al. (2016)).

3.2 Methods

The roadmap of the main methods and analytical steps is illustrated in **Figure 3-1**.

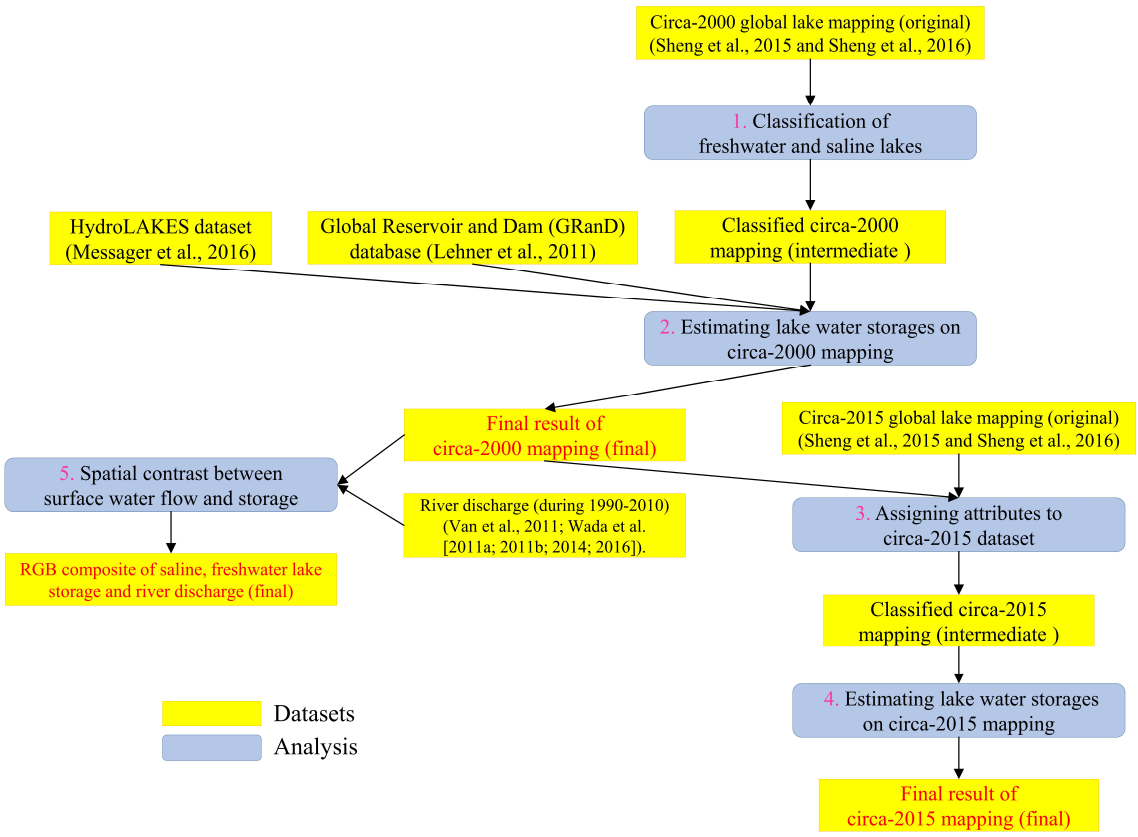


Figure 3-1 Roadmap of the methods applied in this thesis.
Yellow rectangles represent the original, intermediate and final datasets, while blue rectangles represent analytical steps (for Sections 3.2.1 to 3.2.5).

3.2.1 Classifying freshwater and saline lakes

Classifying global lakes into freshwater and saline categories on the Landsat-based circa-2000 inventory is a long-term group work due to a large number of global water bodies and the complex analytical methods (Ding et al. 2017; Wang et al., 2015, 2017). My main contributions were to complete the classification for the circa-2000 inventory based on the progress of other group members, verify the classification results with additional supporting literature, and then replicate the classification on the updated circa-2015 inventory. Four attributes or fields, including saline, salinity, alkalinity, and reference, were added to record the information of each classified lake. The field "saline" gives the classified lake type (freshwater or saline; **Table 3-1**). The fields "salinity" and "alkalinity" record the numeric values of salinity and alkalinity of the lake water, as retrieved from the literature. Numeric salinity or alkalinity values are not always available. If so, the supporting references are given in the "reference" field.

A fundamental way to differentiate freshwater and saline lakes is the salinity of the lake water. The threshold of salinity that separates them is, however, controversial. Various salinity thresholds have been applied according to different applications and purposes (Hammer, 1986). As described in the chapter of "Introduction", my goal of classifying global lakes into freshwater and saline types is to quantify the abundance of surface freshwater storage and thus help water management. In the "Introduction", I elaborated that irrigation water accounts for the largest proportion of all sectoral human water consumption. Therefore, I decided to choose a salinity threshold that determines whether the water in a specific lake can be potentially used for agricultural irrigation. According to the description of water use restriction in Misstear et al. (2006), there may be a severe irrigation problem if the water salinity exceeds 3.0 deciSiemens per meter (dS/m, a unit for electrical conductivity). Thus 3.0 dS/m was considered here as the maximal salinity threshold of irrigation water. This salinity (3.0 dS/m) is equivalent to a "total dissolved solids" (TDS) level of 2000 mg/l, or a salinity of 0.2%. This threshold was used consistently in the subsequent classification process. Accordingly, if the salinity recorded in the literature is more than 0.2%, the lake was considered to be a saline lake, and one lake with a salinity lower than 0.2% was categorized as a freshwater lake. These documented salinity values were also used to verify and assess my classification accuracy (refer to Section 4.1).

Due to the sheer number of global lakes and their diverse environmental settings, it is practically impossible to determine the salinity of each lake. Therefore, the methods with a comprehensive integration of hydrological analysis, spectral remote sensing, and existing references were applied to conduct the classification on both the circa-2000 and circa-2015 global lake datasets (see details in **Figure 3-2**).

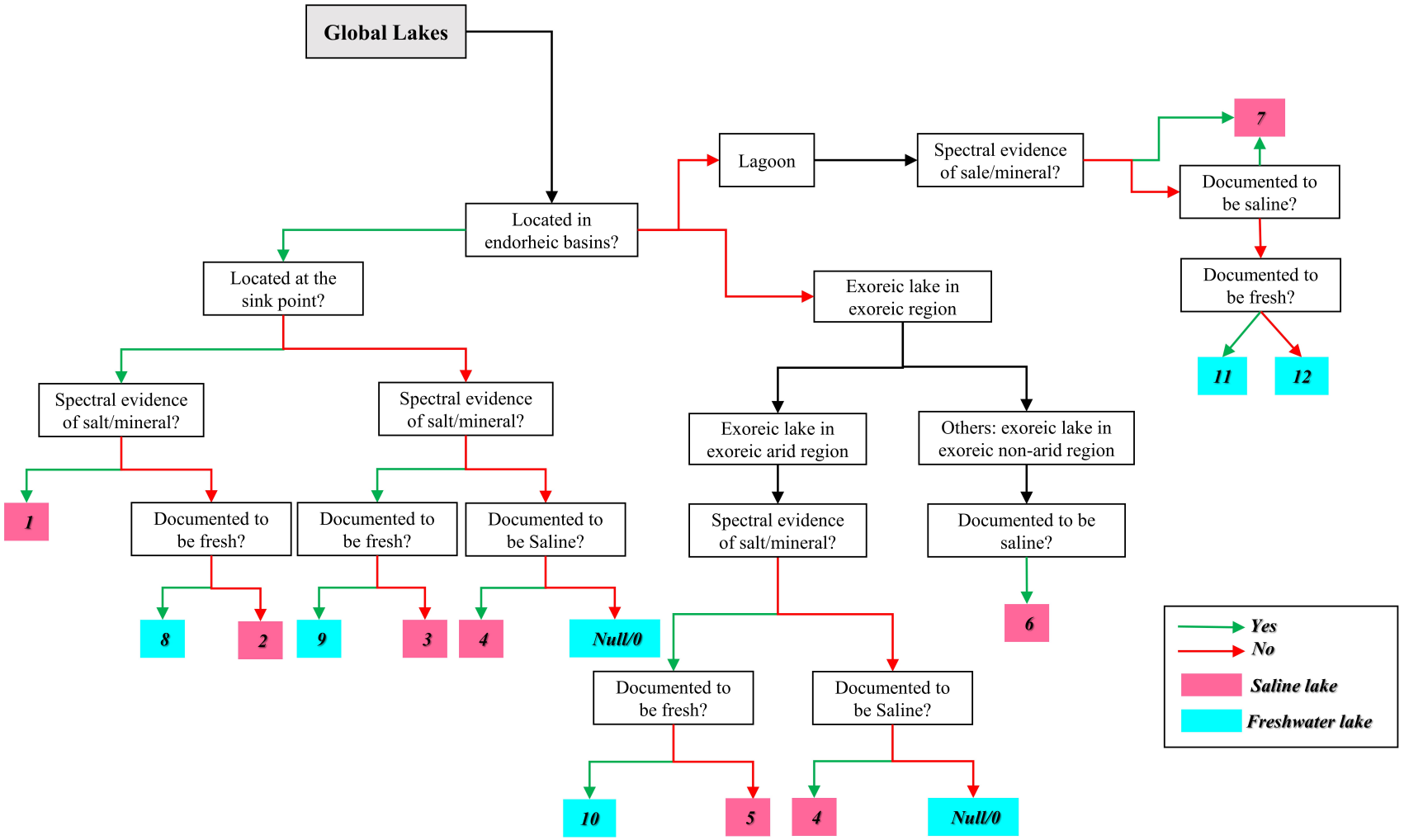


Figure 3-2 Logic flow of global lake classification (freshwater and saline).

Several definitions are helpful to better understand the classification method (**Figure 3-2**). They are highlighted in Italic in the following statements. First, the global continental surface is divided into *exorheic* and *endorheic* basins, depending on whether surface flow from the basin can drain to the ocean. Second, rivers in endorheic basins do not flow into the ocean, but instead, drain to a terminal point which is termed a *sink* or *endorheic point*. Accordingly, a lake that spatially concurs with a sink point is termed an *endorheic* or *terminal lake*. Note that not all lakes in endorheic basins are endorheic or terminal lakes. Third, an *arid region* is a geographic area where potential evaporation exceeds precipitation. Fourth, *coastal lagoons* are considered as shallow waterbodies that are separated from the ocean by barrier islands and/or reefs but remain influenced by the ocean through the inlets (Kjerfve et al., 1996).

The classification scheme illustrated in **Figure 3-2** considers the following aspects of hydrological, spectral evidence, literature evidence (Wang et al., 2015).

Hydrologically, most saline lakes occur in two kinds of places, endorheic basins, and arid regions (Hammer, 1986; Wang et al., 2018; Wurtsbaugh et al., 2017; Yang et al., 2017). In each endorheic basin, since the terminal lake (i.e., the lake at the sink point) has no surface outflow, water storage in this lake is mainly balanced out through evaporation. During the process of evaporation, dissolved minerals remain in the lake water and thus increases the lake salinity after long-term accumulation. Similar to terminal lakes, lake salinity in arid regions is often high due to salt accumulation as a result of excessive evaporation (in comparison to precipitation and surface inflow). In conclusion, lakes in endorheic basins and/or arid regions are candidates for saline lakes. In exorheic basins where rivers drain to the ocean, water budget is balanced out through both evaporation and surface outflow. As dissolved minerals may be largely taken away by surface outflow, the salinity of the lake water often remains low. Therefore, lakes in exorheic basins, except for those in the arid regions, are often found to be freshwater.

Different from terminal lakes, there are many other lakes in the endorheic basins, but not located at the sink points. For this kind of lakes, water can be lost through both evaporation and surface outflow (as those in exorheic basins). Therefore, non-terminal lakes in endorheic basins were not directly considered as saline lakes. In order to determine their types, I integrated the spectral and literature evidence. In order to identify the spectral evidence for each lake, I inspected the surface reflectance in the immediate vicinity of each lake from multi-source spectral images

(see the example in the next paragraph). Because endorheic basins and arid regions generally concur in space (**Figure 3-3**), the inspection for spectral evidence across endorheic basins also overlaps a large proportion of that for arid regions. In conclusion, each lake in endorheic-basins was classified by combing its spectral evidence, related references, and hydrological evidence.

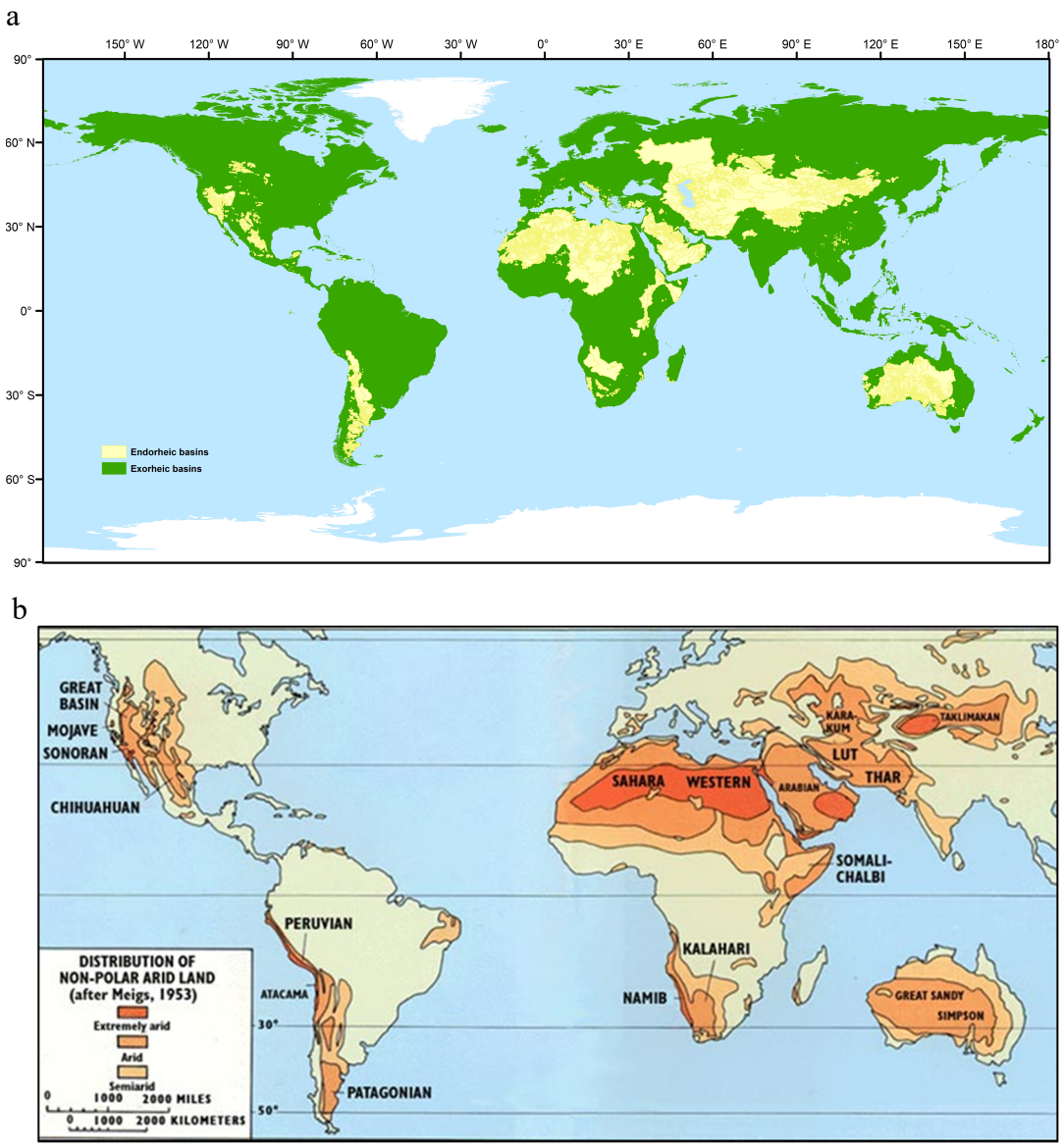


Figure 3-3 Spatial distributions of endorheic basins and arid regions.
Panel a: Global distribution of endorheic basins and exorheic basins. This map was derived from the HydroSHED 15-second Drainage Basins (Lehner et al., 2008). Panel b: Global spatial distribution of arid regions. (Map courtesy of United States Geographical Survey, accessible from http://pubs.usgs.gov/gip/deserts/what/world.html). As illustrated here, endorheic basins (yellow in panel a) generally concur with arid regions (orange in panel b) in space.

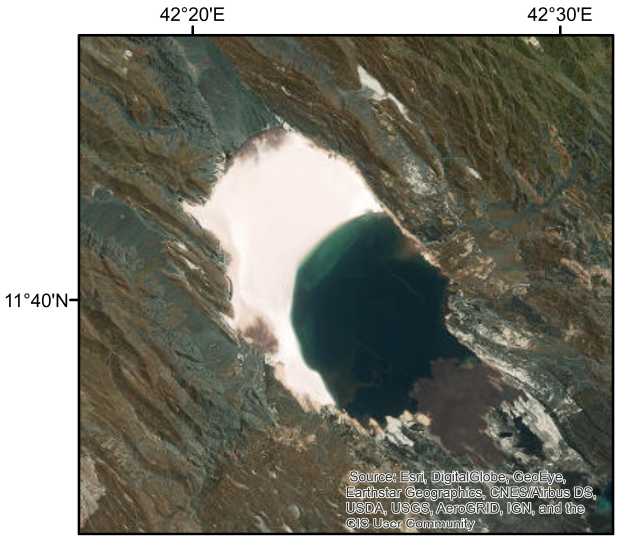


Figure 3-4 Lake Assal used for spectral evidence interpretation.
Lake Assal is located in the eastern part of Africa (11.66°N, 42.41°E). The background World Imagery Map source credits: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community.

Figure 3-4 is an example of how spectral evidence was used in determining the lake type. Lake Assal is one of the hypersaline lakes on the earth, with a salinity of 400 g/L (Brisou et al., 1974). From **Figure 3-4**, one can observe evident evaporates (i.e., deposited minerals through evaporation) in the lake periphery, manifested as high spectral reflectance in comparison with the low reflectance of the lake water. During the entire inspection process, I applied a variety of remote sensing image sources, including Landsat images (those for the circa-2000 lake mapping), Google Earth images, Bing Maps images, and the World Imagery Map provided by the ESRI.

In addition to the hydrological evidence and spectral evidence, a great amount of literature cited in this study helped me to determine the lake type (saline/freshwater) of many lakes. Most of them provided a substantial source of lake type, and some of them also recorded the salinities of several lakes. Based on the salinity criteria that I selected for irrigation water, the salinity within the reference can further confirm the lake is freshwater or saline.

