**University of
Zurich^{UZH}**

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

Glacial lake depth and volume estimation based on a large bathymetric dataset from the Cordillera Blanca, Peru

Muñoz, Randy ; Huggel, Christian ; Frey, Holger ; Cochachin, Alejo ; Haeberli, Wilfried

Abstract: Glacial lakes are most often located in remote places making it difficult to carry out detailed bathymetric surveys. Consequently, lake depths and volumes for unmeasured lakes are often estimated using empirical relationships developed mainly from small bathymetric datasets. In this study, we use the bathymetry dataset of the Cordillera Blanca, Peru comprising 121 detailed lake bathymetries, the most extensive dataset in the world. We assess the performance of the most commonly applied empirical relationships for lake mean depth and volume estimation, but also investigate relationships between different geometric lake variables. We find that lake volume estimation performs better when derived from lake mean depth, which in turn is estimated from lake width. The findings also reveal the extreme variability of lake geometry, which depends on glacio-geomorphological processes that empirical– statistical relationships cannot adequately represent. Such relationships involve characteristic uncertainty ranges of roughly $\pm 50\%$. We also estimate potential peak discharges of outburst floods from these lakes by applying empirical relationships from the literature, which results in discharges varying by up to one-order of magnitude. Finally, the results are applied to the 860 lakes without bathymetric measurements from the inventory dataset of the Cordillera Blanca to estimate lake mean depth, volume and possible peak discharge for all unmeasured lakes. Estimations show that ca. 70% (610) of the lakes have a mean depth lower than 10m and very few longer than 40m. Lake volume of unmeasured lakes represent ca. 32% ($5.18 \times 10^8 \text{ m}^3$) of the total lake volume ($1.15 \times 10^9 \text{ m}^3$) in the Cordillera Blanca. Approximately, 50% of the lakes have potential peak discharges $> 1000 \text{ m}^3/\text{s}$ in case of lake outburst floods, implying a need for additional studies for risk assessment.

DOI: <https://doi.org/10.1002/esp.4826>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-186755>

Journal Article

Accepted Version

Originally published at:

Muñoz, Randy; Huggel, Christian; Frey, Holger; Cochachin, Alejo; Haeberli, Wilfried (2020). Glacial lake depth and volume estimation based on a large bathymetric dataset from the Cordillera Blanca, Peru. *Earth Surface Processes and Landforms*, 45(7):1510-1527.

DOI: <https://doi.org/10.1002/esp.4826>

Glacial lake depth and volume estimation based on a large bathymetric dataset from the Cordillera Blanca, Peru

Randy Muñoz^{1*}, Christian Huggel¹, Holger Frey¹, Alejo Cochachin², Wilfried Haeberli¹

¹ Glaciology and Geomorphodynamics Group, Department of Geography, University of Zurich, Zurich
² Área de Evaluación de Glaciares y Lagunas, Autoridad Nacional del Agua (Peru)

*Corresponding author

Randy Muñoz, Glaciology and Geomorphodynamics Group, Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland
randy.munoz@geo.uzh.ch
044 63 55194

Running head: Glacial lake mean depth and volume estimation

Abstract

Glacial lakes are most often located in remote places making it difficult to carry out detailed bathymetric surveys. Consequently, lake depths and volumes for unmeasured lakes are often estimated using empirical relationships developed mainly from small bathymetric datasets. In this study, we use the bathymetry dataset of the Cordillera Blanca, Peru comprising 121 detailed lake bathymetries, the most extensive dataset in the world. We assess the performance of the most commonly applied empirical relationships for lake mean depth and volume estimation, but also investigate relationships between different geometric lake variables. We find that lake volume estimation performs better when derived from lake mean depth, which in turn is estimated from lake width. The findings also reveal the extreme variability of lake geometry, which depends on glacio-geomorphological processes that empirical-

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/esp.4826

statistical relationships cannot adequately represent. Such relationships involve characteristic uncertainty ranges of roughly $\pm 50\%$. We also estimate potential peak discharges of outburst floods from these lakes by applying empirical relationships from the literature, which results in discharges varying by up to one order magnitude. Finally, the results are applied to the 860 lakes without bathymetric measurements from the inventory dataset of the Cordillera Blanca to estimate lake mean depth, volume and possible peak discharge for all unmeasured lakes. Estimations show that ca. 70% (610) of the lakes have a mean depth lower than 10 m and very few longer than 40 m. Lake volume of unmeasured lakes represent ca. 32% ($5.18 \times 10^8 \text{ m}^3$) of the total lake volume ($1.15 \times 10^9 \text{ m}^3$) in the Cordillera Blanca. Approximately, 50% of the lakes have potential peak discharges $> 1000 \text{ m}^3/\text{s}$ in case of lake outburst floods, implying a need for additional studies for risk assessment.

Keywords

Lake volume, lake mean depth, empirical relationship, bathymetry, Cordillera Blanca, water storage, peak discharge

1 Introduction

High mountain landscapes, including lakes and glaciers, are rapidly changing as a result of climate change (Haeberli et al. 2013; Milner et al. 2017). As glaciers retreat, new lakes can form or existing ones increase in size and volume (Carrivick and Tweed 2013; Cook et al. 2016; Drenkhan et al. 2018). In Peru, the glaciers shrank by 42.6% in area during the last 40 years, resulting in the formation of 996 new glacial lakes (ANA 2014a, b). Additionally, an ice-thickness modeling study identified more than 200 sites with overdeepened depressions in the bed topographies of current Peruvian glaciers, where the formation of new glacial lakes is possible in the course of continued

glacier retreat, most of them likely forming within the coming decades (Colonia et al. 2017).

In the Cordillera Blanca, Peru, some glacial lakes have been used to store water from melting glaciers during the wet season, particularly for hydropower production.

Recently, the interest to implement new infrastructure to use other glacial lakes for water storage is increasing in response to the rising water demand but also for the expected changes in the water supply (Drenkhan et al. 2019a). For water management, the greater the lake volume is, the greater the potential water storage.

Despite their benefits and beauty, lakes in the Cordillera Blanca have caused some of the world's deadliest glacier disasters. In 1941, a glacial lake outburst flood (GLOF) from the Palcacocha Lake killed around 1800 people, destroying large parts of the city of Huaraz (Wegner 2014). Other events with different levels of damage to livelihoods, infrastructure and people occurred in the following decades in different catchments, one of the latest in 2012 in the Santa Cruz Valley (Hubbard et al. 2005; Carey 2010; Frey et al. 2014; Schneider et al. 2014; Mergili et al. 2018). Currently, 40 large lakes in the Cordillera Blanca have different levels of susceptibility to outburst floods (Mergili et al. 2016). For risk assessment, the greater the lake volume is the greater the volume of water available for potential flooding, although the flood volume depends on the cause, mechanism of the processes and the physical characteristics of the lake and dam (Emmer and Vilímek 2014).

Hence, both water storage and potential outburst floods are strongly linked with lake volume, which in turn depends on glacier dynamics and complex geomorphological processes in surrounding landscapes (Cook and Swift 2012; Carrivick and Tweed 2013). For water management lakes with large volumes can be directly identified as

potential places for water storage. For risk assessment, lake volume is not directly an indicator of hazard, therefore other indicators have been used like the potential peak discharge (Haeberli 1983; Bajracharya et al. 2007; Klimeš et al. 2014), which is often estimated through empirical relationships using lake volume as input data (Haeberli 1983; Popov 1991; Huggel et al. 2002).

Glacial lakes are mostly located in remote places, with difficult access, limiting their investigation mainly to remote sensing (Cook and Quincey 2015). Hence, most existing inventories of glacial lakes only include basic characteristics related to the surface such as area, length, and width. Bathymetric information is not always possible to obtain and, therefore, related characteristics like mean lake depth and lake volume are usually estimated.

Several approaches have been developed in the literature for mean lake depth and lake volume estimation (O'Connor et al. 2001; Huggel et al. 2002; Yao et al. 2012; Cook and Quincey 2015; Kapitsa et al. 2017). One approach that is frequently used is an empirical relationship between lake volume and area (V-A approach). A critical issue thereby is the self-correlation between a mathematical product (lake volume = lake depth multiplied by lake area) and one of the variables (lake area) from which this product has been derived. Such self-correlations produce misleading artifacts (unrealistically high coefficients of determination, R^2), which hide the extreme variability of lake geometries in nature. It is therefore adequate to instead estimate mean lake depth as an independent variable from area and only then calculate lake volume by multiplying mean depth with area (D-A approach) (e.g., Haeberli et al. 2016b).

Another issue is the size of the dataset. Most of the existing studies are based on relatively small datasets, around 7 to 30 lakes, from different regions of the world and different dam types. To overcome this issue, Cook & Quincey (2015) compiled 42 different bathymetric measurements from existing references and then assessed the performance of three empirical relations by Huggel et al. (2002), O'Connor (2001) and Evans (1986). They concluded that these empirical relationships are not sufficiently accurate to estimate lake volumes although it is expected that the performance increases when lakes with similar characteristics are grouped together. They also suggested the need for including more bathymetries in order to improve these relationships.

In Peru, since 1941, collaborators of the Office for Glacier and Lake Evaluations - AEGL (former Glaciology and Water Resources Unit - UGRH) have been responsible for the monitoring of glaciers and glacial lakes in the high mountain ranges of the country. They have repeatedly compiled inventories of glaciers and glacial lakes of the Peruvian cordilleras. They have also conducted a series of bathymetric surveys in selected lakes. The first one was Palcacocha Lake, just after the GLOF in 1941. Later, other lakes were selected according to their assumed degree of hazard/risk and to their potential for energy production or water storage. Repeated surveys focused on particularly critical lakes like Palcacocha or Lake 513 for the (re)assessment of their hazard and risk situations. Until today, 190 different glacial lakes in Peru have detailed bathymetric studies, 121 of which are located in the Cordillera Blanca.

Here, we present the detailed and large bathymetry dataset of the Cordillera Blanca obtained between 2005 and 2015, which makes this the most extensive dataset of bathymetric information on glacial lakes of a single mountain range in the world. We use this large dataset to explore new relationships for lake mean depth and volume

estimation that avoid self-correlation, as far as possible, or could provide better estimates than the V-A approach. To do so, we compare the performance of different relationships both from literature and those developed from the dataset in this study. For water management, lake volume is directly used as an indicator of potential water storage; while for risk assessment potential peak discharge derived from lake volume is used to identify possible dangerous lakes. Finally, we apply the findings to estimate the lake mean depth, volume and potential peak discharge for all the 860 glacial lakes in Cordillera Blanca (ANA 2014b), including related uncertainties.

2 Study region

The Cordillera Blanca contains the most extensive area of tropical glaciation in the world. Located in the north of the Peruvian Andes, it has a 120 km extension from north-northwest to south-southeast, comprising 40% (448.81 km²) of the Peruvian glacier extent (INAIGEM 2018). A recent glacier inventory indicates a loss of 38.2% from 726.26 km² in 1962 to 448.81 km² in 2016 (INAIGEM 2018).

The lake inventory of the Cordillera Blanca (ANA 2014b) covers a region larger than the current glacier zone. As seen in Figure 1, lakes with and without bathymetric studies are also located in the northern and eastern regions of the Cordillera Blanca. These regions include lakes not in contact with, or in the proximity of, contemporary glacier extents, but were formed by glacial processes. On the one side, lakes close to glaciers and non-consolidated areas could be exposed to rock or ice avalanches that can infill the lakes with sediment changing the lake geometry. For instance, in April 2002 a landslide at Safuna Alta lake deposited about 5 million m³ at the lake bottom reducing lake depth from 120 m to 80 m (Hubbard et al. 2005). On the other side, lakes far from contemporary glaciers formed as consequence of glacier retreat since

the Little Ice Age may not be exposed to large glacier landslides, but depending on their age could be exposed to decades or centuries of other sedimentation processes.

In addition, several studies have indicated that erosion rates in de-glaciating high mountain catchments tend to increase in a climate change context (Hinderer 2001; Fischer et al. 2013; Haeberli et al. 2016d), influencing the lake geometry.

Based on the expected rise of the freezing-level height (FLH) due to climate change, Schauwecker et al. (2017) showed that 68% (99%) of the current glacier area will be lost in the Cordillera Blanca under RCP2.6 (RCP 8.5) until the end of the 21st century.