The number at the end of each branch in **Figure 3-2** represents a particular pathway of how the lake type may be determined from the available evidence (see details in **Table 3-1**). As shown in **Figure 3-2**, global lakes were divided into five major types: lakes in endorheic basins and also located at the sink points, (2) lakes in endorheic basins but not at the sink points, (3) lakes in

exorheic basins and the arid region, (4) lakes in exorheic basins but not arid region, and (5) lagoons. Each lake was classified into a more specific type which will be assigned into "saline" field (**Table 3-1)**. This rule is helpful not only to understand the saline (represented by red) or freshwater (represented by blue) type of each lake, but also can illustrate more detailed information (e.g. location, spectral evidence, and reference) of them. **Table 3-1** contains the complete information for each specific type, and saline lakes are assigned as 1, 2, 3, 4, 5, 6, and 7 for the specific lake type and freshwater lakes are assigned Null, 8, 9, 10, 11 and 12. The type "Null" indicates the lakes that have none of the evidence (hydrological, spectral, or reference) to be saline, and therefore are considered as "freshwater". Other types were classified by combining all of three evidence sources. For example, type 1 represents the lakes that are located in endorheic basins and at the sink points, and there are also spectral evidence and literature evidence to further prove that they are saline. Another example is given for Types 2 and 8. Compared to type 2, type 8 has the same situations of hydrological and spectral evidence, but the reference records show those lakes (with type 8) as freshwater lakes. The same procedure as in **Table 3-1** was used to assign the results of lake classification on the circa-2000 to the circa-2015 mapping (see Section 3.2.3).

Table 3-1 Detailed information (hydrological, spectral, and reference evidence) for each specific type in the classification of saline and freshwater lakes.

Specific type	*Hydrological evidence*		*Spectral evidence*	*Reference evidence*	*Classification final result*
	Endorheic basin	*Endorheic lake*			
1	Yes	Yes	Yes	Yes	Saline
2	Yes	Yes	No	No	Saline
3	Yes	No	Yes	No	Saline
4	--	No	No	Yes	Saline
5	No	No	Yes	No	Saline
6	No	No	No	Yes	Saline
7	No	No	Yes	Yes	Saline
8	Yes	Yes	No	Yes	Freshwater
9	Yes	No	Yes	Yes	Freshwater
10	No	No	Yes	Yes	Freshwater
11	No	No	No	Yes	Freshwater
12	No	No	No	No	Freshwater
Null	--	No	No	No	Freshwater

Lake type or salinity documented in the existing literature can also help to assess the accuracy of the freshwater/saline lake classification based on hydrological and spectral evidence alone. For example, if one lake was classified to be saline based on the hydrological and/or spectral evidence, but its salinity in the reference is less than 0.2%, the classification was determined to be wrong. Therefore, the classification error was implicated by the number of misclassified lakes as a percentage of all lakes that have at least one type of evidence. Based on **Figure 3-2** and **Table 3-1**, the lakes other than "Null" type represent all the lakes for accuracy assessment, and types 4, 6, 8, 9, 10 are misclassified lakes because of their inconsistent classifications with the salinity

provided in references. The result will be explained in the "results" chapter. It is worth noting again that this assessment aimed to evaluate the accuracy of the classification that only involved hydrological and spectral evidence. Any identified error (where the references do not support the hydrological and/or spectral evidence) has been corrected in the final classification result (which led to types 4, 6, 8, 9, and 10).

3.2.2 Estimating lake water storages

This section describes the methods of calculating the mean depth and the volume of each inventoried lake on the circa-2000 dataset, using peer-reviewed geostatistical models, terrain analysis, and a large volume of existing literature.

As discussed in Chapter 2 (literature review), two recent studies, Messager et al. (2016) and Cael et al. (2017), introduced the methods to estimate the mean depths and volumes of global water bodies. Messager et al. (2016) generated a high-resolution global lake dataset HydroLAKES, and estimated the depth and volume of each inventoried lake by using documentation and geo-statistical models. Cael et al. (2017) estimated the global lake water storage with an area-volume relationship. The estimation methods in both studies were derived from a large volume of literature, which means I referred to these documents indirectly when I applied these methods.

I here combined the documentation and estimation methods in Messager et al. (2016) and the area-volume function in Cael et al. (2017)) to estimate volumes of global lakes on the circa-2000 waterbody dataset. In general, the global lakes and reservoirs as inventoried in the circa-2000 mapping were divided to four size categories: 266 large natural lakes, 6,797 major reservoirs, ~1.4 million natural lakes with intermediate sizes (larger than 0.1 km^2), and the remaining ~6.6 million lakes smaller than 0.1 km^2. I replicated the methods of Messager et al. (2016) to estimate the volumes for the first three types, and applied the scaling equation from Cael et al. (2017) to calculate the volumes for the remaining smaller lakes. In Messager et al. (2016), particularly, the depths and volumes of large natural lakes were retrieved from the documentation (e.g., Herdendorf, 1982), the depths and volumes of reservoirs were extracted from GRanD (Lehner et al., 2011), and the depths and volumes of lakes with intermediate sizes were calculated by the geostatistical models that account for lake size and surrounding terrains. Since HydroLAKES

dataset developed in Messager et al. (2016) did not include the remaining lakes smaller than 0.1 km^2, I used Cael et al. (2017) to calculate their volumes. The models in both Messager et al. (2016) and Cael et al. (2017) involved the fitting uncertainties, which were propagated to the uncertainty calculation for my volume and depth estimations. As a result, four new fields (depth, volume, depth uncertainty, volume uncertainty) were added to all waterbodies in the circa-2000 lake dataset. The entire procedure was automatically completed by Python scripts (**Code 1**).

3.2.2.1 Reservoirs

The first step was to assign the mean depths and volumes of reservoirs included in the GRanD database (Lehner et al., 2011) into the reservoirs mapped in the circa-2000 inventory. GRanD database was the main source for volume estimation of reservoirs in Messager et al. (2016), so I also applied GRanD database to keep the same data source as in Messager et al. (2016). The depth, volume and their uncertainties for each lake in the circa-2000 inventory were calculated based on the values of the reservoir's area and depth that are provided by GRanD. There were two conditions: (1) one reservoir from the circa-2000 mapping is intersected with one reservoir from GRanD (a "one-to-one" intersection relation), and (2) many reservoirs from the circa-2000 mapping are intersected with one reservoir from GRanD (a "many-to-one" relation). For the "one-to-one" condition, the areas of the same lake may vary from different inventories due to different mapping conditions and the changes of lake extents. The "many-to-one" condition happened less often: sometimes as a mapping artifact due to the limitation of image resolution, and sometimes because of the lake inundation extent changes. For example, the Tuttle Creek Reservoir in Kansas, US (usually expected as one single polygon) may appear as several disintegrated water bodies (particularly in the upstream tail) during relatively low water level periods. If these happened, my depth and volume estimation took into account the areal disparity between our own mapping and the reference data (GRanD). In order to better apply the existing values in GRanD, an area adjustment was done for each inventoried lake on both conditions.

Specifically, the following equations were applied to calculate the reservoir water storages and uncertainties on each of the two conditions, and these equations were adapted from Cael et al. (2017):

(1) One reservoir from the circa-2000 mapping is intersected with one reservoir from GRanD (one to one).

$$d = (\frac{a}{A})^{0.2} \times D \quad 1)$$

$$V = a \times d \quad 2)$$

$$E_D = \left|\frac{D \times A - d \times a}{a}\right| \quad 3)$$

$$E_V = a \times E_D \quad 4)$$

where A denotes the area of the reservoir from GRanD (km^2), a the area of the reservoir from the circa-2000 mapping (m^2), D the depth of the reservoir form GRanD (m), d the depth of the reservoir from the circa-2000 mapping (m), V the volume of the reservoir from the circa-2000 mapping, E_D and E_V the uncertainties of the lake depth and volume from the circa-2000 mapping.

Equation 1) was adapted from the relationship between V and A determined by Cael et al. (2017). In Cael et al. 2017, the equation $V = 0.24\ A^{1.2}$ was derived from the ordinary least squares (OLS) linear regression model of $\log_{10}$ (volume) on $\log_{10}$ (area). The volume of a lake is multiplied by the lake area and mean depth. Then, equation 1) was adapted and used for calculating the mean depth of the lake.

Additionally, I also quantified the uncertainties for depth and volume induced by the estimation methods. The models in both Messager et al. (2016) and Cael et al. (2017) involved the fitting uncertainties. The uncertainty of volume was multiplied by the lake area and the uncertainty of the depth (equation 4)). The uncertainty of the depth was calculated by the difference between the recorded volume ($D \times A$) and the estimated volume ($V = a \times d$) to the lake area (a) which is from the circa-2000 mapping (equation 3)).

(2) Multiple reservoirs from the circa-2000 mapping are intersected with one reservoir from GRanD (many to one).

The proportion of each intersected reservoir area (circa-2000) is the principal factor for allocating the volume of the reservoir in GRanD to all intersected reservoirs. The mean depth of all reservoirs (in circa-2000) intersected with one reservoir in GRanD was calculated first

(equation 5)), and then it was allocated to each individual reservoir based on their area proportion (equation 6)). Based on the relationship between V and A determined by Cael et al. (2017), two coefficients α and β are solved to calculate and make the volume allocation.

$$d_{mean} = \left(\frac{a}{A}\right)^{0.2} \times D \quad 5)$$

$$d = (\frac{a}{a_{sum}})^{0.2} \times d_{mean} \times \alpha \quad 6)$$

$$V = a \times d \quad 7)$$

$$E_D = |(\frac{a}{A})^{0.2} \times D \times \beta - d| \quad 8)$$

$$E_V = a \times E_D \quad 9)$$

where d_{mean} is the mean depth of all reservoirs (circa-2000) intersected with one reservoir in GRanD, and it is an intermediate variable that will not be included in the final result. The meanings of other variables are the same as those in equation 1) to equation 4).

3.2.2.2 Large natural lakes

In this step, the volume estimation of 266 large natural lakes was done on the circa-2000 mapping by applying the volume records in HydroLAKES dataset. The general logic in volume estimations for large natural lakes is the same as the calculation for reservoirs, except that different datasets were applied. GRanD was applied for the volume estimation of global reservoirs, and HydroLAKES is the main source to assign the mean depths and the volumes of the 266 large natural lakes. The area adjustment was also necessary to be considered for large natural lakes owing to the areal disparity between circa-2000 mapping and the reference data (HydroLAKES). For example, the area of the Aral Sea decreased dramatically in recent years, so I obtained two values for the area of the Aral Sea from HydroLAKES and the circa-2000 mapping. Then the depth and the volume were assigned based on the area ratio (refer back to Equations 1)).

3.2.2.3 Natural lakes of intermediate sizes

I applied the empirical models derived from Messager et al. (2016) to perform their volume estimation for ~1.4 million intermediate-size lakes greater than 0.1 km^2 in the circa-2000 mapping. The models were geo-statistical relationships among lake mean depth, surface area, and surrounding topography (slope surrounding the lakes). Messager et al. (2016) divided all lakes which are larger than 0.1 km^2 to another four groups, as listed in the table below (**Table 3-2**). For each group, a specific model was fitted to calculating the depth.

Table 3-2 Empirical models for estimating the mean depths of natural lakes larger than 0.1 km^2.

Size class by area (km^2)	***Best-fit equation***	***Residual variance s^2***
0.1–1	$\log 10(D) = 0.3826 + 0.1512 \times \log 10(A) + 0.4820 \times \log 10(S_{100})$	0.0678
1–10	$\log 10(D) = 0.1801 + 0.2985 \times \log 10(A) + 0.8473 \times \log 10(S_{100})$	0.0689
10–100	$\log 10(D) = 0.0379 + 0.2445 \times \log 10(A) + 1.1517 \times \log 10(S_{100})$	0.0692
100–500	$\log 10(D) = 0.0123 + 0.2664 \times \log 10(A) + 1.1474 \times \log 10(S_{100})$	0.1094

Note: These equations were derived in Messager et al. (2016).

In **Table 3-2**, D and A are the depth and area of each inventoried lake, respectively. S_{100} represents the slope within 100 meters surrounding the lake. The area of each lake already existed in the mapping of the circa-2000 global lake dataset. Messanger et al. (2016) applied EarthEnv-DEM90 digital elevation model (Robinson et al., 2014) to develop/regress the equations for lake volume estimations, so for consistency I also used the same DEM in applying thesis equations. Furthermore, I corrected the bias in the mean depth induced by the back-transforming of the logarithmic scale, using equation 10) as per Ferguson (1986).

$$D_{bias-corrected} = 10^{log_{10} D} \times exp(2.65 \times s^2) \qquad 10)$$

where $D_{bias-corrected}$ is the bias-corrected mean depth, $log_{10} D$ is the logarithmic of the uncorrected mean depth as directly estimated from the geostatistical models in **Table 3-2**, and s^2 is the residual variance of these geostatistical models in **Table 3-2**. Therefore, this calculated $D_{bias-corrected}$ is the final depth of each lake, and then this depth was used to calculate the volume by multiplying the lake area.

3.2.2.4 Lakes smaller than 0.1 km^2

For the remaining 6.6 million lakes smaller than 0.1 km^2, which are not concluded in HydroLAKES, I applied the lake area-depth scaling function based on the Hurst coefficient of the Earth's surface, as per Cael et al. (2017). According to the same concept of bias-correction for in equation 10), I corrected the bias of the predicted volume induced by log-log regression, using equation 11).

$$log_{10}(v) = 1.024\, log_{10}(a) - 0.629 + log_{10}(exp(2.65 \times s^2)) \qquad 11)$$

where a and v are the area and the volume of each lake, respectively, s^2 is the model residual variances (0.292^2) (Cael et al., 2017). The lake depth can be calculated based on lake volume and lake area.

3.2.3 Assigning attributes to the updated circa-2015 dataset

To understand the change of the distribution and the abundance of global lakes during the recent 15 years, all lake attributes (freshwater/saline lake, volume estimation, uncertainties of depth and volume) were updated for each inventoried lake on the recently-produced circa-2015 dataset (Sheng et al., 2015; Sheng et al., 2016). The updated circa-2015 mapping contains ~9 million lakes, ~1 million lakes more than the circa-2000 inventory. However, during this period, the total area of global lakes decreased from 3,032 thousand km^2 to 2,988 thousand km^2. The changes also include the addition of many reservoirs that newly emerged during these 15 years. **Figure 3-5** shows only one reservoir that was constructed after 2000 as an example. In addition to the newly constructed reservoirs, the areas and volumes of some large lakes increased or decreased

during the 15 years. For example, evident shrinkages in Lake Chad and the Aral Sea from circa 2000 to 2015 are illustrated in **Figure 3-6**.

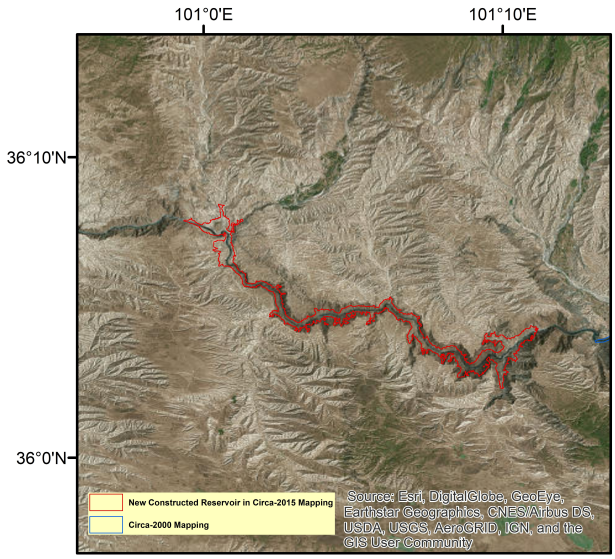


Figure 3-5 One example of a newly constructed reservoir in the circa-2015 lake inventory. This reservoir in the image is located in Qinghai, China (36.11°N, 101.09°E).

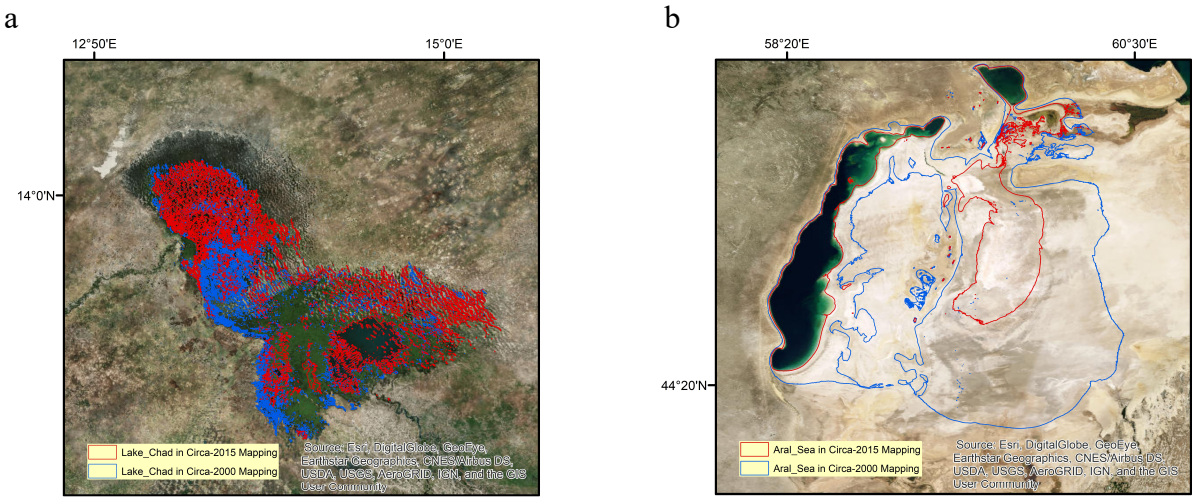


Figure 3-6 Spatial distribution of Lake Chad and the Aral Sea on the circa-2000 and the circa-2015 inventories.
Panels show the differences between Lake Chad (a) and the Aral Sea (b) on both the circa-2000 (Blue) mapping and the circa-2015 (Red) mapping. Areas of the two lakes both decreased dramatically.

Replicating the same freshwater and saline classification and the lake water storage estimation on the updated circa-2015 version is necessary to understand the change of

freshwater/saline lake distribution and lake storage during this 15-year period. But since a high-quality classification on the circa-2000 mapping has been completed, the replication did not have to start from the beginning. To improve the efficiency, I applied the existing result of circa-2000 inventory for circa-2015 classification and lake volume estimation.

Based on the classified circa-2000 results and the spatial relationship between two datasets, I assigned all existing attributes (saline, salinity, alkalinity, and reference) to the global lakes on the circa-2015 dataset. When classifying global lakes into the two types in the circa-2000 mapping, I found that it was a tedious and complicated task. One of the reasons was that based on the different conditions (in the endorheic basin or exorheic basin, in the arid region or non-arid region, lagoons or not lagoons), all lakes have been classified into 12 specific types in total (**Table 3-1**). When I assigned the result of freshwater and saline lake classification from the circa-2000 mapping to the circa-2015 mapping, all specific lake types were assigned to each inventoried lake. How to organize the order of 12 specific types was also considered when I assigned the attribute of different specific types to the new mapping. Based on the flow chart (**Figure 3-1**), the classification result would be more accurate if one lake has more than one type of evidence. Following a sequence from less to more evidence, my general steps were given in **Figure 3-7**. This figure illustrates that the circa-2015 lakes located within exorheic basins were considered first, followed by lakes in endorheic basins and then terminal lakes at the sink points without circa-2000 lakes. Then I assigned the attributes by using the order listed in **Table 3-3**. The automated assignments have been manually inspected against the original classification rule.