In the course of this glacier retreat, new lakes are likely to form where overdeepened parts of the glacier beds become exposed due to glacier retreat (Frey et al. 2010; Haeberli et al. 2016a). As result, Colonia et al. (2017) have identified 31 sites with overdeepenings in the current glacier beds of the Cordillera Blanca, where new lakes will potentially form with a total volume of $6.34 \times 10^7 \text{ m}^3$.

3 Data and methods

In this paper, we use two lake datasets from the Cordillera Blanca. The first one is the bathymetry dataset with 121 detailed bathymetric surveys. The second dataset is the lake inventory from satellite imagery with 860 lakes (ANA 2014b) including the 121 lakes from the bathymetry dataset

3.1 The lake bathymetry and inventory datasets

Bathymetric surveys took place between 2005 and 2015 and were developed in three steps: topographic survey, lake depth measurement, and data processing. The topographic survey started with the identification of points from where the topographic

station could overlook most of the lake to avoid moving the topographic base station.

At some lakes like Palcacocha with repeated bathymetries, a concrete benchmark was installed using differential GPS. From these points, the lake perimeter was delineated, the slopes surrounding the lake mapped and any present infrastructure documented.

The scale of the topographic surveys vary within the same bathymetric survey. As a rule, the infrastructure and dam have a more detailed survey (1:500) than the hillside (1:2000) and potentially unstable glacier tongues and up-valley slopes (1:10000).

Measurements of lake depth were collected using an echo sounder (with an integrated GPS, +/- 5 m) attached to a boat. The boat crossed the lakes longitudinally and transversely every 10 or 30 meters, depending on the size of the lake (on smaller lakes shorter distances were used). The location east and north of every measurement point was stored in the memory of the echo sounder and corrected with data from the topographic station during the data processing. This resulted in the collection of about 200 to 600 individual depth measurements for every bathymetric survey.

The data processing consisted of the drawing of the bathymetric map with the corresponding depth isolines from the bathymetric survey using ArcGIS and the topo to raster interpolation method.

The bathymetry dataset of the Cordillera Blanca includes administrative and topographic locations. Table 1 provides an overview of the available information using Palcacocha Lake as example (see supplementary material S1 for complete dataset and description), while Figure 2 outlines the geometric variables.

Locations east and north are given in projected coordinates WGS 1984 – 18S and correspond to the centroid of the lake. Altitude refers to the altitude of the water level when it was measured. Date indicates when the bathymetry survey was carried out.

Maximum length (MxL) is the maximum distance between two points on the axis of the lake using the lake outlet and the direction of the valley. Maximum width (MxW) is the maximum distance between two points perpendicular to the MxL. Both definitions are specific to the Peruvian inventory dataset, since width is usually defined as the longest possible axis in a lake, and width is the longest axis perpendicular to the length. Therefore, in the Peruvian dataset, lake width could be longer than length. Maximum depth (MxD) is the maximum measured depth value. Mean depth (MD) is calculated from the 3D model using ArcGIS. Maximum and average front slope (MFS, AFS), and maximum and average back slope (MBS, ABS), are obtained from the longitudinal profile. Front refers to the side next to the lake outlet and back refers to the opposite, in both cases the average slope is measured from the top of the water level to the maximum depth (see Figure 2). Dam material could be moraine, rock or a moraine-rock combination. Dam type could be with or without infrastructure (Table 2) to regulate the water level and volume, most of them for safety reasons (Portocarrero 2014). Dam material and type will be used later to identify lake groups for further analysis. Finally, volume values were bathymetrically obtained.

The second dataset used here is the Lake Inventory of the Cordillera Blanca with 860 different lakes, including the 121 lakes with bathymetric studies (ANA 2014b). The inventory includes the same variables as the bathymetric dataset, except for the variables that directly depend on the bathymetric survey such as MxD, MD, MFS, AFS,

MBS, ABS, volume, dam material, and dam type. Lakes from the Cordillera Blanca were digitized from satellite images obtained by the Spot 4 and Aster sensors, with 10 and 15 m spatial resolution, respectively. Every image corresponds to the dry season (June - September) between 2001 and 2006. Base cartography is from the Peruvian National Map at a 1:100 000 scale from the Instituto Geográfico Nacional. This was used to generate the digital elevation model and the corresponding watershed delimitation according to the Pfafstetter method (ANA 2012). Satellite images were processed through radiometric and geometric correction, where the latter used the cartographic base meanwhile the radiometric correction was based on the methodology described by Milder (2008). The Normalized Difference Water Index NDWI (Gao 1996) was applied to identify the lakes. Post-processing consisted of the elimination of effects caused by shadow and vegetation. The multitemporal analysis was carried out to differentiate between temporary and permanent lakes. The inventory consider only permanent lakes larger than 5000 m² (ANA 2010, 2014b).

3.2 Data analysis

Lakes from the large bathymetry dataset were classified into different sub datasets (Figure 3), which were used to assess mean depths and volumes: 1) the entire bathymetry dataset (EBDS), 2) the moraine-dammed bathymetry dataset without infrastructure (MBDS) and 3) the rock dam bathymetry dataset (RBDS). Infrastructure includes tunnels, syphons or any other measures implemented mainly for safety reasons.

The EBDS is composed of all the 121 different bathymetries and incorporates all the lake variability, like dam material and the presence or absence of infrastructure. The MBDS is composed of 69 moraine-dam lakes excluding those with infrastructure because such infrastructure in cases modifies lake depths and volumes, possibly

affecting the results of the statistical analysis. The RBDS is composed of 12 different lakes with or without infrastructure.

During the assessment, a fourth lake group from the MBDS was identified: the moraine dataset with mean lake depths exceeding 10 m (MBDS10) composed of 42 lakes. The value of 10 m represents a threshold above which the coefficient of determination R^2 between mean depth with lake area increases. We used this approach to try to avoid, as far as possible, a statistical bias due to large-area lakes with shallow depth and relatively small volumes.

Initially, the lake inventory and bathymetry datasets were compared through the use of descriptive statistics, box-plots and histograms. Later we focused on the depth and volume estimation process for each of the lake groups that we have defined. For this, we divided the whole procedure into four stages: I) correlation analysis, II) mean depth estimation, III) lake volume estimation and, IV) peak discharge estimation (Figure 4). Stages I, II and III were applied to every lake group EBDS, MBDS, RBDS, and MBDS10, while stage IV was applied only to EBDS. Lastly, findings were used to estimate mean depth, lake volume and peak discharge for every lake in the inventory dataset of the Cordillera Blanca.