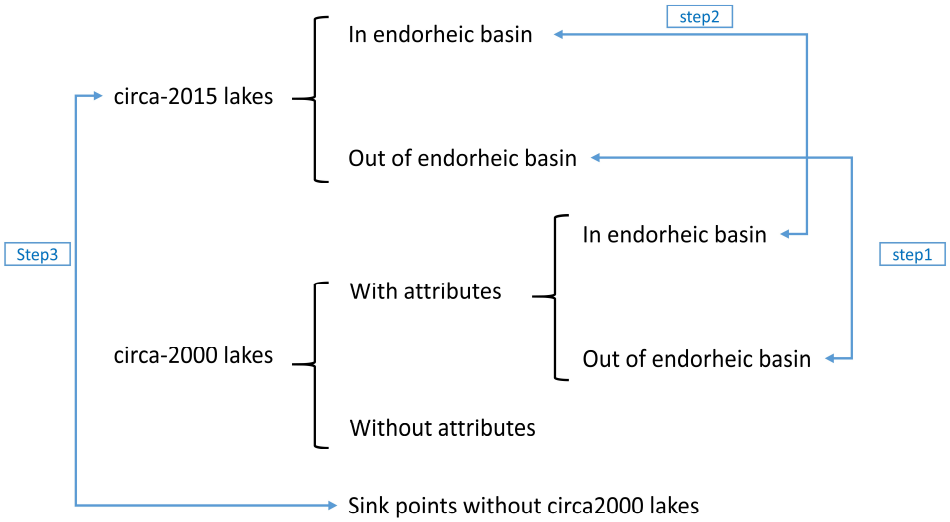


Figure 3-7 Steps for updating saline/freshwater types of global lakes from the circa-2000 to the circa-2015 inventories.

Table 3-3 The sequence of lake type (from less to more evidence) assigning procedure from the circa-2000 to the circa-2015 mappings.

Sequence	*Type*	*With literature reference (R) /Without literature reference (NR)*
1	6 (saline)	NR
2	5 (saline)	NR
3	Null (freshwater)	NR
4	10 (freshwater)	--
5	6 (saline)	R
6	5 (saline)	R
7	12 (freshwater)	--
8	7 (saline)	NR or “Spectral” in the reference field
9	11 (freshwater)	--
10	7 (saline)	R
11	3 (saline)	NR
12	Null (freshwater)	R
13	9 (freshwater)	--
14	3 (saline)	R
15	4 (saline)	--
16	2 (saline)	NR
17	1 (saline)	NR
18	8 (freshwater)	--
19	2 (saline)	R
20	1 (saline)	R

Similar to the previous steps, Python script **Code 1** was generated to assign the attributes automatically.

3.2.4 Lake water storage estimation on the circa-2015 mapping

The volumes, depths, and associated uncertainties for the circa-2015 lakes were calculated using the same method as for circa-2000 lakes described in Section 3.2.2. This replication was automated in Python script **Code 3**.

3.2.5 Spatial contrast between surface flow and storage

As introduced in Section 2.3, the annual river discharge averaged from 1990 and 2010, as simulated by the PCR-GLOWBW hydrological model (Van et al., 2011; Wada et al., 2011a, 2011b, 2014, 2016), was used to represent the abundance of annual surface flow in the circa-2000 era. The river flow and lake storage were compared in two geographical scales including 0.1° grid (consistent with the model resolution) and individual drainage basins. The HydroSHEDS watersheds dataset was used for the lake type classification, and was used here again to represent the boundaries of individual drainage basins. Some drainage basins, such as the Yangtze River Basin and Mississippi River Basin (**Figure 3-8**), are rather large in the geographic area. These large drainage basins may not be sufficient to reveal spatially-explicit variations of water storage and water flow. Therefore, I also downscaled the lake water storages and discharges into 0.1º × 0.1º grid for improved spatial details. In conclusion, there are two maps that were applied for assigning the values of average annual discharge, one is the HydroSHEDS basins (**Code 5**) and the other is the map with 0.1º × 0.1º geographic grid. The total storage in freshwater lakes and saline lakes were next aggregated in each HydroSHEDS basin and each 0.1° grid. Because the available river flow data described the annual river flow in the circa-2000 era, I only used circa-2000 lake storage data to do spatial contrast between surface flow and storage. All procedures were implemented by Python scripts automatically (**Code 2** and **Code 3**).

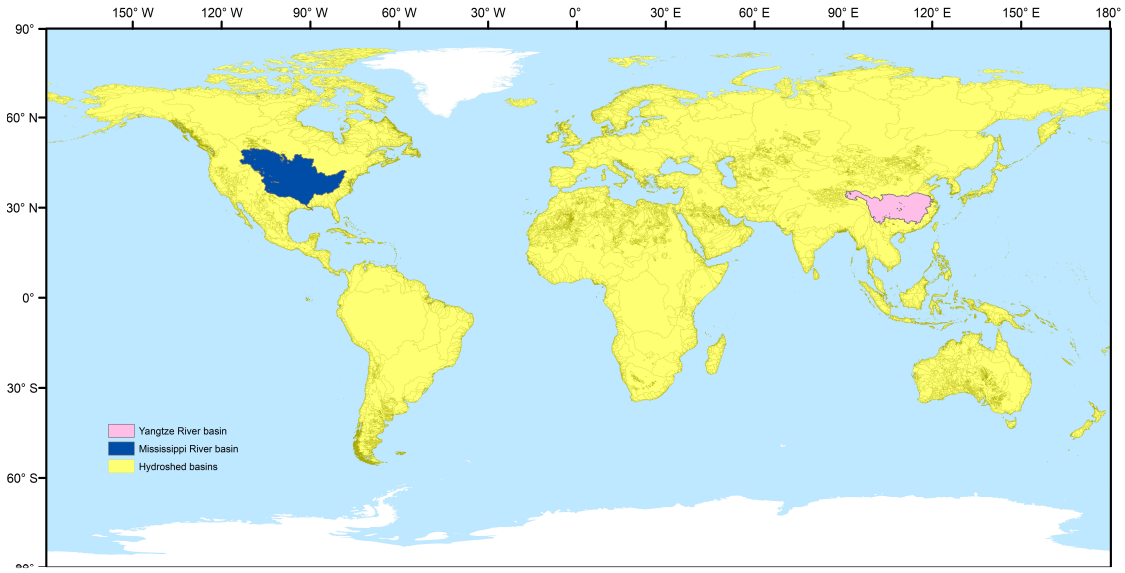


Figure 3-8 HydroSHEDS basins (15 arc second version) with the Yangtze River Basin and Mississippi River Basin highlighted.

Chapter 4 - Results

4.1 Global lake classification (freshwater and saline)

The first proposed question is where freshwater lakes and saline lakes are. **Figure 4-1** and **Figure 4-2** show the spatial distributions of these two lake types on circa-2000 and circa-2015 datasets, respectively. The general patterns of saline/freshwater lake distribution are similar in both two mappings. The number of freshwater lakes is much larger than saline lakes, and freshwater lakes are mainly distributed across the pan-Arctic region and the majority of saline lakes are located in endorheic and semi/arid regions.

In terms of regional distribution, there are a large number of freshwater lakes in North America (especially in Canada), China, South Asia, and South America. Due to maritime influences, there is higher humidity in most of the coastal regions, so another phenomenon on both two mappings is that freshwater lakes are also densely distributed in coastal areas. However, some other coastal areas (such as part of the Pacific state of the US, the Mediterranean, and North Africa) are dry due to atmospheric inversion and the patterns of ocean currents, so many saline lakes are distributed in those regions. In general, most saline lakes are highly concentrated on the subtropical high-pressure zone (near 30° S and 30° N, e.g., Central Asia, Sahara, Dry Andes, western US, and the Arabian Peninsula) where the climate is dry and deserts are widespread.

Differences between these two maps (circa-2000 and circa-2015 mapping) exist although their general spatial patterns are similar. For example, the area of saline lakes experienced a widespread global decline from circa 2000 to circa 2015. The areas of some saline lakes, such as the Aral Sea, Lake Urmia, Lake Poopó, and the Great Salt Lake, decreased significantly from 2000 to 2015. The possible causes include increasing human water consumption and drying climates (Pokhrel et al., 2012; Wada et al., 2017; Wang et al., 2018; Wurtsbaugh et al., 2017). However, there is a significant expansion of the saline lakes in the Tibetan Plateau due to increased precipitation as well as warming-induced glacial melting (Song et al., 2014; Yao et al., 2018; Zhang et al., 2017). These patterns are overall consistent with the recent studies of Donchyts et al. (2016), Pekel et al. (2016), Wang et al. (2018) and Wurtsbaugh et al. (2017). For example, Wang et al. (2018) indicated that most of the endorheic land-mass experienced a widespread water loss

during 2002-2016 with the exception of water gains in the Inner Tibetan Plateau, which is fairly consistent with the results between the circa-2000 and circa-2015 datasets presented here.

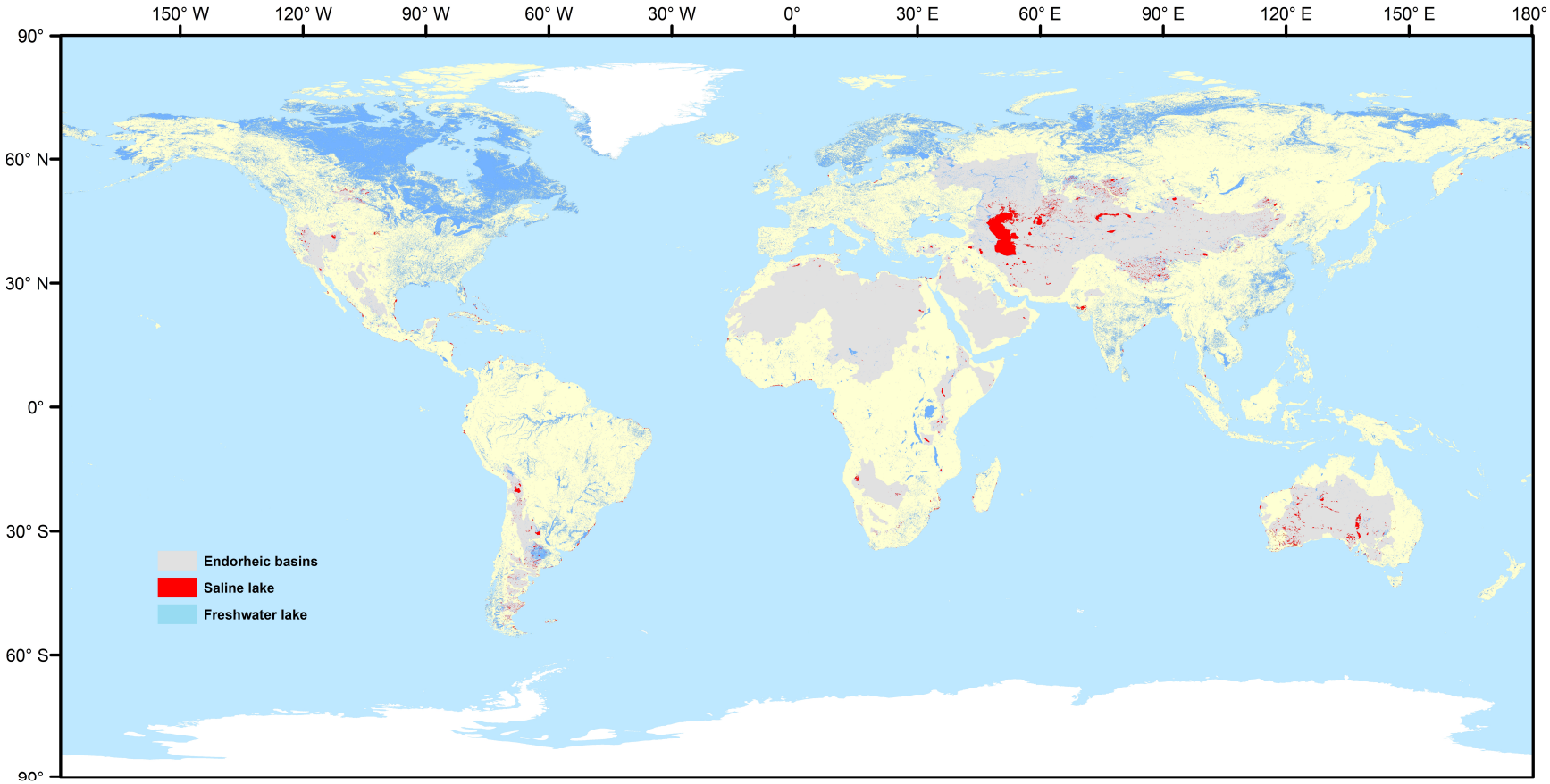


Figure 4-1 Global spatial distribution of freshwater and saline lakes on the circa-2000 mapping.

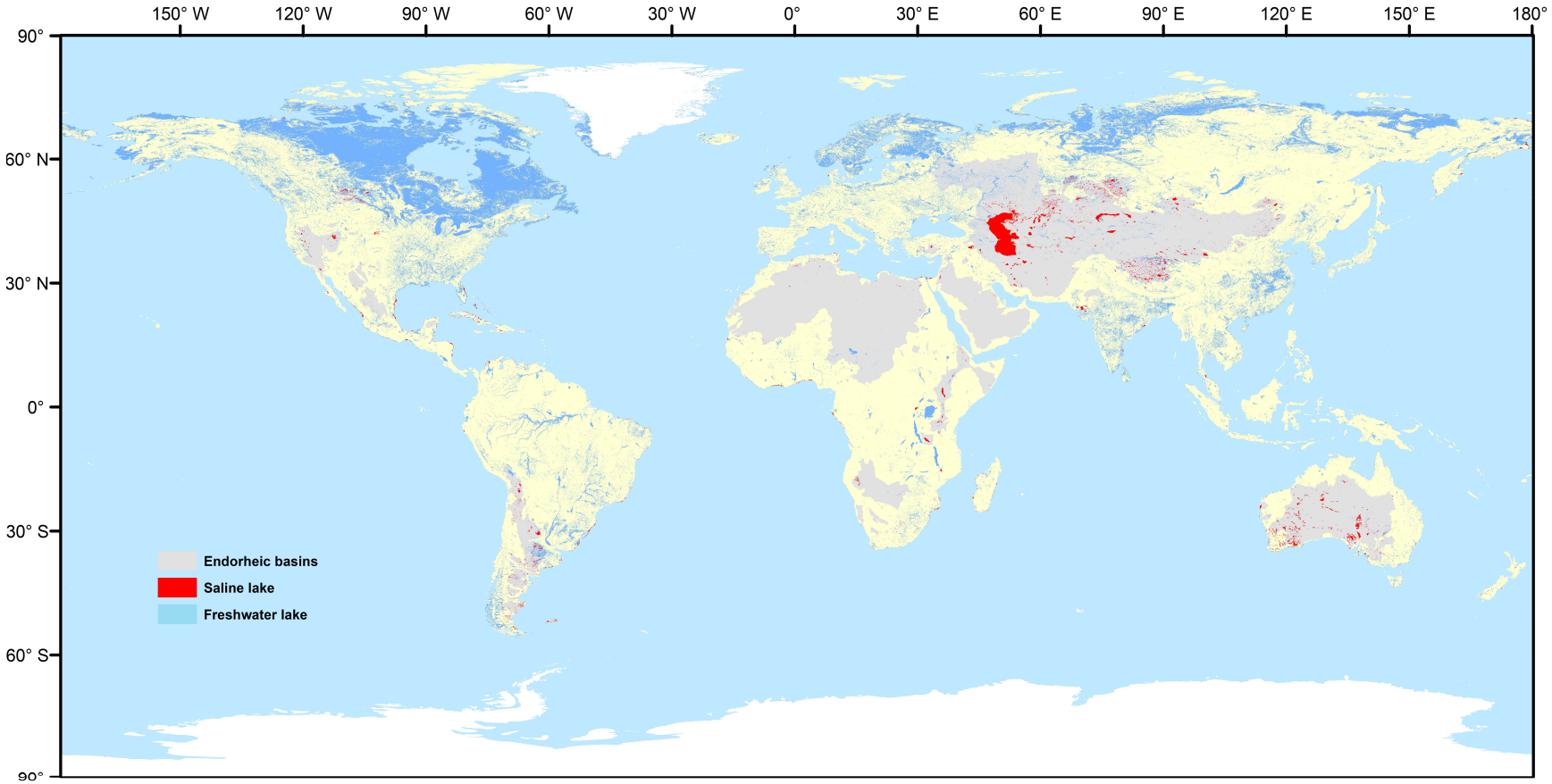


Figure 4-2 Global spatial distribution of freshwater and saline lakes on the circa-2015 mapping.

By combining spectral evidence, hydrological, and literature evidence, millions of lakes have been classified into 12 specific types, including 1, 2, 3, 4, 5, 6, 7 for saline lakes, and 8, 9, 10, 11, "Null" for freshwater lakes. The following graph (**Figure 4-3**) shows the number of each type (instead of lake area) on both circa-2000 and circa-2015 mappings. Despite millions of lakes that have been mapped on both two datasets, saline lakes only account for a small percentage in the total count. The percentages of saline lakes are 1.51% and 1.38% on the circa-2000 and circa-2015 mappings, respectively. After comparing **Figure 4-1 and 4-2**, it is clear that the area of saline lakes decreased. However, the following **Figure 4-3** shows that the numbers of several different-types saline lakes increased. For example, there are 30,470 lakes classified as type 1 (i.e., saline lakes at the sink points) on the circa-2000 mapping, and 34,455 lakes for type 1 on the circa-2015 mapping.

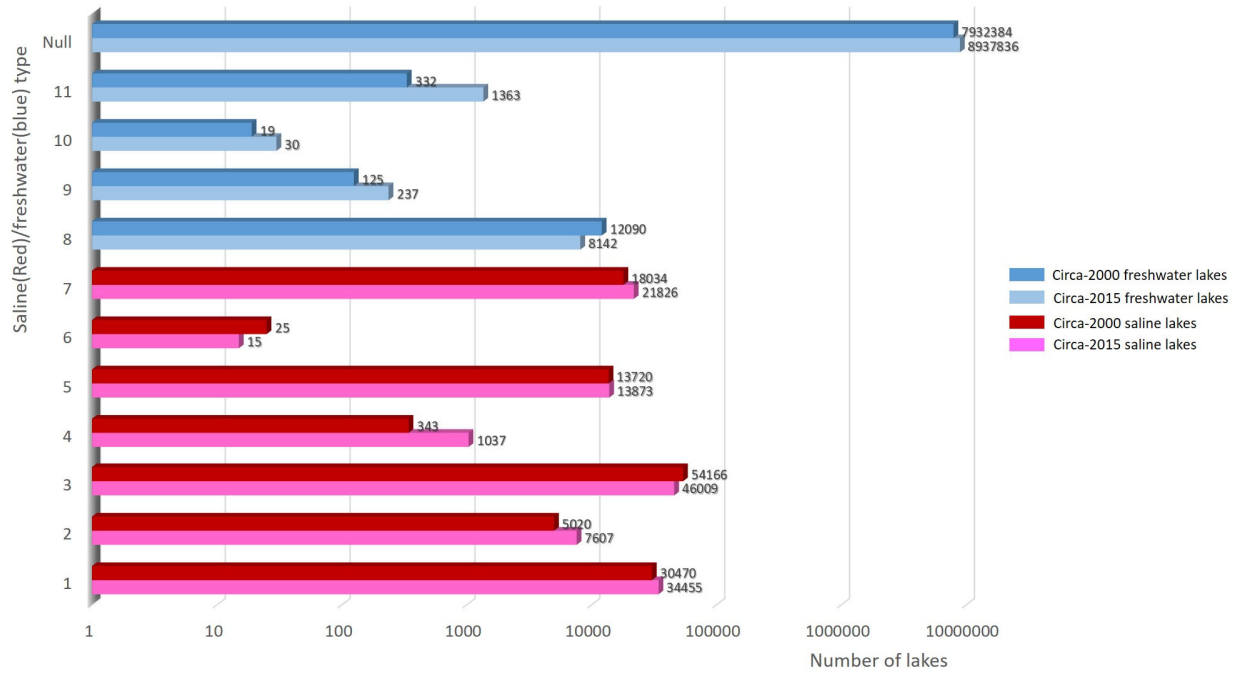


Figure 4-3 Quantities of freshwater and saline lakes on the circa-2000 and the circa-2015 inventories.

Blue bars and red bars represent the quantities of freshwater lakes and saline lakes, respectively.

According to the classification results, the classification error was assessed by calculating the number of misclassified lakes (as verified by literature) as a percentage of the number of all classified lakes where types are not "Null" (refer back to Section 3.2.1 for details). **Table 4-1** shows that the number of all classified lakes is 134,344, and the number of inaccurately classified lakes is 12,934. The ratio of these two gives 0.096 (or 9.6%), which suggests if only hydrological evidence and/or spectral evidence were applied, at least 9.6% of the lakes might be misclassified. The classification accuracy is thus about 90%. I acknowledge that this accuracy assessment is dependent on the number of references collected for verification. Although I tried to collect as many as possible (with details provided in the "reference" attribute), more references may always be added to further improve the verification. Therefore, the error given here is likely conservative. As described in Section 3.2.1, these identified errors were marked as types 4, 6, 8, 9, and 10 depending on their evidence scenarios (**Figure 3-2**) and have been associated with the correct freshwater or saline types in the final classification result. In **table 4-1**, type 8 representing the terminal lakes but verified to be freshwater by the literature, accounts for the majority of the misclassified lakes. Most of these lake bodies, however, come from several terminal regions where lake water surfaces appear disintegrated.

Table 4-1 Quantities of specific lake types for classification accuracy evaluation.

All types	***Count***	***Inaccurately-classified types***	***Count***
1 (saline)	30,470	*4 (saline)*	343
2 (saline)	5,020	*6 (saline)*	25
3 (saline)	54,166	*8 (freshwater)*	12,090
4 (saline)	343	*9 (freshwater)*	125
5 (saline)	13,720	*10 (freshwater)*	19
6 (saline)	25		
7 (saline)	18,034		
8 (freshwater)	12,090		
9 (freshwater)	125		
10 (freshwater)	19		
11 (freshwater)	332		
Total	134,344	***Total***	12,934
Error	12,934/134,344 = 9.6%		

Figure 4-4 presents the number of lakes under different area levels for each of the six continents (except Antarctica). These figures were generated by a joint use of Python scripts (**Code 4**) and R scripts (**Code 5**). In general, lake count and size follow a log-log linear relationship, also termed as Pareto or power-law distribution. The quantity of lakes decreases as the lake area increases, and the decreasing rate stays fairly constant on a logarithmic scale. Accordingly, a power-law equation was fitted for the plot of each continent in order to further quantify the pattern.

As the relations in **Figure 4-4** implies, the number of global lakes has an approximately 10-fold decrease from one logarithmic size to the next one. All six graphs show that when the lake area gradually increases, the number of freshwater lakes is close to saline lakes until the area is roughly 10^4 km^2. For most continents, the numbers of freshwater lakes are substantially higher than saline lakes (except Oceania), which is consistent with their spatial distribution shown in the global freshwater/saline lake maps (**Figure 4-1** and **Figure 4-2**). At the beginning of the curves where lake sizes are small, the numbers of freshwater lakes are about 100 times greater than those of saline lakes in Asia, Europe, and North America. However, in Africa and South America, the

difference is approximately 10 times. In Oceania (**Figure 4-4d**), the counts of freshwater and saline lakes are overall comparable under different area levels, owing to its large areas of endorheic basins and arid climate.

According to the global spatial distribution of freshwater and saline lakes (**Figure 4-1**), the region with the highest freshwater lake density is North America. By observing **Figure 4-4e**, the number of freshwater lakes is more than 100 times of saline lakes when the lake area is smaller than 10^3 km^2, and the number of freshwater lakes smaller than 10^{-2} km^2 (or 1 hectare) area exceeds 10^5. However, the numbers of freshwater and saline lakes are close to each other when the lake area lies between 10^4 to 10^5 km^2.

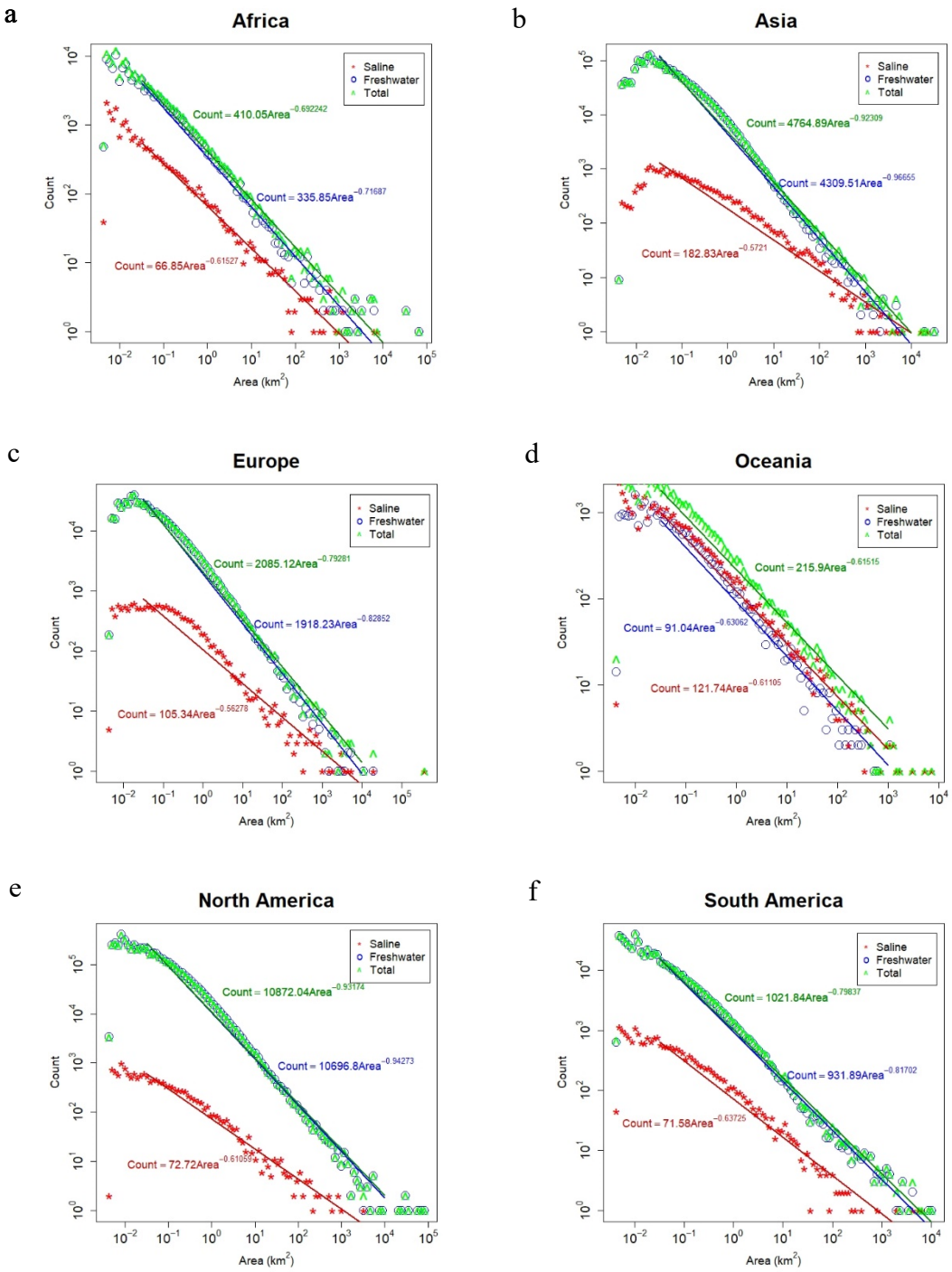


Figure 4-4 Relationships between lake area and abundance (count) in each continent. Panels show the relationships between lake area and lake count (both in logarithmic scales) in Africa (a), Asia (b), Oceania (c), Europe (d), North America (e), and South America (f). Red, blue and green dots represent saline lakes, freshwater lakes and both of them, respectively.

4.2 Global lake and reservoir volumes

The third research objective is to estimate the total water storages in freshwater and saline lakes. The methods produced by Messager et al. (2016) have been applied on reservoirs, large natural lakes, and lakes of intermediate sizes (see Section 3.2.2). In order to evaluate the accuracy of the volume estimation results, I compared my results with those in HydroLAKES and GRanD (**Table 4-2** and **Table 4-3**).

Table 4-2 Comparison of estimated lake volumes (for lakes larger than 0.1 km^2) between circa-2000/2015 inventories and HydroLAKES.

Result	*This study (Gt)*	*HydroLAKES (Gt)*	*Difference (Gt)*	*Difference (%)*
Circa-2000	183,055.69	182,900	+155.69	+0.08
Circa-2015	183,624.84	182,900	+724.84	+0.39

Note: "+" means the estimated volume in this thesis greater than the result in HydroLAKES.

Table 4-3 Comparison of estimated reservoir volumes between circa-2000/2015 inventories and GRanD.

Result	*This study (Gt)*	*GRanD (Gt)*	*Difference (Gt)*	*Difference (%)*
Circa-2000	6,057.850	6,128.93	–71.080	–1.16
Circa-2015	6,065.239	6,128.93	–63.691	–1.05

Note: "–" means the estimated volume in this thesis less than the result in GRanD.

Given the comparisons in **Table 4-2**, the estimated total volumes for lakes larger than 0.1 km^2 in both circa-2000 and circa-2015 inventories slightly exceed the result in HydroLAKES, i.e., +0.08% in circa 2000 and +0.39% in circa 2015. The estimated total volumes for reservoirs in both

circa-2000 and circa-2015 inventories (**Table 4-3**) are slightly below the result in GRanD, i.e., –1.16% in circa 2000 and –1.05% in circa 2015. These minor differences indicate that my volume estimations are highly consistent with HydroLAKES. The total volume on the circa-2000 mapping is ~156 Gt more than that of HydroLAKES, and circa-2015 mapping is ~725 Gt more than HydroLAKES (**Table 4-2**). Due to the finer spatial resolution (0.004 km^2) in circa-2000/2015, the lakes smaller than 0.1 km^2 were mapped in both these two inventories. Therefore, the volume of these small lakes leads to the exceeded part between circa-2000/2015 and HydroLAKES. The reservoir areas mapped in circa-2000/2015 inventories do not always represent the maximal areas under the reservoir storage capacities, which likely explains why my estimated reservoir volumes (**Table 4-3**) are smaller than that in GRanD.

In order to present and compare the areas and storages of natural lakes, saline lakes and reservoirs on both mappings, four pie charts (**Figure 4-5**) were created. In general, the area of saline lakes accounts for a quarter of the total area of global lakes and reservoirs, but the volume of saline lakes is approximately one half of the global total. Moreover, the area and the storage of saline lakes decreased from 2000 to 2015, which is consistent with the previous results related to the decrease of saline lake area in Section 4.1. The Caspian Sea, Great Lakes, and Lake Baikal are also highlighted in the four charts. The total volume of these three large lakes accounts for ~63% of the global storage. The Caspian Sea accounts for ~90% of the total storage of global saline lakes, and other saline lakes contribute the remaining ~10%. Reservoirs are important freshwater resource, and they play a significant role in biophysical environments and human societies. The total area and volume of reservoirs are comparable to those of the saline lakes excluding the Caspian Sea. This result further supports that the classification of freshwater/saline lakes is necessary if we want to accurately assess the abundance of global freshwater lake storage.

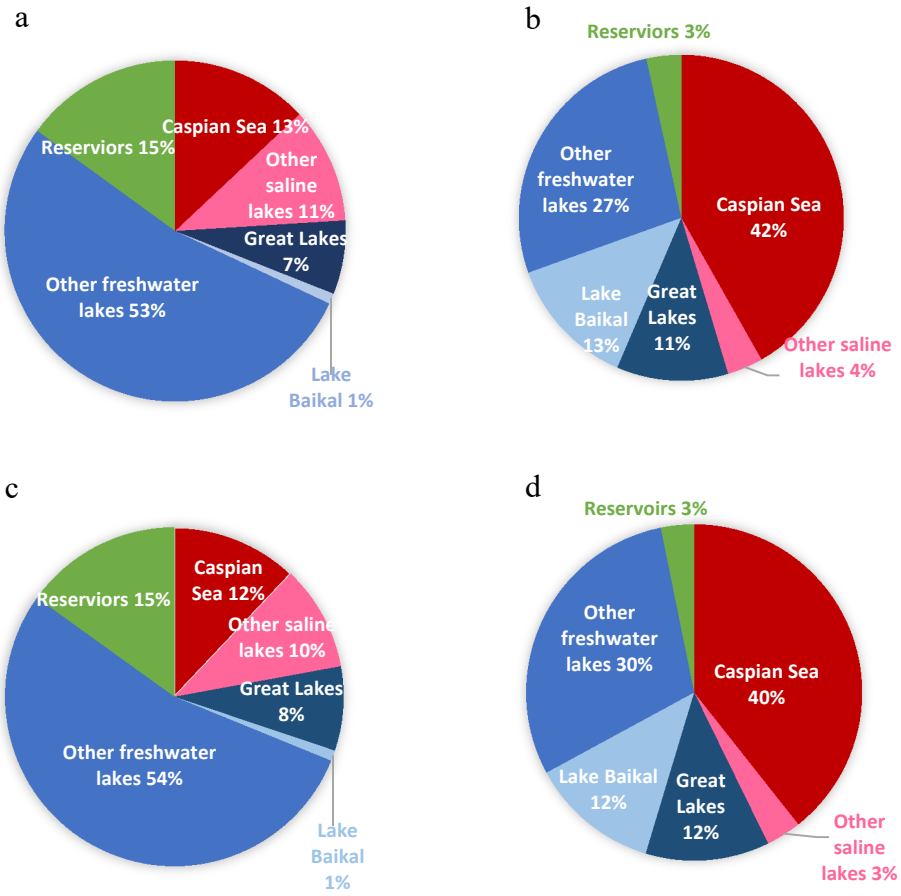


Figure 4-5 Area/Volume percentages for saline and freshwater lakes on two mappings. Panels show the area percentages of saline and freshwater lakes on both circa-2000 (a) and circa-2015 (c) mappings, volume percentages of saline and freshwater lakes on both circa-2000 (b) and circa-2015 (d) mappings. The total areas are 3.02 and 2.97 million km^2 on the circa-2000 and circa-2015 mappings, respectively. The total volume is 189 (±6) thousand Gt on the circa-2000 mapping, and 190 (±6) thousand Gt on circa-2015 mapping. Red, blue, and green colors represent saline lakes, freshwater lakes, and reservoirs, respectively.

Similar to the distributions between lake area and lake count, six graphs (**Figure 4-6**) were created using **Code 6** and **Code 7** to show the numbers of lakes under different volumes for each continent (except Antarctica). These plots show that the lake number decreases as the volume increases in a Pareto or power-law distribution. Similar to the relationship between lake area and count, the abundance of freshwater and saline lakes is comparable under different volume levels in Oceania. However, in other continents, it is apparent that the numbers of freshwater lakes are substantially higher than those of saline lakes. This is especially true in Asia, Europe, South

America, and North America, where the numbers of freshwater lakes are about 100 times of those of saline lakes when the volumes lie between 10^{-5} Gt (10^{4} m^{3}) and 10^{-3} Gt (10^{6} m^{3}).

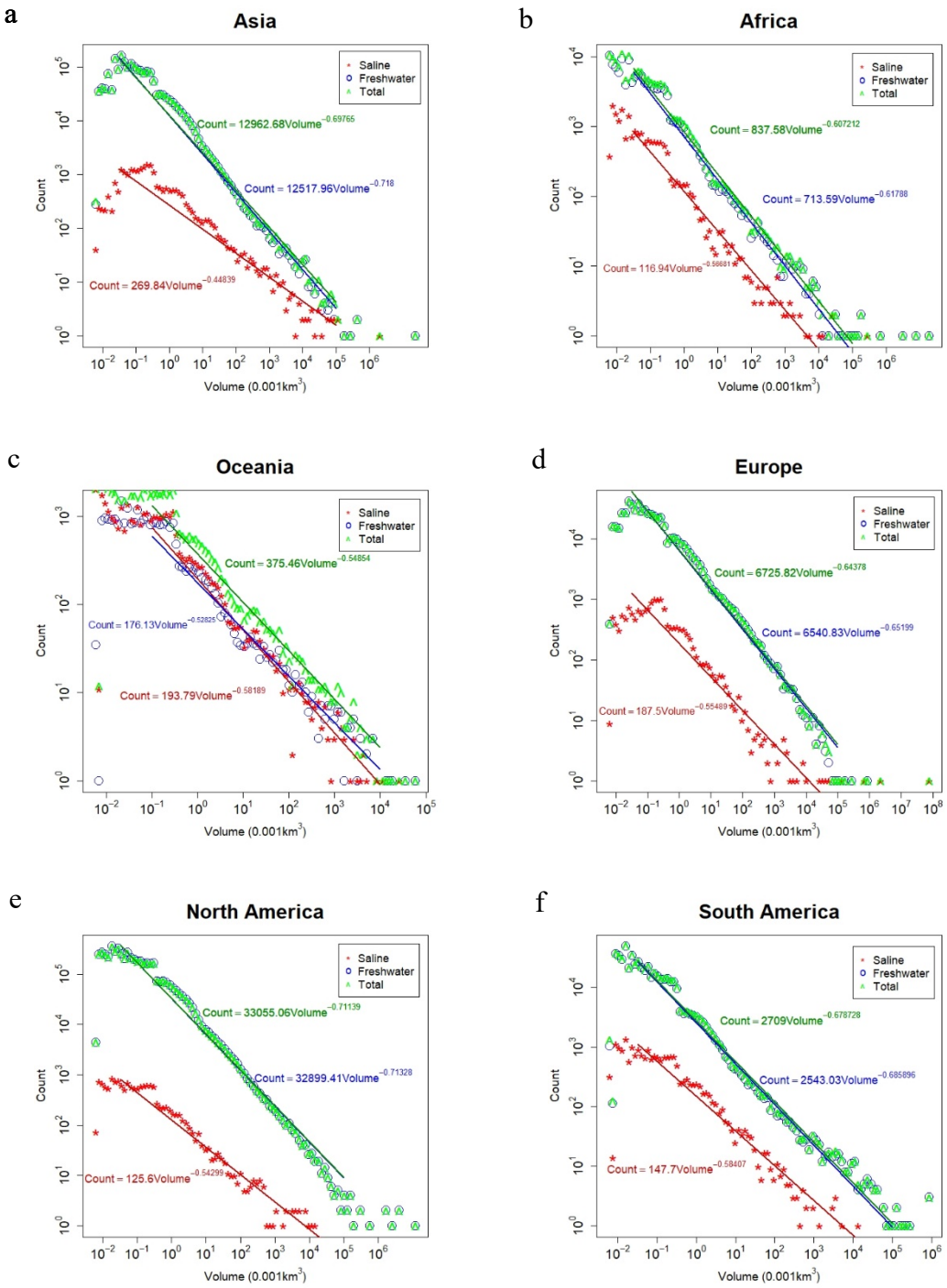


Figure 4-6 Relationships between lake volume and abundance (count) in each continent. Panels show the relationships between lake volume and lake count (both in logarithmic scales) in Africa (a), Asia (b), Oceania (c), Europe (d), North America (e), and South America (f). Red, blue and green dots represent saline lakes, freshwater lakes and both of them, respectively.