In stage I, we assessed the correlation between mean lake depth and area, length, width, elongation, MFS (maximum front slope), AFS (average front slope), MBS (maximum back slope), or ABS (average back slope) through scatter plots. Later we compared simple linear, power, or exponential relationships between mean depth with length or width. Multiple linear relationships were also assessed between mean depth

and area, length and/or width to evaluate if more complex relationships could perform better estimations. The identification of the best relationship was based on the highest value of the coefficient of determination R^2 , and the lowest value of the mean squared error (MSE).

In stage II, we estimated the mean depth applying the best relationship from stage I and some of the most commonly applied area-relationships from the literature, which are listed in Table 3. Since only Wang et al. (2012) and Huggel et al. (2002) provided a separate relationship for calculating mean depth from lake area, we converted the originally published volume-area self-relations to the format $D = a \cdot A^b$ (where D: mean lake depth, A: lake area, a and b: constants) by dividing reported volumes by the corresponding lake areas. Resulting mean depth estimates were compared with measurements from the bathymetry dataset through consideration of R^2 values and an error function from Huggel et al. (2002). The latter is the difference between measurements and estimates divided by estimates. The error function generates positive or negative values depending on whether lake depth is over- or underestimated in comparison to the measurements. To allow for comparison among different relationships, we calculated the total error as the absolute average of every single error. Then we selected the best mean depth relationships, meaning higher R^2 and lower total error values.

In stage III, we estimated lake volumes by applying the existing area-relationships in Table 3 and, by multiplying reported lake area with estimated mean depth ($V=D \cdot A$). Here, mean depths are the best estimates from stage II. Then, we assessed the performance, through the R^2 and total errors values, comparing the estimated volumes

with volumes reported in the bathymetry dataset for every lake group. Additionally, we assessed the influence of the distance between lakes and the current glacier outlines through three distance thresholds: lakes up to 1 km, 5 km and 10 km away from the current glacier margins. For this assessment, we applied only the best relationships for mean depth and volume estimation.

In stage IV we estimated the peak discharge of potential lake outburst floods applying three relationships from the literature, which use volume as the independent variable (Table 4). The relationships were derived for failure of moraine or ice-dammed lakes (mechanical failure of ice dams), known to have the highest peak discharge (Huggel et al. 2004). The relationship proposed by Haeberli (1983) represents an upper-bound estimate of peak discharge, while relationships from Huggel et al. (2002) and Popov (1991) are based on regressions of available data and, hence, consider mean rather than extreme values for given lake volumes. Since in the Cordillera Blanca there are no independent data to compare, it is not possible to assess the performance, but it is possible to evaluate the range of estimates. Here we used all the bathymetrically derived lake volumes from the bathymetry dataset.

4 Results

Since the inventory dataset includes all the lakes from the bathymetry dataset, it is worth assessing if the latter could be considered as a sample for the inventory dataset. That is why we started by comparing the inventory and bathymetry datasets through basic statistics (Table 5), box-plots (Figure 5) and histograms (Figure 6). The bathymetry dataset includes some of the largest lakes in the Cordillera Blanca because the aim of the bathymetric surveys was to monitor large lakes with large

expected volumes, although this expectation was not always met. The total area of all lakes of the bathymetry dataset is 29.3 km², ca. half of the total area for the inventory dataset (60.4 km²).

In Figure 5, the length of the whisker (box) groups 95% (50%) of the data points, while every point outside this region is considered an outlier. Many of the statistically extreme data points of the inventory dataset (above the 95th percentile) plot in the 25-75th percentile of the bathymetry dataset. Therefore, the bathymetry dataset would be more representative for lakes with large area than a sample for the entire bathymetry dataset. Box-plots also show that a majority of lakes range within a relatively narrow array, but outliers reflect what is widely observed in nature. Additionally, most of the ratio width/length values are lower than one (<1), showing that most of the lakes are elongated in the direction of the valley.

Histograms in Figure 6 show that most of the lakes in the Cordillera Blanca are smaller than 0.2 km². However, several lakes are one order of magnitude larger than these smaller lakes. Most of the lakes have a width smaller than 0.2 km and a length smaller than 0.5 km. Most bathymetries reported volume values lower than 1.5×10^7 m³ with a maximum depth between 10 and 50 m. The mean depth for most of the lakes is below 20 m.

4.1 Correlation analysis

In stage I, we assessed the correlation between lake mean depth and other lake geometric variables. Here we used the bathymetry dataset and the related lake groups EBDS, RBDS, MBDS, and MBDS10. Scatter plots, like Figure 7 for the EBDS, show no strong correlation between mean lake depth with any other geometric lake variable. This reveals the high heterogeneity of the lakes in nature, even for the same region like the Cordillera Blanca. The same heterogeneity is observed in the other lake groups RBDS, MBDS, and MBDS10 (Table 6 and supplementary material S2).

Simple linear, power and exponential relationships between mean depth with length or width are shown in Figure 8 for every lake group. We found that linear relationships have a slightly better performance than more complex relationships but the differences are small. Coefficients of determination (R^2) are mostly lower than 0.5 and tend to increase with decreasing number of assessed bathymetries. The maximum R^2 values across all investigated relationships between lake depth and other variables is as low as 0.36. In the homogeneous lake groups MBDS (69 bathymetries) and MBDS10 (42 bathymetries) R^2 values are as low as 0.34 and 0.48 respectively, while for the RBDS with 12 bathymetries R^2 is 0.73. This shows that the many small (shallow) lakes cause the large scatter and significantly lower the correlation. In most cases, the highest, but still weak correlations, are observed when using lake width as an independent variable through simple linear relationship, however, differences are mostly small.

To identify which relationship performs better to estimate mean depth, we summarized the R^2 and MSE values in Table 6 for every lake group, comparing linear and multiple

relationships. For simple linear relationships, lake width as independent variable generates the highest R^2 and, at the same time, has the lowest MSE for every lake group. Lake area and length rank second and third, respectively, with small differences. Bed-slope related parameters (AFS, MFS, ABS or MBS) all have weak correlations with mean depth for every lake group ($R^2 < 0.3$). For multiple linear relationships, the best result occurs when using both lake area and width as independent variables but only in the EBDS, in other lake groups results show no clear trend. By far the strongest correlation ($R^2 = 0.8$) is found between mean depth and lake width for rock-dammed lakes (RBDS). However, this result could be influenced by the small sample size.