Summarized from the continental distributions, **Figure 4-7** shows the freshwater and saline lake abundance as relations of area and volume on a global scale. The general pattern shows that with the increase of area or volume, the numbers of global freshwater lakes and saline lakes decrease significantly. All three plots for global lakes also follow the Pareto distribution. Furthermore, the difference of lake count between global freshwater and saline lakes is large under the smallest size class (the number of freshwater lakes is about 100 times than saline lakes), but under the large area or volume, the numbers of two types are comparable to each other.

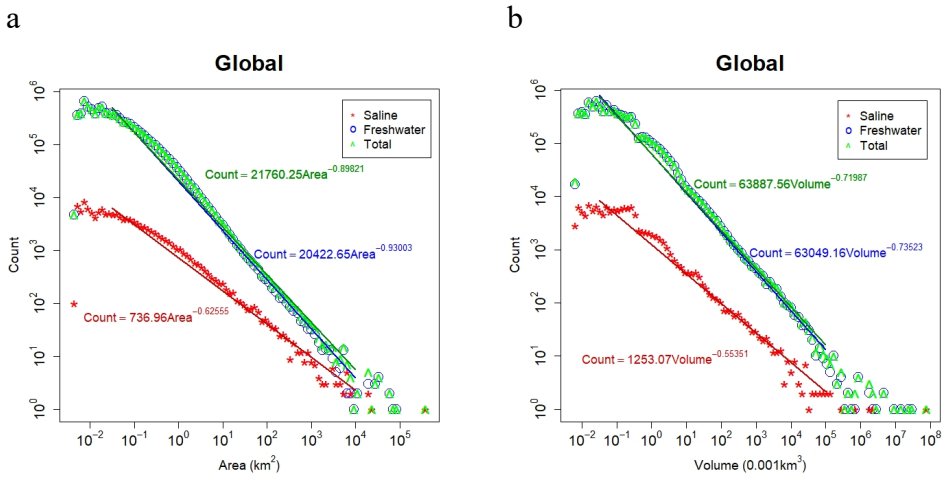


Figure 4-7 Relationships between global lake area/volume and abundance (count). Panels show the relationships between lake area (a)/volume (b) and lake count (both in logarithmic scales) on a global scale. Red, blue and green dots represent saline lakes, freshwater lakes and both of them, respectively.

The fitted Pareto (power-law) relationship are provided for each of the area-count and volume-count plots on both continental and global scales (**Figures 4-4**, **4-6**, **4-7**) and the equation parameters are listed in **Table 4-4**. Except for Oceania, the number of freshwater lakes substantially exceeds that of saline lakes, so the plot for freshwater lakes appears to overlap that for the both combined. Therefore, due to the same trend and the almost overlapped patterns of the fitted lines, the coefficients a and b for freshwater power-law equations are close to those for both types combined. As shown in **Figures 4-4**, **4-6** and **4-7**, the power-law relationship does not hold for those points when the lake areas are sufficiently small. In other words, the actual number of

small lakes is below the fitted power-law relation. The thresholds below which the power-law relation may not hold are provided in **Table 4-4**. These thresholds for different curves are not always the same. This finding is attributed to the limitation of lake mapping or the law of nature. The resolution for both circa-2000 and circa-2015 mappings is 0.004 km^2, so the lake smaller than 0.004 km^2 were not mapped in those maps. It is not possible to observe the relationship between lake area and lake count when the lake area is smaller than 0.004 km^2, so we cannot compare the actual distribution to the fitted equation. Another reason may be the law of nature, which means it is true that when the lake area is extremely small, their area and count do not follow a power-law relationship.

Table 4-4 Power law distributions for lake abundance (count) under area and volume levels for each continent and the globe.

Tables a, b, c show the parameters of power-law distribution ($Y=aX^b$ where X represents lake area or volume and Y lake count) for saline lakes, freshwater lakes and both. Thresholds represent the values below which the power-law relations do not appear to hold.

a. Saline lakes	***Area-Count***			***Volume-Count***		
	Threshold	***a***	***b***	***Threshold***	***a***	***b***
Africa	0.008313	66.85	-0.61527	0.007771	116.94	-0.56681
Asia	0.020925	182.83	-0.57210	0.037128	269.84	-0.44839
Europe	0.01897	105.34	-0.56278	0.016141	187.5	-0.55489
Oceania	0.004813	121.74	-0.61105	0.008148	193.79	-0.58189
South America	0.004846	71.58	-0.63725	0.015607	147.7	-0.58407
North America	0.008399	72.72	-0.61059	0.018277	125.6	-0.54299
Global	0.007535	736.96	-0.62555	0.00799	1253.07	-0.55351

b. Freshwater lakes	*Area-Count*			*Volume-Count*		
	Threshold	**a**	**b**	**Threshold**	**a**	**b**
Africa	0.008313	335.85	-0.71687	0.006234	713.59	-0.61788
Asia	0.020925	4309.51	-0.96655	0.037128	12517.96	-0.71800
Europe	0.01897	1918.23	-0.82852	0.016141	6540.83	-0.65199
Oceania	0.00995	91.04	-0.63062	0.021639	176.13	-0.52825
South America	0.004846	931.89	-0.81702	0.015607	2543.03	-0.685896
North America	0.008399	10696.8	-0.94273	0.018277	32899.41	-0.71328
Global	0.007535	20422.65	-0.93003	0.010101	63049.16	-0.73523

c. Both	*Area-Count*			*Volume-Count*		
	Threshold	***a***	***b***	***Threshold***	***a***	***b***
Africa	0.008313	410.05	-0.692242	0.006234	837.58	-0.607212
Asia	0.020925	4764.89	-0.92309	0.037128	12962.68	-0.69765
Europe	0.01897	2085.12	-0.79281	0.016141	6725.82	-0.64378
Oceania	0.004813	215.9	-0.61515	0.021639	375.46	-0.54854
South America	0.004846	1021.84	-0.79837	0.015607	2709.0	-0.678728
North America	0.008399	10872.04	-0.93174	0.018277	33055.06	-0.71139
Global	0.007535	21760.25	-0.89821	0.010101	63887.56	-0.71987

Table 4-5 provides more detailed information (count, area, volume, uncertainties, and the percentages of count/area/volume) about the results of lake classification and volume estimation. From this table, the ratio of saline lake area to freshwater lake area is close to 2:7, so saline lakes

account for about one-quarter of global lakes in the area. In terms of lake volume, the total freshwater lakes and reservoirs account for 56% of the global lakes and reservoirs. By comparing the attribute information between circa-2000 and circa-2015 lake mappings, I found that the number of saline lakes increased by ~3000, but the area of total saline lakes decreased by ~1.1%. This result is also consistent with my previous explanation for the difference between circa-2000 and circa-2015 freshwater/saline distributions (see details in Section 4.1).

Table 4-5 Summary of global freshwater and saline lake area and volume on both circa-2000 and circa-2015 mappings.

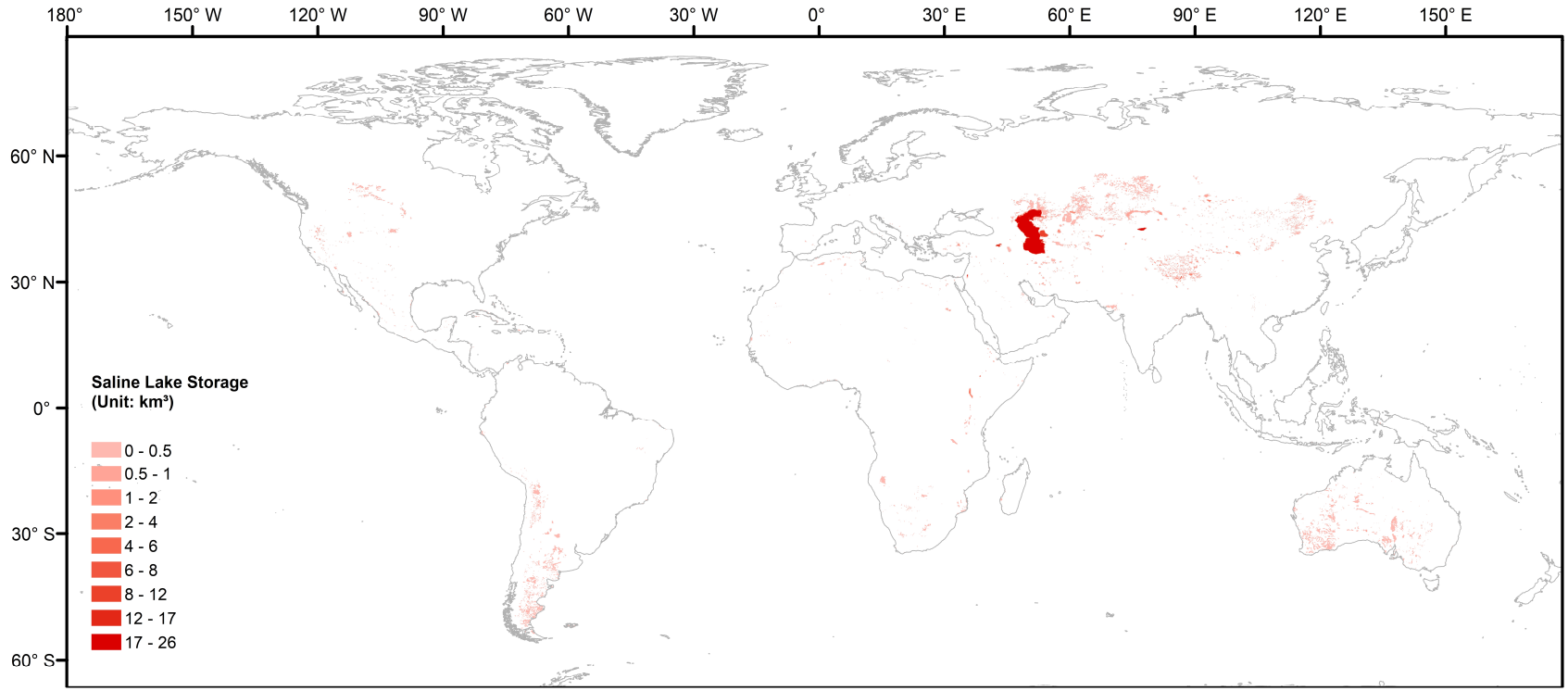
	Circa-2000 lake mapping						
	Number	***Area (km^2)***	***Volume (Gt)***	***Volume error (Gt)***	***Number of lakes percentage***	***Area percentage***	***Volume percentage***
Saline	121,778	689,088.46	81,837.73	615.05	1.51%	22.81%	43.34%
Freshwater	7,944,950	2,330,986.54	107,006.85	13,582.72	98.49%	77.19%	56.66%
Total	8,066,728	3,020,075.00	188,844.58	14,197.77	100%	100%	100%
	Circa-2015 lake mapping						
Saline	124,822	647,299.98	81,223.40	316.19	1.38%	21.75%	42.74%
Freshwater	8,947,608	2,329,038.00	108,807.60	10,752.91	98.62%	78.25%	57.26%
Total	9,072,430	2,976,337.98	190,031.00	11,069.10	100%	100%	100%

4.3 Comparison between water flow and lake water storage

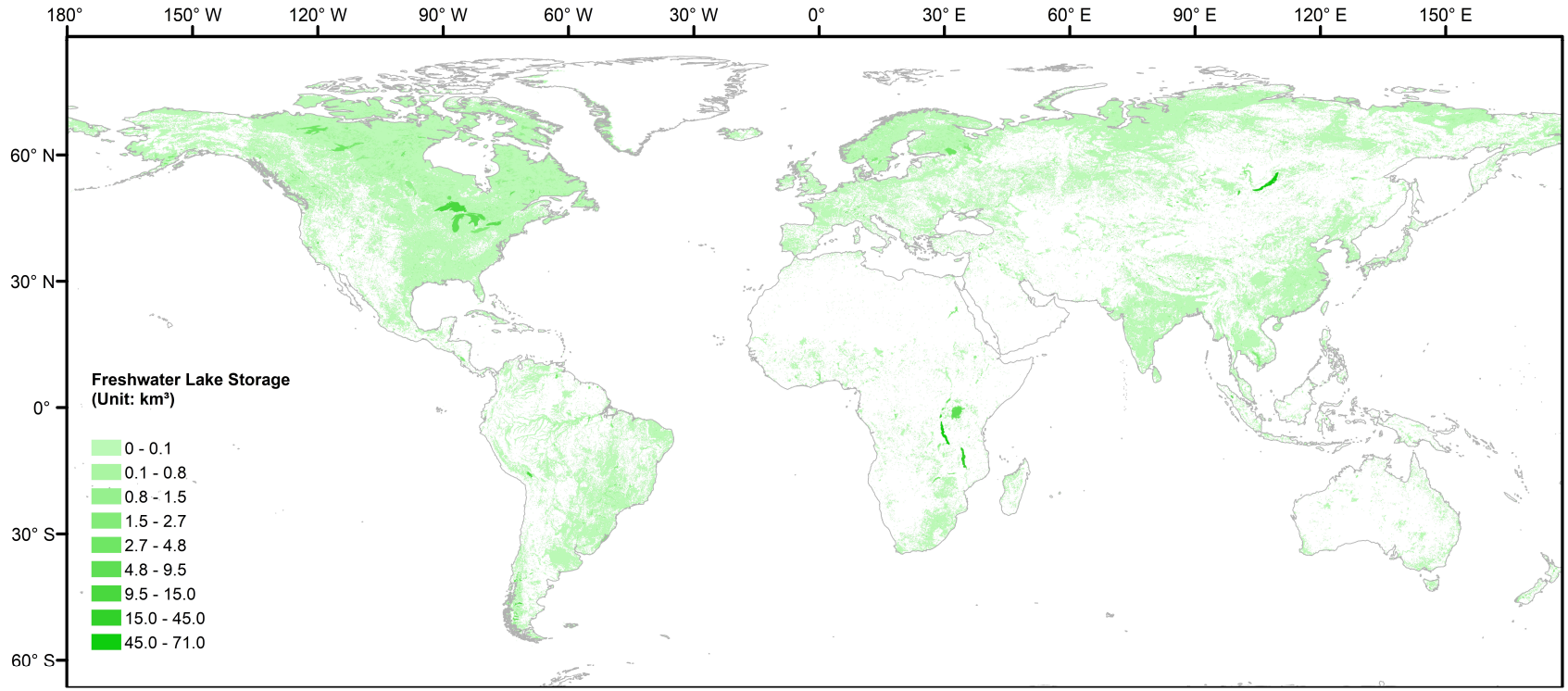
My estimated lake water storage and model-simulated water flow were compared on two different scales (one is 0.1° geographic grid (**Figure 4-8**), and the other is drainage basins (**Figure 4-9**)) for understanding the spatial relationship between surface water flux and storage on a global scale. These maps display the global patterns of contemporary surface water storage (in both freshwater and saline lakes) and river discharge around the year 2000. **Figure 4-8** contains three high-resolution maps with 0.1° geographic grid. They are helpful to show spatially-explicit

variations of river flow and lake storage. **Figure 4-9** contains three maps to show three categories in HydroSHEDS basins. They can better explain the global general patterns of freshwater lake storage, saline lake storage, and discharge.

a


180°
150° W
120° W
90° W
60° W
30° W
0°
30° E
60° E
90° E
120° E
150° E
60° N
30° N
0°
30° S
60° S
Saline Lake Storage
(Unit: km³)
0 - 0.5
0.5 - 1
1 - 2
2 - 4
4 - 6
6 - 8
8 - 12
12 - 17
17 - 26

b


180°
150° W
120° W
90° W
60° W
30° W
0°
30° E
60° E
90° E
120° E
150° E
60° N
30° N
0°
30° S
60° S
Freshwater Lake Storage
(Unit: km³)
0 - 0.1
0.1 - 0.8
0.8 - 1.5
1.5 - 2.7
2.7 - 4.8
4.8 - 9.5
9.5 - 15.0
15.0 - 45.0
45.0 - 71.0