4.2 Lake mean depth estimation

In stage II, we estimated mean depth applying relationships from (i) the literature in Table 3 and (ii) the linear regression between mean depth and width (denoted by Md_Wi) from stage I. Then we compared these estimations with mean depth determined from the bathymetry dataset to identify which relationships perform best for each lake group. The performance was assessed through the R^2 and total error values, summarized in Table 7.

Coefficients of determination R^2 are lower than 0.5 in most cases, except when assessing the RBDS lake group (up to 0.73). Once again, correlation decreases as the number of assessed bathymetries increases. Total errors are around 50% meaning that the mean depth estimation uncertainty is $\pm 50\%$ average; even in the smallest lake group RBDS with the highest R^2 (0.73) the error is 40%. The lowest total error values are mainly related to the most homogeneous lake group, MBDS10.

When considering both higher R^2 and lower TE, the best estimates result when applying the linear regression between mean depth and width (relationship Md_Wi). Additionally, relationships such as Huggel et al. (2002) and Wang et al. (2012) seem to provide better results than other relationships, however due the small differences could be also considered in line with the others

The box-plots in Figure 9 allow comparison between the measured and estimated mean depth for each lake group. The estimations tend to compress the results around the mean values, which can be seen both from the boxes (50% of the data) and from the whiskers (95% of the data).

4.3 Lake volume estimation

In stage III, we estimated lake volumes using relationships from the literature and from this study, and compared them with lake volumes reported in the bathymetry dataset in order to identify what relationships perform better. The performance was assessed again through the R^2 and total errors values, summarized in Table 8. Lake volumes were estimated by applying both the existing area-relationships in Table 3 and by multiplying reported lake area with estimated mean depth ($V=D \cdot A$). For mean depth, we used values previously estimated in stage II by applying the linear relationship Md_Wi.

All coefficients of determination (R^2) in Table 8 are around 0.85, however, for relationships from the literature, these high values are artifacts caused by the area-

volume self-correlation. Lake volume estimated as $V = A \cdot Md_{Wi}$ provides also high R^2 values nonetheless mean depth is estimated from lake width avoiding self-correlation. Despite this, the total error values are high, all around 50%, implying an uncertainty of about $\pm 50\%$ average. When taking into account both R^2 and TE, lake volume estimated from $V = A \cdot Md_{Wi}$ performs better for lake groups EBDS, RBDS, and MBDS10. The latter, the most homogeneous lake group, in particular presents the lowest total error value (26%). It is notable that the relationship proposed by Wang et al. (2012) provides by far the largest total error values. Finally, lake volume measurements and estimations were plotted as box-plots in Figure 10. Results show that estimated volumes barely catch the highest values but tend to represent the mean values reasonably well.

We also assessed the influence of distance between lakes and current glacier margins for mean depth and volume estimations. Three distance thresholds were applied: 1 km, 5 km and 10 km away from current glaciers. For mean depth estimations we applied the relationship from Wang et al. (2012), Huggel et al. (2002) and the linear relationships using lake width as the independent variable. Linear relationships were defined by using only the lakes for each threshold and lake group. For volume estimation, we applied the relationship from Huggel et al. (2002) and the estimated lake mean depth with lake area.

Results show that for every threshold and lake group mean depth estimated with linear relationship provides higher R^2 and lower TE% than relationships from literature, except for MBDS with lakes in up to 10 km distance from the glacier (see supplementary material S3). In addition, in most of the lake groups, mean depth

estimations by linear relationship provide higher R^2 and lower TE% values than estimations without distance thresholds (Table 9). However, volume estimations do not improve in the same way; in most of cases TE% values reduce but also R^2 values. Both, lake mean depth and volume estimations, improve only for the MBDS10 (1 km and 10 km distance thresholds) and the EBDS (10 km distance threshold). In the latter, the improvement could be considered negligible as R^2 and TE% values are very close to those obtained without applying distance thresholds.

4.4 Peak discharge estimation

In stage IV we use the lake volumes from the bathymetry dataset to estimate the peak discharge for potential lake outbursts. Estimations were performed using relationships in Table 4, allowing assessment of the range of estimations. As seen in Figure 11, peak discharge values using the Haeberli (1983) relationship for extreme values are substantially larger (ca. two times) than mean-value estimates from Huggel et al. (2002) and Popov (1991). The results show the extremely high potential discharge of lake outbursts of up to thousands of cubic meters per second, and thus the high destructive potential of such catastrophic events. It should be noted that the literature with the underlying data of lake outburst events is often not clear whether volumes represent total lake volume or outburst volume.

4.5 Upscaling the estimations to the Cordillera Blanca

Findings from stages I, II and III of the analysis were used to estimate mean depth, lake volume and peak discharge for every lake in the inventory dataset of the Cordillera

Blanca, which also includes lakes from the bathymetry dataset. Table 10 summarizes the applied relationships.

To estimate mean depth we applied the linear relationship using lake width as the independent variable (Md_Wi). We used the relationship derived from the EBDS lake group since the dam material is unknown for most of the lakes in the inventory dataset. Then lake volumes were estimated as $V = \text{area} \cdot \text{Md_Wi}$, where area is the reported lake area in the inventory dataset and Md_Wi is the mean depth previously estimated. The error estimation for mean depth and volume is assumed to be similar to the one obtained from the EBDS, 44 and 45% respectively. For peak discharge estimation, we applied the three relationships from Haeberli (1983), Popov (1991) and Huggel et al. (2002).