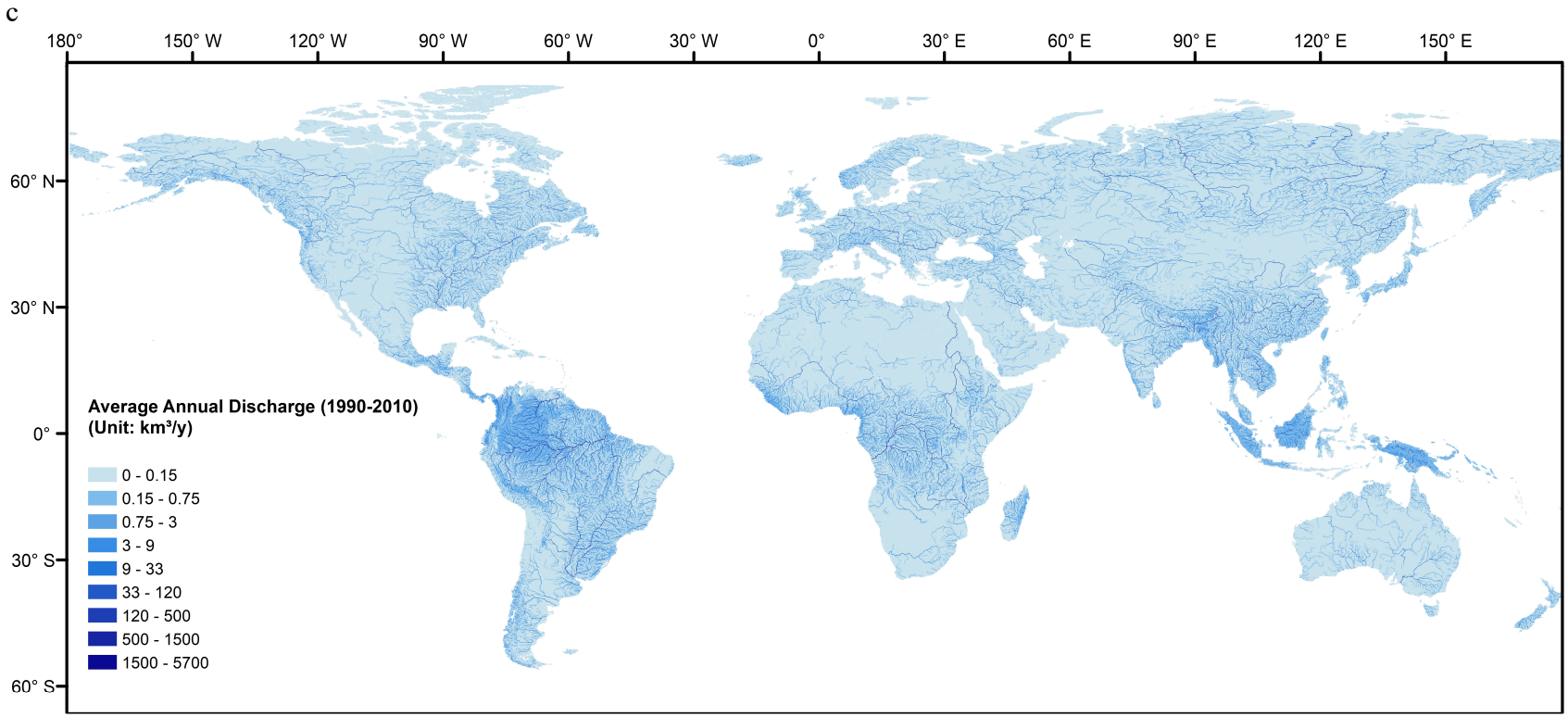
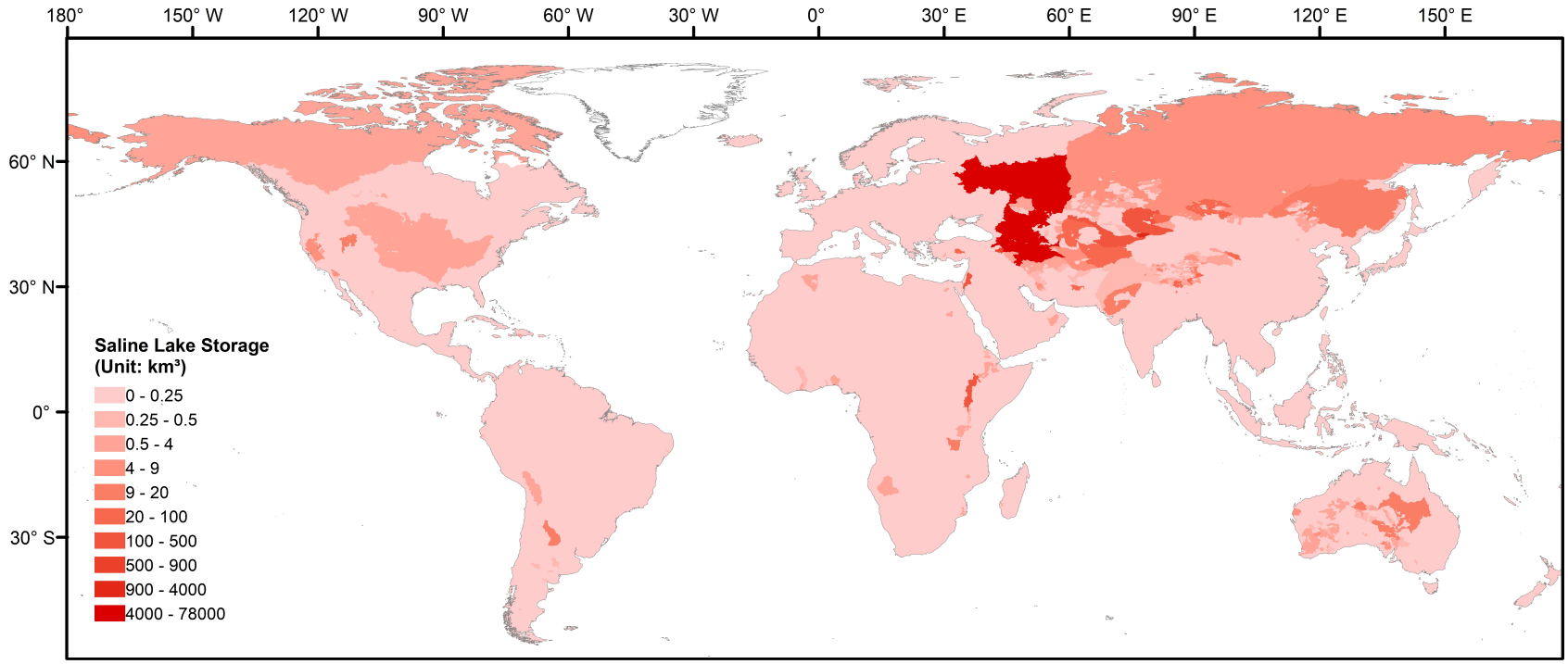
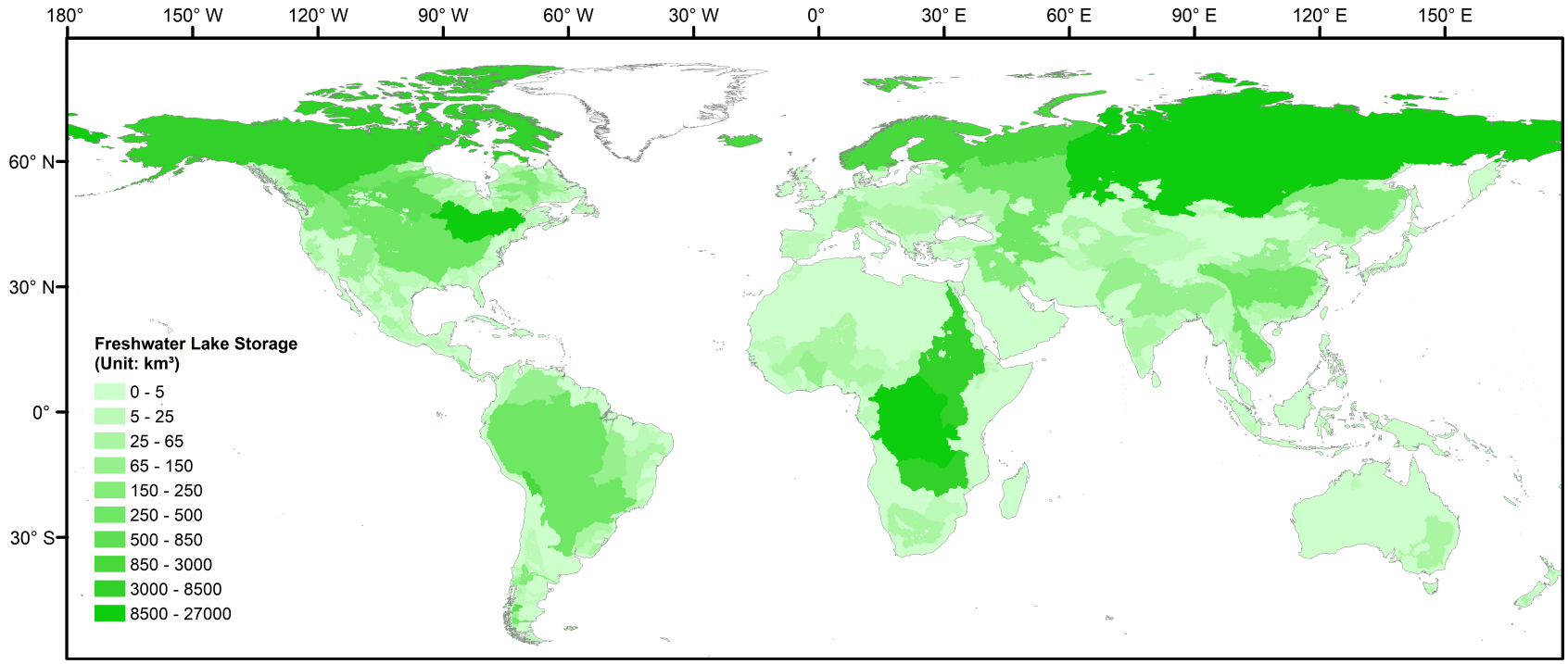


Figure 4-8 Spatial distributions of saline lake storage, freshwater lake storage and annual discharge on 0.1º grid. Panels represent the global volume distribution of saline lakes (a), freshwater lakes (b), and global annual average (1990–2010) discharge on 0.1° global grid (based on the circa-2000 mapping).

a


180°
150° W
120° W
90° W
60° W
30° W
0°
30° E
60° E
90° E
120° E
150° E
60° N
30° N
0°
30° S
Saline Lake Storage
(Unit: km³)
0 - 0.25
0.25 - 0.5
0.5 - 4
4 - 9
9 - 20
20 - 100
100 - 500
500 - 900
900 - 4000
4000 - 78000

b


180°
150° W
120° W
90° W
60° W
30° W
0°
30° E
60° E
90° E
120° E
150° E
60° N
30° N
0°
30° S
Freshwater Lake Storage
(Unit: km³)
0 - 5
5 - 25
25 - 65
65 - 150
150 - 250
250 - 500
500 - 850
850 - 3000
3000 - 8500
8500 - 27000

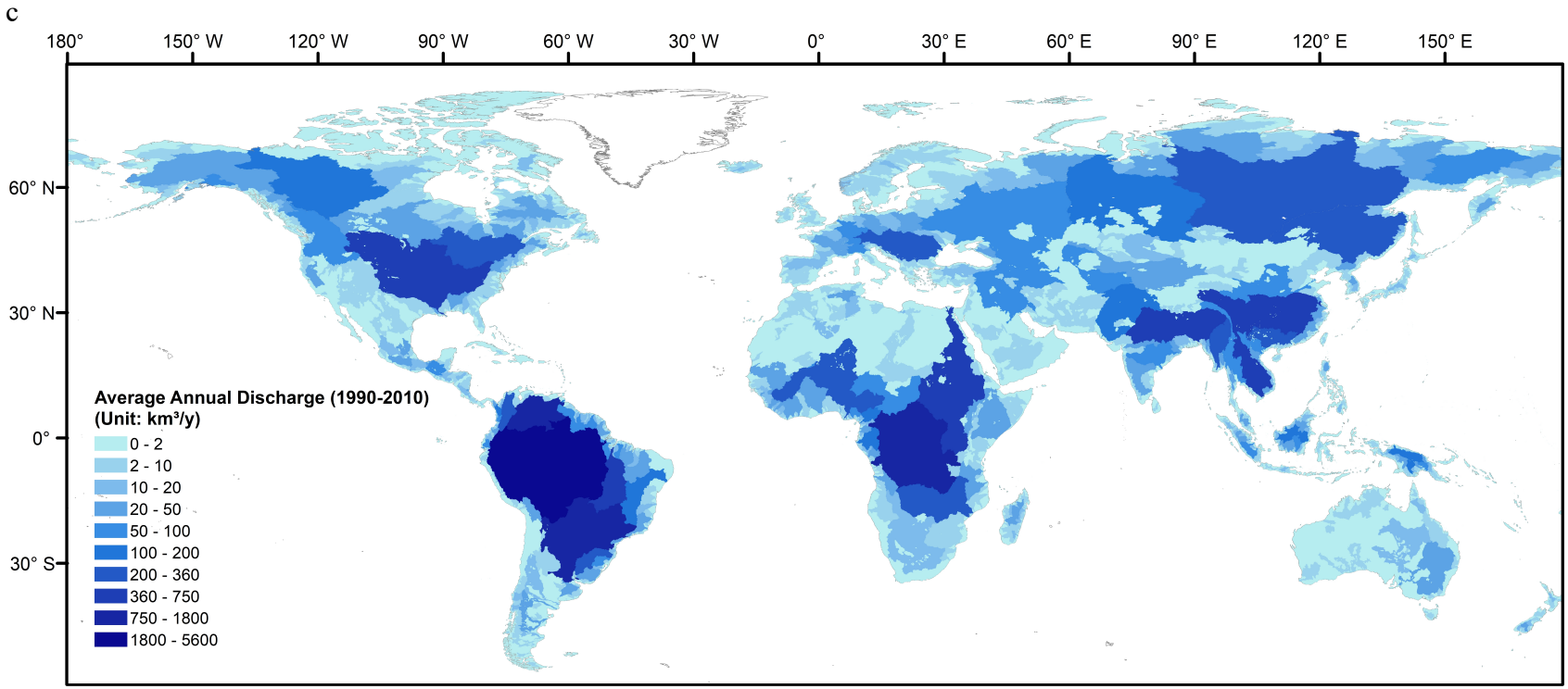


Figure 4-9 Spatial distributions of saline lake storage, freshwater lake storage and average annual discharge in HydroSHEDS basins.

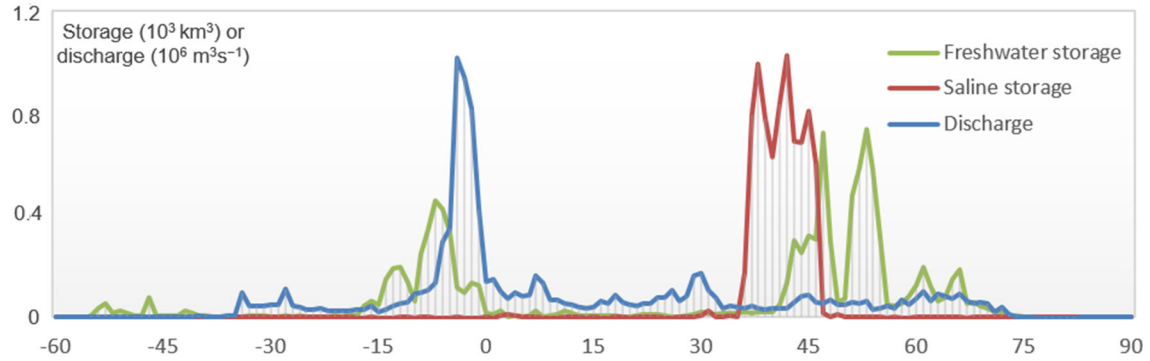
Panels represent the global volume distribution of saline lakes (a), freshwater lakes (b), and global annual average (1990–2010) discharge in HydroSHEDS basins (based on the circa-2000 mapping).

The six maps above display the global distribution patterns of freshwater lakes, saline lakes, and discharge. Saline lake storage is concentrated in several regions. Because of the largest saline lake, Caspian Sea, Central Eurasia has the largest saline lake storage. Oceania is another highly concentrated area of saline lakes because dry area and endorheic basins account for a large proportion. Although the northern part of North America and the northern part of Asia do not have many saline lakes, they also show darker red. Because in those regions, drainage basins are relatively large, darker red in the whole basin does not mean everywhere has high a density of saline lakes. Moreover, some saline lagoons also contribute to the saline storage in those regions. Freshwater lake storage focuses on the northern part of Asia (such as Lake Baikal), the northern part of North America (e.g., the Great Lake and most of Canada), and central Africa (e.g., Lake Victoria and Lake Tanganyika). The regions with large discharge are the northern part of South America (e.g., the Amazon River Basin), the eastern part of Asia (e.g., the Yangtze River Basin) and Central Africa (e.g., the Nile River Basin and the Congo River Basin).

In order to further contrast these three categories in a quantitative way, the following two graphs (**Figure 4-10**) were generated for both directions of latitude and longitude (**Code 8**). Combining with the previous spatial distribution maps (**Figure 4-1**, **4-2**), it is helpful to better explain the distribution of lake water reserves and river flows on a global scale. To be more specific, **Figure 4-10b** shows the volume distribution on the longitude level. There is one peak of saline lake storage at ~50°E and three peaks of freshwater lakes at 30° E, 108.2° E, and 82.5° W. Based on the spatial distribution of freshwater and saline lakes on the circa-2000 mapping, the volume of the Caspian Sea (41.9° N 50.7° E) leads to this peak of saline lakes, and the three peaks of freshwater lakes correspond to the Great Lakes (45.05° N 82.5° W) of North America, the large freshwater lakes in Africa (such as Lake Victoria, Lake Tanganyika, and Lake Malawi), and Lake Baikal (53.4° N 108.2°E) in Russia. Based on the global discharge distribution, the northern region of South America (e.g., Amazon River Basin, central coordinates: 2.15° S 55.1° W), the central Africa (e.g., Congo River Basin, central coordinates: -6° S 12.5° E) and East Asia (e.g., Yangtze River Basin, central coordinates: 31° N 109.5° E) have high average discharge during 1990 to 2010. Therefore, there are also three peaks (55.1° W, 12.5°E, and 109.5° E) for global discharge on the longitude level. For the peaks of the three categories on the latitude scale, the reasons are the same as the longitude scale. From the graphs, I concluded that the freshwater and saline lake

storages are concentrated on the mid-high latitudes in the Northern hemisphere, and the discharge more concentrated on the equatorial regions.

a



b

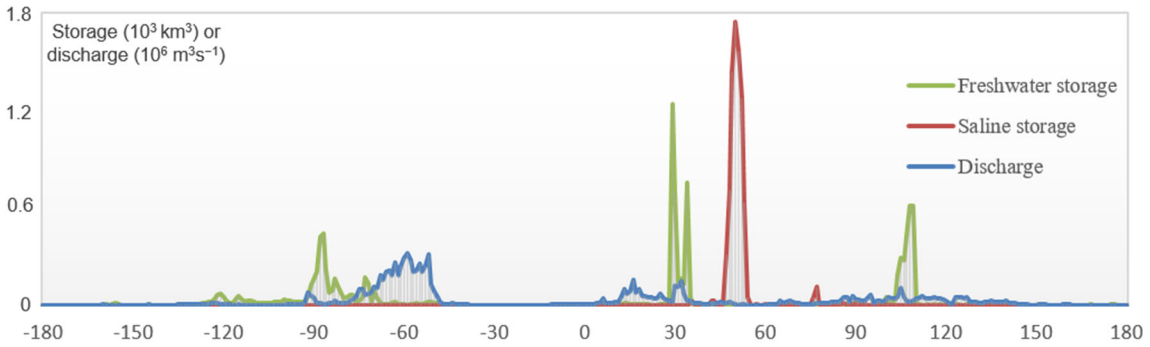


Figure 4-10 Latitudinal and longitudinal distributions of global freshwater lake storage, saline lake storage, and average annual river discharge (1990–2010).
Panels show the spatial distribution of global freshwater and saline lake storage, and average annual discharge (1990–2010) at latitude (a) and longitude (b).

Saline lake storage (red), freshwater lake storage (green) and the average annual discharge are shown in an RGB composite image (**Figure 4-11**).

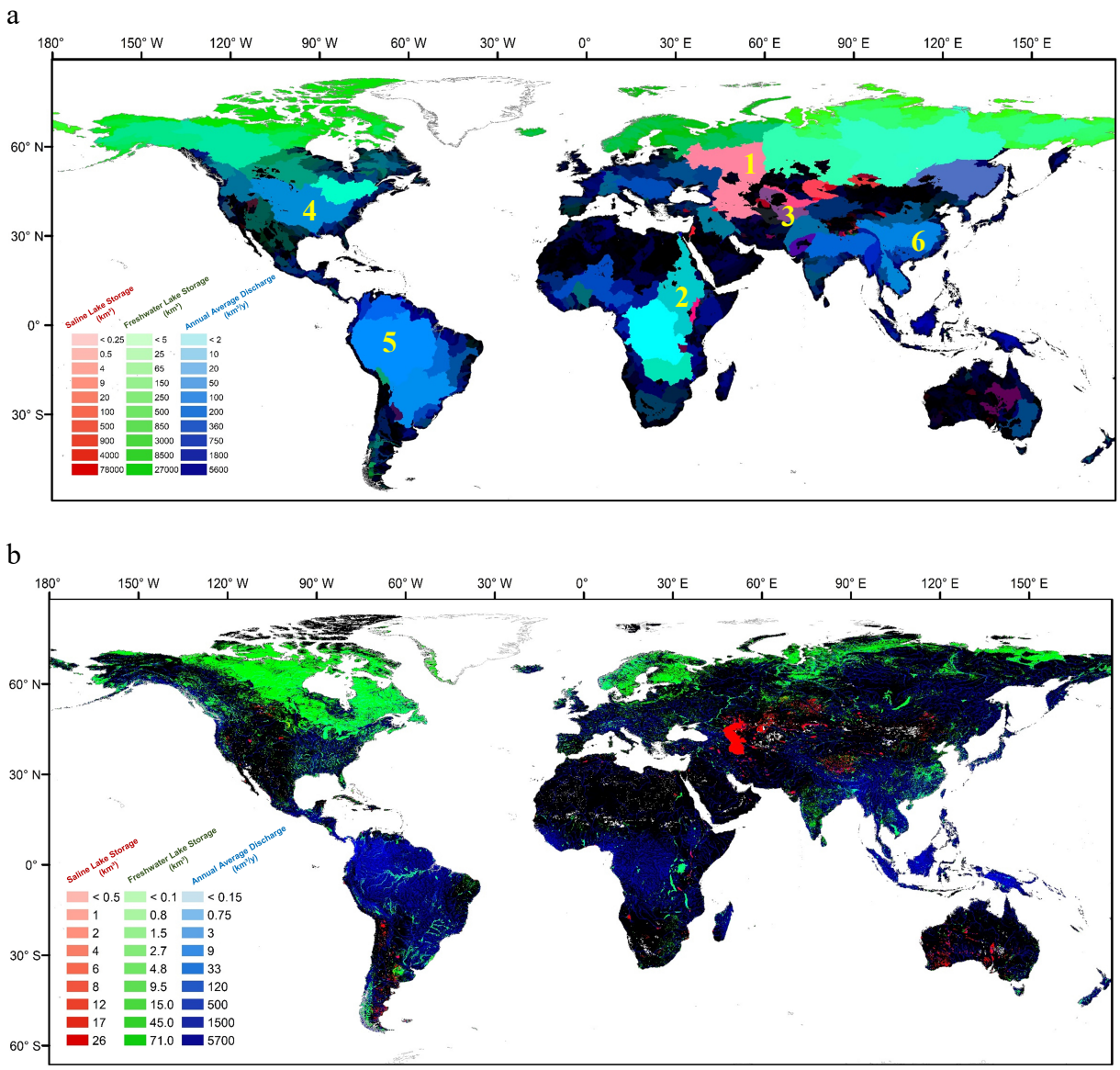


Figure 4-11 RGB composite of saline lake storage, freshwater lake storage, and river discharge.

Panels show the global spatial distribution of freshwater lakes, saline lake, and discharge based in HydroSHEDS basins (a) and 0.1° global grid (b).

From **Figure 4-11**, saline lakes are mainly located in endorheic basins and arid/semi-arid regions. At the high latitude of the northern hemisphere, the dominant color is green (freshwater lake storage) which is consistent with previous figures. River discharge is mainly distributed in the Amazon plain, the southeastern part of China, the southeastern part of Asia and some other humid regions.

In each basin, the saturation of each color indicates the quantity of storage or flow. For example, when both water flow and saline lake storage are large in the same region, this region will be presented by the combination of high blue and high red, which tends to be purple.

In order to explain the results more quantitatively, six basins were selected (**Table 4-6**) to provide the detailed values of their lake storage and river flow. These basins are also marked in **Figure 4-11a** to better illustrate the color composite. The Caspian Sea basin has extreme high saline storage (77,455.91 km^3) but small freshwater storage (301.76 km^3) and discharge (96.34 $km^3\ yr^{-1}$), so it presents reddish/pink which is represented saline storage. For the Nile River basin, several large freshwater lakes lead to high freshwater storage (3,199.71 km^3) and the Nile River leads to a high discharge (366.60 $km^3\ yr^{-1}$). Therefore, the color of the Nile River basin appears cyan which is a composite of green and blue. In the Aral Sea Basin (**Table 4-6**), the magnitudes of river discharge, freshwater storage, and saline water storage are comparable to each other, but due to freshwater lake storage has a much larger scale, the resulting color is purple which is a composite of red and blue. The other three basins, the Mississippi River basin, the Amazon River basin, and the Yangtze River basin have high discharges, but lower freshwater storages and almost zero saline water storage, resulting in saturated blue colors for all of them.

Table 4-6 Saline/freshwater storage and discharge in six large river basins.

	Saline storage (km^3)	***Freshwater storage (km^3)***	***Discharge ($km^3\ yr^{-1}$)***
1. Caspian Sea	77,455.91	301.76	96.34
2. Nile River	0	3,199.71	366.60
3. Aral Sea	27.72	39.93	53.86
4. Mississippi River	0.59	414.98	516.44
5. Amazon River	0.05	441.05	5,505.75
6. Yangtze River	0	293.62	734.90

Chapter 5 - Discussion

As the final result of my research, our understanding of the abundance and the distribution of surface water resources was improved from four aspects (i.e., freshwater/saline distribution, freshwater/saline abundance, global lake storage, the spatial contrast between river flow and lake storage). Two high-resolution global lake datasets have been classified into freshwater and saline categories, and the depth and volume of each inventoried lake have been estimated. The produced global lake inventories have now included comprehensive attributes including lake area, salinity, alkalinity, depth, volume, uncertainties for depth and volume. Moreover, for the comparison between the global annual discharge and total lake storage, two RGB composites have been generated to show the distributions of freshwater storage, saline storage, and river flow together. They also demonstrated the relationship between river flow and storage. These results help us better understand the quantity and quality of global lakes and reservoirs.

Before my research, due to the development of geographic information systems and existing archives, we have a basic understanding of several large lakes. However, our knowledge of global lakes was lacking, such as the global distribution of freshwater and saline lakes. The main goal of my thesis is to further improve our understanding of the abundance of global water bodies. Based on the produced lake inventories in this thesis, we also can continue exploring and studying other aspects to have a more comprehensive understanding of global lakes.

5.1 Comparison with existing studies

The freshwater and saline lake classification helps to better understand the location and abundance of freshwater lakes, but this classification was not considered in the existing global water body datasets. Compared to other studies, another breakthrough of my thesis research is an explicit separation of freshwater lake storages from saline lake storages.

Extending from **Table 2-1**, **Table 5-1** compares the results of my thesis to several major global waterbody datasets that were discussed previously. Both circa-2000 and circa-2015 inventories contain the freshwater/saline classification and lake volume estimation. Circa-2000

and circa-2015 lake mappings documented 8-9 million lakes larger than 0.4 hectares. This spatial resolution is much finer than most existing datasets. The most significant contribution of these two datasets is separating freshwater lakes from those that are saline. Another important progress is that the volume of global freshwater lakes and saline lakes were estimated on both two global waterbody masks.

The result of lake storage estimation shows that there is a difference of ~6,000 km^3 between HydroLAKES and our circa-2000/circa-2015 inventories. This difference stems from the lakes smaller than 0.1 km^2 that was missing from HydroLAKES. Various studies indicate the importance of small lakes in the water cycle (Messager et al., 2016; Cael et al., 2017). For example, small lakes make more contributions to the aquatic-terrestrial interface (Messager et al., 2016). However, HydroLAKES, by documenting water bodies larger than 0.1 km^2, included only part of the global lake system. A total of more than 6 million lakes smaller than 0.1 km^2 were not mapped and their volumes not estimated in this lake inventory. Therefore, the result of lake volume estimation in my thesis is likely more complete and accurate.

Table 5-1 Comparison between major global lake datasets and our circa-2000 and circa-2015 mappings.

Dataset	*Spatial resolution*	*Total number (million)*	*Total area (million km^2)*	*Classification of freshwater and saline lakes*	*Total volume (thousand km^3)*	*Reference*
Global Lakes and Wetlands Dataset (GLWD)	0.1 km^2	1.5	2.7	No	---	Lehner and Döll (2004)
Merge of SRTM Water Body Dataset v2.0 (SWBD 2005) and GTOPO30 Lakes	~0.1 km^2	0.70	2.76	No	---	Slater et al. (2006)
GlObal WAter BOdies Database (GLOWABO)	0.002 km^2	117	5.0	No	199 (±3) (for > 0.01 km^2)	Verpoorter et al. (2014); Cael et al. (2017)
Global Inland Surface Waterbody Dataset (GIW)	~0.001 km^2	---	3.65	No	---	Feng et al. (2016)
HydroLAKES	0.1 km^2	1.42	2.67	No	182.9	Messager et al. (2016)
Circa-2000 Mapping	~0.004 km^2	8.08	3.02	Yes	188.8	Sheng et al. (2015); Sheng et al. (2016); this thesis
Circa-2015 Mapping	~0.004 km^2	9.09	2.97	Yes	190.3	Sheng et al. (2016); this thesis

Note: The dataset list may not be exhaustive, and some numbers are approximations.

5.2 Dataset limitations

5.2.1 Freshwater/saline classification

The first attempt of freshwater and saline classification on global lakes in my thesis identified the specific locations of freshwater lakes and provided guidance for future management of surface freshwater resources.

The entire global lake classification process contains both quantitative and qualitative methods. For example, salt and minerals tend to have high reflectance and are often visible in spectral images as salt crust deposits around saline lakes. During the manually checking of the spectral evidence, we not only focused on the reflectance value (quantitative) but also considered the texture of the high-reflectance objects (qualitative).

The methods I identified saline lakes were overall conservative and generally categorical. A lake would be classified into a saline lake only if it has solid evidence (hydrological, spectral, literature, or combined). Among the saline lakes I identified, many of them were also provided with a numeric salinity value which was collected from the literature. These salinity values were used to verify and assess my classification accuracy (see detail in "Results" chapter, **Table 4-1**). As I described in my thesis, the classification was based on the combination of three kinds of evidence: hydrological, spectral, and literature. If a lake was classified based on the first two evidence types exclusively, there are no salinity or alkalinity values. However, some of the global lakes have salinities if they were classified with consideration of literature or documentation. Although I tried the best to collect as many papers, reports, and documentation as I could, given the sheer number of lakes all over the world, those with salinity documentations are still relatively rare.

During my research, I focused on both circa-2000 and circa-2015 global inventories. For future work, perhaps a most useful dataset will be the recently-updated circa-2015 dataset because it is the latest version and it is more comprehensive on lake mapping. However, the circa-2000 mapping was carefully classified by following each step of the entire classification procedure, and it is valuable for future research as well. For example, by comparing the differences between those two lake inventories, it is feasible to observe some of the trends of the global lake area and volume

over the past 15 years. Furthermore, to compare river flow and lake storage, the circa-2000 waterbody mask is more consistent with the available discharge data.

As another limitation, the lake classification consumed several years to complete. Many team members contributed to this process due to the difficulties of creating/adjusting the classification scheme and the huge amount of data. Therefore, even though there is only one logic flow chart that was followed by everyone, subjective judgments still exist. In order to reduce the errors caused by different decisions on lake classification, I checked the result several times to ensure the accuracy as much as I could.

My classification on the circa-2015 dataset mainly depended on the spatial relationship with the well-classified circa-2000 dataset. Apparently, lake salinity is unlikely to remain unchanged during 15 years. Some of them may have higher or lower salinities on the circa-2015 mapping. I performed manual quality assurance after the automated classification. It is not feasible in time, and likely not necessary, either, to repeat the entire classification based on the logic chart. Therefore, although I have considered as many aspects as possible to assign the classification result from 2000 to 2015, it is inevitable to contain some misclassified lakes in the updated mapping classification.

5.2.2 Lake volume estimations

The volumes of lakes and reservoirs were assigned through both documentation and empirical estimations. Documents from different periods and different authors can produce some inconsistent results. For example, the records in different datasets contain temporal and seasonal variations in lake area, depth and volume, and those variations may generate uncertainties in the final lake volume estimation. In the process of volume estimation, part of the results of HydroLAKES was produced by geostatistical regression, which generated an estimated result, not the actual volume. The uncertainties of volume and depth for each inventoried lake in both circa-2000 and circa-2015 mappings were quantified in my research, and the regression uncertainties also have been considered in my error propagation.

The methods for estimating global lake volumes in Messager et al. (2016) and Cael et al. (2017) are novel and meaningful. I applied the spatial relationship and the area ratio between our mappings and theirs to calculate the lake depth and volume of each inventoried lake. The area ratios were applied for the volume estimation due to the area variation between our global lake inventories and HydroLAKES. Therefore, another novel aspect of my research is the consideration of the area variation, so the uncertainties of mean depth and volume generated by the estimation methods were calculated for each inventoried lake.

5.3 Future research

According to the ratio between water withdrawal and water availability, it is feasible to identify the adequacy of freshwater resources in different regions. Then the areas with severe water stress can be determined as regions that governments and water authorities can take effective measures for. River discharge provides a renewable water resource that can be accessed and utilized by human society (Vörösmarty et al., 2000). However, there are several types of water sources, such as desalinated water, virtual water, groundwater, and other terrestrial water including river discharge and surface storage in lakes and reservoirs. Global lakes and man-made reservoirs are of vital socioeconomic and environmental importance in the contemporary ecosystem. Due to climate change and rapid population growth, a potentially important issue is the water stress of lakes and reservoirs on a global scale. Therefore, the long-term goal of my research is to evaluate the global freshwater lake water stress by combining the produced lake classifications, lake water storage, water flow, and groundwater with human water consumption together. In terms of the data acquisition, total human water consumption from domestic, industrial, irrigation, livestock sectors has been estimated over the period 1990–2010 on a 0.1° global grid (Wada et al., 2013). The dataset for discharge, the main water resource, is also available (Van et al., 2011; Wada et al., 2011a, 2011b, 2014, 2016) (see details in Section 3.1). In my thesis, I completed a global lake inventory that contains the location and the total volume of global freshwater and saline lakes, separately. This is the most vital contribution of my thesis study.

5.4 Broader impacts

The generated circa-2000 and circa-2015 global waterbody datasets with water types and volumes will make a significant contribution to limnology research and support an advanced understanding of the abundance and distribution of surface water resources. For example, in the field of limnology, the storage of global lakes will be helpful to detect the changes of lake water storage in the future. Furthermore, given the location and volume of each freshwater lake, researchers can conduct a series of studies to connect human activities to lake distribution and understand how human activities influence lake storage changes. This research has great social significance. My work also encourages women to pursue remote sensing and geographic information science skills and obtain the research experience.

The educational plan of this project includes creating new datasets that can help students to better understand the fundamental attributes of global lakes. The final classification maps of global freshwater and saline lakes can be potentially added into the textbook of geography classes so that students can identify the location and the comprehensive attributes of each inventoried lake from these maps. They can also help understand the general pattern of the spatial distribution of lake storage and river discharge.

For the governments and research institutes, knowing the locations and the storage of freshwater and saline lakes has a positive impact on future water management and utilization.

Chapter 6 - Conclusions

In general, my research improves our understanding of physical geography of global lakes and reservoirs, and it also provides two high-resolution global mappings with comprehensive attributes, including lake area, freshwater/saline type, depth, volume, and their uncertainties. These two datasets may be applicable to other research related to surface water resource in the future.

6.1 Assessment of water abundance of global lakes and reservoirs

My research contributes to a better understanding of the quantity (area, depth, and volume) and quality (saline or freshwater types) of global open water bodies (lakes and reservoirs). The first three questions emphasize the water abundance of global lakes and reservoirs. To obtain the distribution, quantities, and the total storage of freshwater and saline lakes is the main objective of those questions. In general, among both circa-2000 and circa-2015 global lake mappings, saline lakes account for less than 2% in number, a quarter in area, and about half in volume. Moreover, excluding the Caspian Sea, saline lakes are comparable to reservoirs in water volume.

6.1.1 Distribution of freshwater and saline lakes

From the results of the global lake classification maps (circa-2000 and circa-2015 datasets), a great majority of the global saline lakes are located in endorheic basins and arid/semi-arid regions. Freshwater lakes account for a considerable number in global lakes and most of them are located in exorheic basins, humid regions, and/or pan-Arctic regions.

6.1.2 Abundance of freshwater and saline lakes

I provided an enhanced geography of freshwater/saline lake distributions on two temporal snapshots (circa-2000 and circa-2015 mappings). The circa-2000 global inventory concludes 8.1 million lakes larger than 0.4 hectares with a total area of 3,002 thousand km^2, and saline lakes account for 1.5% in the number of lakes and 22.8% in the area of the global lakes. The circa-2015

lake dataset has 9.1 million lakes with a total area of 2,976 thousand km^2, and saline lakes account for 1.4% in the number and 21.8% in the area.

6.1.3 Storages of freshwater and saline lakes

Global lakes in the circa-2000 and circa-2015 inventories have a total volume of 189 (±6) thousand Gt and 190 (±6) thousand Gt, respectively. Saline lakes account for 43.3% and 42.7% on the circa-2000 and circa-2015 mappings, separately.

6.2 Comparison between water flow and lake water storage

Two main components of surface water that are directly accessible by human societies are river flow and lake water storage. Freshwater lakes are a significant part of the global lake system because it can be utilized as the main freshwater resource and can mitigate the stress of river flow. The classification of freshwater and saline lakes and the estimation of their water volumes allow for a spatially-explicit comparison between surface water flow and freshwater storage.

Both river flow and lake water storage provide essential water resources for human survival. The distribution of river flow and lake storage in different regions are uneven, so water flow and lake storage experience different-level water stress. Comparison between lake storage and water flow reveals the spatial disparity among surface water resources. As the result of the lake water storage and river flow comparison, two-scale datasets (drainage basins and 0.1° geographic grid) that contain freshwater lake storage, saline lake storage and average annual discharge (1990–2010) were generated to evaluate the spatial distribution relationship among three factors. The created RGB composite maps show that river flow and freshwater lakes are the main water sources in humid regions, but the saline lakes are dominant in arid/semi-arid regions. Moreover, high freshwater storage tends to coexist with high river flow, except high-latitude (pan-Arctic) regions (where storage dominates). Saline water storage is spatially less correlated to freshwater storage or discharge, thus suggesting arid/semiarid climates and water-stressed regions. The detailed

spatial cross-tabulations among surface river flow, saline lake storage, and freshwater lake storage provide an important guidance for potential water resource managements in the future.

References

Abbott, B. W., Bishop, K., Zarnetske, J. P., Minaudo, C., Chapin, F. S., Krause, S., Hannah, D. M., Conner, L., Ellison, D., Godsey, S. E., Plont, S., Marçais, J., Kolbe, T., Huebner, A., Frei, R. J., Hampton, T., Gu, S., Buhman, M., Sayedi, S., Ursache, O., Chapin, M., Henderson K. D., & Pinay G. (2019). Human domination of the global water cycle absent from depictions and perceptions. *Nature Geoscience*, 12, 533–540.