A statistical summary of the results is presented in Table 11, showing minimum, maximum and mean values of 860 lakes in the Cordillera Blanca. The total volume estimated for the lakes in the Cordillera Blanca is $1.29 \times 10^9 \text{ m}^3$ including lakes with ($7.71 \times 10^8 \text{ m}^3$) and without ($5.18 \times 10^8 \text{ m}^3 \pm 2.33 \times 10^8$) measured bathymetries. The lakes with measured bathymetries contain about 60% ($\pm 11\%$) of the total water volume of all lakes in the Cordillera Blanca. The total volume of all lakes without bathymetries is about eight times higher than the estimated volume of potential future lakes ($6.34 \times 10^7 \text{ m}^3$) that could emerge as consequence of glacier retreat and climate change in the Cordillera Blanca of about $6.34 \times 10^7 \text{ m}^3$ (Colonia et al., 2017). When applying the corresponding relationships from RBDS, MBDS and MBDS10, the total lake volumes for the Cordillera Blanca are $1.62 \times 10^9 \pm 4.42 \times 10^8$, $1.15 \times 10^9 \pm 1.74 \times 10^8$ and $1.46 \times 10^9 \pm 3.16 \times 10^8 \text{ m}^3$ respectively (see supplementary material S2).

The distribution of estimated mean depth for the inventory dataset is observed in both, the box-plot and histogram in Figure 12-A. Results show that ca. 70% (610) of the lakes have a mean depth lower than 10 m, and very few are deeper than 40 m. Additionally, ca. 95% of lakes have mean depths lower than 20 m.

In the same way, Figure 12-B shows the distribution of estimated lake volumes. Results indicate that ca. 50% of the values are smaller than 600 000 m³ and ca. 95% are smaller than 5x10⁶ m³. Additionally, there is one lake with a volume lower than 20 000 m³. This is an effect of the lake area defined in the inventory dataset (area < 5000 m²) and not a natural phenomenon.

Finally, an overview of peak discharge estimates is presented in Figure 12-C through box-plots. Results clearly distinguish the extreme values from Haeberli (1983) that could represent the worst case, compared with estimates from Huggel et al. (2002) and Popov (1991) that could better represent average conditions. Generally, we found that ca. 50% of the lakes have estimated peak discharges below 1000 m³/s, but that several cases with values higher than 10⁴ m³/s.

5 Discussion

In this study, we presented and analyzed the largest bathymetric dataset of glacial lakes globally available to date. For the international research community, an important value of this dataset from Peru lies in: 1) better understanding of the statistical relations and associated variability of geometric lake variables; 2) the evaluation of existing empirical relationships which were typically derived from

measured or estimated parameters from a limited number of lakes; 3) establishing more robust and new relationships with a better quantification of uncertainties associated with lake volumes. Relationships for assessing volumes of glacial lakes are often relatively uncritically applied in research and engineering practice without having a clear understanding of related uncertainties, which is, however, essential for applications like lake outburst hazard and water resource assessments.

Three main findings highlight: 1) empirical relationships does not capture the high heterogeneity of lake geometries and therefore, the related lake mean depth and volume; 2) if an empirical volume estimation is needed, then it is recommendable to obtain it by multiplying lake area times mean depth, where mean depth has been estimated as a function of lake width through linear relationship; 3) however, it should be noted that empirical relationships generally provide large error estimations around $\pm 50\%$, and thus offer an order of magnitude perspective.

We found that the correlation between mean depth and any other lake variable is predominantly weak, confirming findings from previous studies (Haeberli et al. 2016c). Moreover, this is independent of the tested empirical relationship (simple/multiple linear relationships or a more complex power or exponential relation). This is because the lake geometry is linked to its formation that in turn depends on glacier dynamics and complex geomorphological processes in surrounding landscapes (Cook and Swift 2012; Carrivick and Tweed 2013).

Our results show that the R^2 values from relationships increases as the size of the sample decreases. This could explain why some previous studies based on relatively small datasets (Huggel et al. 2002; Wang et al. 2012; Yao et al. 2012) obtained relatively high values of R^2 . However, error estimations provided high values, even for

relationships with apparently good performance due to their high R^2 values. Therefore, it is highly recommended to calculate the error of estimations.

In this study, the mean depth estimated from linear relationship, with lake width as the independent variable, provided the best performance. The single differences between the R^2 and MSE values from different relationships based on lake area, length and width are small. However, only lake width generates the highest R^2 and the lowest MSE values at the same time, and for every lake group. It is important to notice that the resulting estimates were compared with the same dataset from which that relationship was derived. As a result, the corresponding R^2 and error estimation values will perform better compared with relationships in the literature. Relationships should preferably be derived from the region where the estimations will be applied; using relationships from other regions could result in a more substantial error in estimates.

Grouping lakes according to their distance to current glaciers seems to provide better estimations but only for lake mean depth and if the lakes are close to glaciers. However, these results must be taken carefully. As previously indicated, the number of lakes used for estimations plays an important role for results. Estimates for mean depth from lakes up to 1 km away from glaciers provide the highest R^2 and lower TE% values but also involve the least number of lakes (i.e. MBDS10 up to 1 km involves only 13 lakes). This is quite an unexpected result, since the lakes close to glaciers and in non-consolidated environments (like MBDS and MBDS10) are susceptible to rock or ice avalanches. Possibly, in the case of the Cordillera Blanca, most of the lakes close to glaciers were recently formed and the geomorphological processes do not have yet an important influence. For lakes further away than 1 km, distance does not seem to provide any additional information. Such lakes could be older and be exposed to larger sediment fluxes, therefore the lake geometry could change according to

particular landscapes and geomorphic dynamics. Unfortunately, the time of formation is unknown for most of the lakes in the Cordillera Blanca, and thus further research is needed.

We highlight that empirical volume estimates not only provide mean values but also tend to compress estimates around them, while extremes are barely represented. Estimations provide results related to a standard lake shape, that is, large lake area with large values of mean depth and volume. This average behavior must be taken into account when using results for further assessment. Consequently, for water resources, inadequate values could provide an overestimation of the potential water storage, while for risk assessment underestimations of possible peak discharge could be dangerous. In consequence, for water resources management, conservative relationships yielding lower values would be more appropriate, while for hazard and risk assessment higher estimates should also be considered.