Adrian, R., O'Reilly, C. M., Zagarese, H., Baines, S. B., Hessen, D. O., Keller, W., Livingstone, D. M., Sommaruga, R., Straile, D., Donk, E., Weyhenmeyer, G. A., & Winder, M. (2009). Lakes as sentinels of climate change. *Limnology and oceanography*, 54(6part2), 2283–2297.

Armah, F. A., Luginaah, I., Kuitunen, M., & Mkandawire, P. (2012). Ecological health status of the Fosu Lagoon, Southern Ghana II: environmental and human health risk assessment. *Journal of Ecosystem and Ecography*, 2: 107.

Arnell, N. W. (2004). Climate change and global water resources: SRES emissions and socio-economic scenarios. *Global Environmental Change*, 14(1), 31–52.

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438(7066), 303–309.

Birkett, C. M., & Mason, I. M. (1995). A new global lakes database for a remote sensing program studying climatically sensitive large lakes. *Journal of Great Lakes Research*, 21(3), 307–318.

Brisou, J., Courtois, D., & Denis, F. (1974). Microbiology study of a hypersaline lake in French Somaliland. *Applied and Environmental Microbiology*, 27(5), 819–822.

Cael, B. B., Heathcote, A. J., & Seekell, D. A. (2017). The volume and mean depth of Earth's lakes. *Geophysical Research Letters*, 44(1), 209–218.

Carroll, M. L., Townshend, J. R., DiMiceli, C. M., Noojipady, P., & Sohlberg, R. A. (2009). A new global raster water mask at 250 m resolution. *International Journal of Digital Earth*, 2(4), 291–308.

Chahine, M. T. (1992). The hydrological cycle and its influence on climate. *Nature*, 359(6394), 373–380.

Davie, J. C. S., Falloon, P. D., Kahana, R., Dankers, R., Betts, R., Portmann, F. T., Wisser, D., Clark, D. B., Ito, A., Masaki, Y., Nishina, K., Fekete, B., Tessler, Z., Wada, Y., Liu, X., Tang, Q., Hagemann, S., Stacke, T., Pavlick, R., Schaphoff, S., Gosling, S. N., Franssen, W., & Arnell, N. (2013). Comparing projections of future changes in runoff from hydrological and biome models in ISI-MIP. *Earth System Dynamics*, 4(2), 359–374.

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. *Aquatic Procedia*, 4, 1001–1007.

Ding, M., Wang, J., Song, C., Sheng, Y., Urano, T., Bailey, C., Walter, B., and Satori, P.J. (2017). Considering water quality in global lake inventory: freshwater or saline? In *Association of American Geographers*, Boston, Massachusetts, 5–9 April, 2017.

Dodds, P. S., & Rothman, D. H. (2000). Scaling, universality, and geomorphology. *Annual Review of Earth and Planetary Sciences*, 28(1), 571–610.

Donchyts, G., Baart, F., Winsemius, H., Gorelick, N., Kwadijk, J., & Van De Giesen, N. (2016). Earth's surface water change over the past 30 years. *Nature Climate Change*, 6(9), 810–813.

Feng, M., Sexton, J. O., Channan, S., & Townshend, J. R. (2016). A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic–spectral classification algorithm. *International Journal of Digital Earth*, 9(2), 113–133.

Ferguson, R. I. (1986). River loads underestimated by rating curves. *Water Resources Research*, 22(1), 74–76.

Giorgi, F. (2006). Climate change hot-spots. *Geophysical Research Letters*, 33(8), L08707.

Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z. D., Wada, Y., & Wisser, D. (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences*, 111(9), 3251–3256.

Hammer, U. T. (1986). *Saline Lake Ecosystems of the World*. Berlin/Heidelberg, Germany: Springer Science & Business Media.

Herdendorf, C. E. (1990). Distribution of the world's large lakes. *Large Lakes*, 3-38. Berlin/Heidelberg, Germany: Springer

Herdendorf, C. E. (1982). Large lakes of the world. *Journal of Great Lakes Research*, 8(3), 379–412.

Herschy, R. W. (2014). *Streamflow Measurement*. Boca Raton, Florida: Chemical Rubber Company (CRC) Press,

Hollister, J., & Milstead, W. B. (2010). Using GIS to estimate lake volume from limited data. *Lake and Reservoir Management*, 26(3), 194–199.

Jackson, R. B., Carpenter, S. R., Dahm, C. N., McKnight, D. M., Naiman, R. J., Postel, S. L., & Running, S. W. (2001). Water in a changing world. *Ecological Applications*, 11(4), 1027–1045.

Kameyama, S., Zhang, J., Wang, Q., Xu, K., Katoh, T., & Watanabe, M. (2004). An approach to estimate the water level and volume of Dongting Lake using Terra/MODIS data. *Acta Geographica Sinica Chinese Edition*, 59(1), 88–94.

Kashid, S. G., & Pardeshi, S. A. (2014, August). A Survey of Water Distribution System and new approach to intelligent water distribution system. In *2014 First International Conference*

on Networks & Soft Computing (ICNSC2014) (pp. 339-344), Institute of Electrical and Electronics Engineers (IEEE), Seattle, Washington, 8–15 November, 2014.

Kihn, R. G., & Gómez, E. A. (2015). Predation on recent marine and fresh-water ostracod populations of the central region of the republic Argentina. *Munis Entomology & Zoology*, 10, 86–93.

Kjerfve, B., Schettini, C. A. F., Knoppers, B., Lessa, G., & Ferreira, H. O. (1996). Hydrology and salt balance in a large, hypersaline coastal lagoon: Lagoa de Araruama, Brazil. *Estuarine, Coastal and Shelf Science*, 42(6), 701–725.

Korobkov, A. A., Kotseruba, V. V., & Probatova, N. S. (2014). Chromosome numbers of some species of Artemisia L. from Altai region, South Siberia. Botanica Pacifica: *A Journal of Plant Science and Conservation*, 3(1), 61–66.

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, 296(1–4), 1–22.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., & Wisser, D. (2011). High ‐ resolution mapping of the world's reservoirs and dams for sustainable river ‐ flow management. *Frontiers in Ecology and the Environment*, 9(9), 494–502.

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union*, 89(10), 93–94.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., & Wood, E. F. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, doi:10.1029/2019WR025287.

Mendoza, E., Castillo-Rivera, M., Zárate-Hernández, R., & Ortiz-Burgos, S. (2009). Seasonal variations in the diversity, abundance, and composition of species in an estuarine fish community in the Tropical Eastern Pacific, Mexico. *Ichthyological Research*, 56(4), 330–339.

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7, 13603.

Milly, P. C., Dunne, K. A., & Vecchia, A. V. (2005). Global pattern of trends in streamflow and water availability in a changing climate. *Nature*, 438(7066), 347–350.

Misstear, B., D. Banks & L. Clark (2006) Appendix 5 FAO Irrigation water quality guidelines. *Water Wells and Boreholes*, 473–474.

Nijssen, B., Lettenmaier, D. P., Liang, X., Wetzel, S. W., & Wood, E. F. (1997). Streamflow simulation for continental-scale river basins. *Water Resources Research*, 33(4), 711–724.

Oki, T., & Kanae, S. (2006). Global hydrological cycles and world water resources. *Science*, 313(5790), 1068–1072.

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422.

Pérez, L., Lorenschat, J., Massaferro, J., Pailles, C., Sylvestre, F., Hollwedel, W., Brandorff, G., Brenner, M., Islebe, G., Lozano, M. S., Scharf, B., & Schealb, A. (2013). Bioindicators of climate and trophic state in lowland and highland aquatic ecosystems of the Northern Neotropics. *Revista de Biología Tropical*, 61(2), 603–644.

Pokhrel, Y. N., Hanasaki, N., Yeh, P. J., Yamada, T. J., Kanae, S., & Oki, T. (2012). Model estimates of sea-level change due to anthropogenic impacts on terrestrial water storage. *Nature Geoscience*, 5(6), 389–392.

Robinson, N., Regetz, J., & Guralnick, R. P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 57–67.

Rost, S., Gerten, D., Bondeau, A., Lucht, W., Rohwer, J., & Schaphoff, S. (2008). Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research*, 44(9), W09405.

Scheffran, J., & Battaglini, A. (2011). Climate and conflicts: the security risks of global warming. *Regional Environmental Change*, 11(1), 27–39.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., & Kabat, P. (2014). Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences*, 111(9), 3245–3250.

Schlesinger, W. H., & Bernhardt, E. S. (2013). *Biogeochemistry: An Analysis of Global Change*. Cambridge, Massachusetts: Academic Press.

Schmitt, R. W. (1995). The ocean component of the global water cycle. *Reviews of Geophysics*, 33(S2), 1395–1409.

Sheng, Y., Song, C., Wang, J., Garibay, D., Woods, J., Lyons, E. A., & Smith, L. C. (2015). Automated quality assurance and quality control (QA/QC) in developing decadal global lake dynamics products using Landsat-7 and 8. In *American Geophysical Union Fall Meeting*, San Francisco, California, 14–18 December, 2015.

Sheng, Y., Song, C., Wang, J., Lyons, E. A., Knox, B. R., Cox, J. S., & Gao, F. (2016). Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery. *Remote Sensing of Environment*, 185, 129–141.

Sima, S., & Tajrishy, M. (2013). Using satellite data to extract volume–area–elevation relationships for Urmia Lake, Iran. *Journal of Great Lakes Research*, 39(1), 90–99.

Slater, J. A., Garvey, G., Johnston, C., Haase, J., Heady, B., Kroenung, G., & Little, J. (2006). The SRTM data "finishing" process and products. *Photogrammetric Engineering & Remote Sensing*, 72(3), 237–247.

Smith, L. C., Sheng, Y., MacDonald, G. M., & Hinzman, L. D. (2005). Disappearing arctic lakes. *Science*, 308(5727), 1429–1429.

Song, C., Huang, B., Richards, K., Ke, L., & Hien Phan, V. (2014). Accelerated lake expansion on the Tibetan Plateau in the 2000s: Induced by glacial melting or other processes?. *Water Resources Research*, 50(4), 3170–3186.

Song, C., Sheng, Y., Ke, L., Nie, Y., & Wang, J. (2016). Glacial lake evolution in the southeastern Tibetan Plateau and the cause of rapid expansion of proglacial lakes linked to glacial-hydrogeomorphic processes. *Journal of hydrology*, 540, 504–514.

Sood, A., & Smakhtin, V. (2015). Global hydrological models: a review. *Hydrological Sciences Journal*, 60(4), 549–565.

Stricker, V. (1983). *Base Flow of Streams in the Outcrop Area of Southeastern Sand Aquifer, South Carolina, Georgia, Alabama, and Mississippi*. Atlanta, Georgia: US Geological Survey.

SWBD, 2005. Shuttle Radar Topography Mission Water Body Data set (SWBD). Accessed from www2.jpl.nasa.gov/srtm.

Van Beek, L. P. H., Wada, Y., & Bierkens, M. F. (2011). Global monthly water stress: 1. Water balance and water availability. *Water Resources Research*, 47(7), W07517.

Veldkamp, T. I., Wada, Y., de Moel, H., Kummu, M., Eisner, S., Aerts, J. C., & Ward, P. J. (2015). Changing mechanism of global water scarcity events: Impacts of socioeconomic changes and inter-annual hydro-climatic variability. *Global Environmental Change*, 32, 18–29.

Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 41(18), 6396–6402.

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global water resources: vulnerability from climate change and population growth. *Science*, 289(5477), 284–288.

Wada, Y., de Graaf, I. E., & van Beek, L. P. (2016). High-resolution modeling of human and climate impacts on global water resources. *Journal of Advances in Modeling Earth Systems*, 8(2), 735–763.

Wada, Y., Reager, J. T., Chao, B. F., Wang, J., Lo, M. H., Song, C., Li, Y., & Gardner, A. S. (2017). Recent changes in land water storage and its contribution to sea level variations. *Surveys in Geophysics*, 38(1), 131–152.

Wada, Y., Van Beek, L. P. H., & Bierkens, M. F. (2011). Modelling global water stress of the recent past: on the relative importance of trends in water demand and climate variability. *Hydrology and Earth System Sciences*, 15(12), 3785–3805.

Wada, Y., Van Beek, L. P. H., Viviroli, D., Dürr, H. H., Weingartner, R., & Bierkens, M. F. (2011). Global monthly water stress: 2. Water demand and severity of water stress. *Water Resources Research*, 47(7), W07518.

Wada, Y., Van Beek, L. P., Wanders, N., & Bierkens, M. F. (2013). Human water consumption intensifies hydrological drought worldwide. *Environmental Research Letters*, 8(3), 034036.

Wada, Y., Wisser, D., & Bierkens, M. F. (2014). Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dynamics Discussions*, 5(1), 15–40.

Wanders, N., Wada, Y., & Van Lanen, H. A. J. (2015). Global hydrological droughts in the 21st century under a changing hydrological regime. *Earth System Dynamics*, 6(1), 1–15.

Wang, J., & Sheng, Y. (2010). A basin-based analysis of global lake stress from scarcity of sustainable water resource. In *American Geophysical Union Fall Meeting*, San Francisco, California, 13–18 December, 2010.

Wang, J., Sheng, Y., Hinkel, K. M., & Lyons, E. A. (2012). Drained thaw lake basin recovery on the western Arctic Coastal Plain of Alaska using high-resolution digital elevation models and remote sensing imagery. *Remote Sensing of Environment*, 119, 325–336.

Wang, J., Sheng, Y., Song, C., Urano, T., Satori, P. J., & Ford, S. J. (2015). A high-resolution global lake inventory with classified freshwater and saline types. In AGU Fall Meeting Abstracts. In *American Geophysical Union Fall Meeting*, San Francisco, California, 14–18 December, 2015.

Wang, J., Sheng, Y., & Tong, T. S. D. (2014). Monitoring decadal lake dynamics across the Yangtze Basin downstream of Three Gorges Dam. *Remote Sensing of Environment*, 152, 251–269.

Wang, J., Sheng, Y., & Wada, Y. (2017). Little impact of the Three Gorges Dam on recent decadal lake decline across China's Yangtze Plain. *Water Resources Research*, 53(5), 3854–3877.

Wang, J., Song, C., Reager, J. T., Yao, F., Famiglietti, J. S., Sheng, Y., MacDonald, G. M., Brun, F., Schmied, H. M., Marston R. A., & Wada. Y., (2018). Recent global decline in endorheic basin water storages. *Nature Geoscience*, 11(12), 926–932.

Wang, J., Song, C., Sheng, Y., Yao, F., Walter, B., Urano, T., Ding, M., and Bailey, C. (2017). Global surface water storage in the new Millennium: dynamics, implications, and challenges. In *Association of American Geographers*, Boston, Massachusetts, 5–9 April 2017.

Wurtsbaugh, W. A., Miller, C., Null, S. E., DeRose, R. J., Wilcock, P., Hahnenberger, M., Howe, F., & Moore, J. (2017). Decline of the world's saline lakes. *Nature Geoscience*, 10(11), 816–821.

Yamazaki, D., Trigg, M. A., & Ikeshima, D. (2015). Development of a global~ 90 m water body map using multi-temporal Landsat images. *Remote Sensing of Environment*, 171, 337–351.

Yang, K., Yao, F., Wang, J., Luo, J., Shen, Z., Wang, C., & Song, C. (2017). Recent dynamics of alpine lakes on the endorheic Changtang Plateau from multi-mission satellite data. *Journal of Hydrology*, 552, 633–645.

Yao, F., Wang, J., Wang, C., & Crétaux, J. F. (2019). Constructing long-term high-frequency time series of global lake and reservoir areas using Landsat imagery. *Remote Sensing of Environment*, 232, 111210.

Yao, F., Wang, J., Yang, K., Wang, C., Walter, B. A., & Crétaux, J. F. (2018). Lake storage variation on the endorheic Tibetan Plateau and its attribution to climate change since the new millennium. *Environmental Research Letters*, 13(6), 064011.

Zhang, G., Yao, T., Shum, C. K., Yi, S., Yang, K., Xie, H., Feng, W., Bolch, T., Wang, L., Behrangi, A., Zhang, H., Wang, W., Xiang, Y., & Yu, J. (2017). Lake volume and groundwater storage variations in Tibetan Plateau's endorheic basin. *Geophysical Research Letters*, 44(11), 5550–5560.

Appendix A - Summary of Codes

Table A-1 List of codes for each analytical step.

Codes	*Research steps*	*Codes name*	*Purposes*	*Description*
Code 9	3.2.2	*Lake_volume_estimation_2000 (Python)*	Calculate the depth and volume of each inventoried circa-2000 lake.	This script contains four steps based on the analytical methods (Section 3.2.2). The new attributes of the circa-2000 global lake dataset have been updated. (LakeName: the name of the lake, Ave_Dpt: the depth of the lake, Err_Dpt: the uncertainty of the lake depth, Vol: the volume of the lake, Err_Vol: the uncertainty of the lake volume).
Code 10	3.2.3	*Global_lake_classification_2015 (Python)*	Assign the exiting attributes that are included in circa-2000 to the updated circa-2015 global lake mapping.	This script helps me to update the freshwater/saline classification and lake volume estimation on the circa-2015 mapping.
Code 11	3.2.4	*Lake_volume_estimation_2015 (Python)*	The same codes to step 2, only different inputs.	As code 1 but for the circa-2015 inventory.

Code 12	3.2.5	*Annual_average_discharge (IDL)*	Calculate the average annual daily discharge from 1990 to 2010.	This script aims to calculate th average annual discharge whic was compared with lake storage.
Code 13	3.2.5	*Discharge_to_basin (Python)*	Assign the results of the average annual discharge to each 0.1° HydroSHEDSS basin and geographic grid.	A new field ("max") was adde into the global lake dataset. represents the mean value annual discharge from 1990 2010.
Code 14	3.2.5	*Storage_and_flow_comparison_basins (Python)*	Calculate the volume of freshwater and saline lakes in each HydroSHEDSS basin.	This script assigns the informatio related to freshwater lakes ar saline lakes separately. Sever new attributes are added into th HydroSHEDSS basin attribu table. (FreStor: total freshwat lake storage in each drainag basin, SalStor: total saline lak storage in each drainage basi FreArea: total freshwater lake are in each drainage basin, SalAre total saline lake area in eac drainage basin, FreV_Err: tot uncertainty of freshwater lak storage in each drainage basi SalV_Err: total uncertainty saline lake storage in eac drainage basin, Unc_Vol: tot uncertainty of total lake storage each drainage basin).

Code 15	3.2.5	*Storage_and_flow_comparison_grid (Python)*	Calculate the volume of freshwater and saline lakes in each geographic grid cell.	As Code 6 but for 0.1° geograph grid.
Code 16	Figure 4-4	*Area_count_extraction (Python)*	Establish the relationship between lake area and lake count for each continent.	--
Code 17	Figure 4-4	*Area_count_plot (R)*	Plot the relationship between lake area and lake count based on the result of code 8.	--
Code 18	Figure 4-6	*Volume_count_extraction (Python)*	Establish the relationship between lake volume and lake count for each continent.	--
Code 19	Figure 4-6	*Volume_Count_plot (R)*	Plot the relationship between lake volume and lake amount based on the result of code 10.	--
Code 20	Figure 4-10	*Storage_Lon_Lat (Python)*	Create two figures that show the distribution of freshwater lake storage, saline lake storage and average annual discharge in latitude and longitude directions.	--