For hazard assessment in particular, there is a large range of peak discharge estimates based on the different approaches. Relationships produce extremely high peak discharges for the large lake volumes, hardly realistic for typical dam breach processes but possible for overtopping waves triggered by avalanches or landslides impacting on lakes. More precise outburst flood estimates can be obtained by physically based models simulating mass flow impact and wave generation in a lake (Schneider et al. 2014; Westoby et al. 2014), or simpler empirical estimation equations such as presented by Heller et al. (2009). Lake bathymetries are important input data for such models and estimation approaches because wave generation crucially depends on lake geometry. However, large uncertainties may be involved due to the range of possible mass flow impact scenarios producing outburst floods (Schaub et al. 2016) which then needs to be translated into hazard, risk or evacuation maps (Frey et

al. 2018; Hofflinger et al. 2019). Defining scenarios is also a way to deal with the uncertainties; Kougkoulos et al. (2018), for instance, defined 20%, 50% and 100% drainage to account for different GLOF severities in the Bolivian Andes. Such an approach as well as the one presented in this paper could be considered as a middle ground between simple empirical estimations and complex physical modelling.

In terms of water management, volume estimates for all the lakes in the Cordillera Blanca indicate a large amount of water impounded in current lakes. Considering the declining dry season water supply in Peru as a consequence of glacier recession (Condom et al. 2012; Vuille et al. 2018), these lakes could represent an important opportunity for water storage but only if they were to be regulated. Currently, in the upper part of the Santa River catchment, the water demand for agriculture, hydropower, and domestic use were estimated as 7.23×10^7 , 2.32×10^7 and 2.78×10^8 m³/year, respectively (Drenkhan et al. 2019b), which equals one third of the total (lower) estimate of the lake volume. Further assessment is required in order to identify whether future changes in precipitation and temperature, and reduced glacier meltwater contribution (Juen et al. 2007; Chevallier et al. 2011; Schauwecker et al. 2014, 2017) can still refill the lakes.

Another important aspect is whether natural lakes can be regulated and how such regulation can be managed, an aspect that has become critical in the Cordillera Blanca in Peru and elsewhere as well. Many years of conflict over lake water use and management in the Cordillera Blanca continue unresolved into the present day (Carey et al. 2012; Drenkhan et al. 2015; French et al. 2015). Furthermore, many of the lakes are located inside the Huascarán National Park, a protected area where infrastructure development is prohibited or strongly restricted. Many of the lakes have a cultural meaning for local people (Bolin 2009; Cadena 2015; Jurt et al. 2015), which must be

taken into account to avoid further social conflicts. These statements support the need for a more comprehensive approach for water management, incorporating the social dynamics. Methods like agent-based model (Wens et al. 2019), multi-agent networks (Dermody et al. 2018), or other socio-hydrology approaches (Sivapalan et al. 2012) aim at merging hydrological, climate and social aspects. Additionally, a management concept that appropriately considers multiple uses is becoming increasingly urgent. Multi-purpose water projects in lakes have been proposed as a possible infrastructure and management solution that can accommodate several users from sectors such as agriculture, hydropower, domestic, tourism, combined with GLOF risk reduction efforts and ecosystem services (Haeberli et al. 2016b; Drenkhan et al. 2019a).

6 Conclusions

Estimation of glacial lake depth and volume represents a significant challenge where bathymetric data are not available. Based on the large bathymetry dataset of the Cordillera Blanca, Peru, we showed that lake mean depth and volume are generally difficult to predict with other lake morphometric variables. The best results are obtained when estimating lake mean depth as a function of lake width through linear relationship (although a function of lake length cannot be completely discarded). Lake volume estimated as the product of reported lake area and estimated lake mean depth provides results at least as good as existing relationships, but offers several advantages to the processes and avoid self-correlation. We furthermore found with the dataset that our new empirical relationships, based on higher correlations and lower error estimations, improve the accuracy of the mean values whereas extreme values are barely captured. The bathymetric dataset revealed that lake volumes based on a given lake depth can vary by as much as one order of magnitude. Hence, it needs

to be carefully considered whether mean or extreme values are used, depending on the purpose of the application, such as water resource management or hazard and risk assessments. Improvements of mean depth and lake volume estimations will depend on the understanding of the glacio-geomorphological processes of lake formation, and bathymetric precision may be required depending on the exact purposes. However, for a first-order estimate of the total water volume impounded in lakes, in a region like the Cordillera Blanca, empirical relationships might be useful. We found that the total lake volume in the Cordillera Blanca is ca. $1.15 \times 10^9 \text{ m}^3$, six times more than the expected volume of potential future lakes, while lakes without bathymetric surveys represent ca. 32% of the total volume.

Acknowledgments

This study was developed in the framework of the 'Proyecto Glaciares+', an initiative by the Swiss Agency for Development and Cooperation (SDC) and the government of Peru, implemented by CARE Peru and a Swiss consortium led by the University of Zurich. We acknowledge the collaboration with CARE Peru and the Área de Evaluación de Glaciares y Lagunas (AEGL) from the National Water Authority (ANA) of Peru, in particular to Cesar Salazar and Eduardo Sánchez. The careful and constructive reviews by Duncan Quincey and Simon Cook helped to further improve an earlier version of this article. The authors have declared no conflicts of interest for this article.

Data availability statement

The data that supports the findings of this study are available in the supplementary material of this article.

References

- ANA (2014a) Inventario Nacional de Glaciares y Lagunas. Glaciares, Primera. Huaraz, Perú
- ANA (2014b) Inventario de Lagunas Glaciares del Perú. Autoridad Nacional del Agua, Lima, Perú
- ANA (2012) Delimitación y codificación de unidades hidrográficas del Perú, Primera. Ministerio de Agricultura, Lima, Perú
- ANA (2010) Inventario de glaciares de la Cordillera Blanca. Huaraz
- Bajracharya B, Shrestha AB, Rajbhandari L (2007) Glacial Lake Outburst Floods in the Sagarmatha Region. *Mt Res Dev* 27:336–344. doi: 10.1659/mrd.0783
- Bolin I (2009) The glaciers of the Andes are melting: indigenous and anthropological knowledge merge in restoring water resources. *Anthropol Clim Chang From Encount to actions* 228–239
- Cadena M de la (2015) *Earth beings : ecologies of practice across Andean worlds*. Duke University Press
- Carey M (2010) *In the Shadow of Melting Glaciers: Climate Change and Andean Society*. Oxford University Press
- Carey M, French A, O'Brien E (2012) Unintended effects of technology on climate change adaptation: An historical analysis of water conflicts below Andean Glaciers. *J Hist Geogr* 38:181–191. doi: 10.1016/j.jhg.2011.12.002
- Carrivick JL, Tweed FS (2013) Proglacial Lakes: Character, behaviour and geological importance. *Quat Sci Rev* 78:34–52. doi: 10.1016/j.quascirev.2013.07.028
- Chevallier P, Pouyaud B, Suarez W, Condom T (2011) Climate change threats to environment in the tropical Andes: glaciers and water resources. *Reg Environ Chang* 11:179–187. doi: 10.1007/s10113-010-0177-6
- Colonia D, Torres J, Haeberli W, Schauwecker S, Braendle E, Giraldez C, Cochachin A (2017) Compiling an Inventory of Glacier-Bed Overdeepenings and Potential New Lakes in De-Glaciating Areas of the Peruvian Andes: Approach, First Results, and Perspectives for Adaptation to Climate Change. *Water* 9:336. doi: 10.3390/w9050336
- Condom T, Escobar M, Purkey D, Pouget JC, Suarez W, Ramos C, Apaestegui J, Tacsí A, Gomez J (2012) Simulating the implications of glaciers' retreat for water management: A case study in the Rio Santa basin, Peru. *Water Int* 37:442–459. doi: 10.1080/02508060.2012.706773
- Cook SJ, Kougkoulos I, Edwards LA, Dortch J, Hoffmann D (2016) Glacier change and glacial lake outburst flood risk in the Bolivian Andes. *Cryosphere* 10:2399–2413. doi: 10.5194/tc-10-2399-2016
- Cook SJ, Quincey DJ (2015) Estimating the volume of Alpine glacial lakes. *Earth Surf Dyn* 3:559–575. doi: 10.5194/esurf-3-559-2015
- Cook SJ, Swift DA (2012) Subglacial basins: Their origin and importance in glacial systems and landscapes. *Earth Sci Rev* 115:332–372. doi: 10.1016/j.earscirev.2012.09.009
- Dermody B, Van Vuuren DP, Wassen MJ, Dekker SC, Sivapalan M, Stehfest E, Bierkens MFP (2018) A framework for modelling the complexities of food and water security under globalisation. *Earth Syst Dyn* 9:103–118. doi: 10.5194/esd-9-103-2018

- Drenkhan F, Carey M, Huggel C, Seidel J, Oré MT (2015) The changing water cycle: climatic and socioeconomic drivers of water-related changes in the Andes of Peru. *Wiley Interdiscip Rev Water* 2:n/a-n/a. doi: 10.1002/wat2.1105
- Drenkhan F, Guardamino L, Huggel C, Frey H (2018) Current and future glacier and lake assessment in the deglaciating Vilcanota-Urubamba basin, Peruvian Andes. *Glob Planet Change* 169:105–118. doi: 10.1016/j.gloplacha.2018.07.005
- Drenkhan F, Huggel C, Guardamino L, Haeberli W (2019a) Managing risks and future options from new lakes in the deglaciating Andes of Peru: The example of the Vilcanota-Urubamba basin. *Sci Total Environ* 665:465–483. doi: 10.1016/j.scitotenv.2019.02.070
- Drenkhan F, Muñoz R, Huggel C, Frey H, Valenzuela F, Motschmann A, Guardamino L (2019b) Socio-economic losses and impacts from glacier retreat in the Santa River catchment. Lima, Perú
- Emmer A, Vilímek V (2014) New method for assessing the susceptibility of glacial lakes to outburst floods in the Cordillera Blanca, Peru. *Hydrol Earth Syst Sci* 18:3461–3479. doi: 10.5194/hess-18-3461-2014
- Evans SG (1986) Landslide damming in the Cordillera of Western Canada. Seattle, Washington, pp 111–130
- Fischer L, Huggel C, Kääb A, Haeberli W (2013) Slope failures and erosion rates on a glacierized high-mountain face under climatic changes. *Earth Surf Process Landforms* 38:836–846. doi: 10.1002/esp.3355
- French A, Barandiarán J, Rampini C (2015) Contextualizing conflict. Vital water and competing values in glaciated environments. In: Huggel C, Carey M, Clague JJ, Kääb A (eds) *The High-Mountain Cryosphere - Environmental Changes and Human Risks*. Cambridge University Press, Cambridge and New York, pp 315–336
- Frey H, García Hernández J, Huggel C, Schneider D, Rohrer M, Gonzales C, Muñoz R, Mesa L, Cochachin A, Price K, Masias P (2014) An Early Warning System for lake outburst floods of the Laguna 513, Cordillera Blanca, Peru. In: *Analysis and Management of Changing Risks for Natural Hazards*. Padua, Italy, pp 1–8
- Frey H, Haeberli W, Linsbauer A, Huggel C, Paul F (2010) A multi-level strategy for anticipating future glacier lake formation and associated hazard potentials. *Nat Hazards Earth Syst Sci* 10:339–352. doi: 10.5194/nhess-10-339-2010
- Frey H, Huggel C, Chisolm RE, Baer P, Mcardell BW, Cochachin A, Portocarrero C (2018) Multi-Source Glacial Lake Outburst Flood Hazard Assessment and Mapping for Huaraz, Cordillera. *Front Earth Sci* 6:1–16. doi: 10.3389/feart.2018.00210
- Gao BC (1996) NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Environ* 58:257–266. doi: 10.1016/S0034-4257(96)00067-3
- Haeberli W (1983) Frequency characteristics of glacier floods in The Swiss Alps. *Ann Glaciol* 4:6. doi: 10.3189/S0260305500005280
- Haeberli W, Buetler M, Huggel C, Friedli TL, Schaub Y, Schleiss AJ (2016a) New lakes in deglaciating high-mountain regions – opportunities and risks. *Clim Change* 139:201–214. doi: 10.1007/s10584-016-1771-5
- Haeberli W, Huggel C, García Hernández J, Guillén Ludeña S (2016b) *Proyectos Multipropósito en Recursos Hídricos en la Alta Montaña Un documento orientador*. Lima
- Haeberli W, Huggel C, Paul F, Change C (2013) *Glacial Responses to Climate Change*. Treatise

Geomorphol 13:152–175. doi: 10.1016/B978-0-12-374739-6.00350-X

Haerberli W, Linsbauer A, Cochachin A, Salazar C, Fischer UH (2016c) On the morphological characteristics of overdeepenings in high-mountain glacier beds. *Earth Surf Process Landforms* 41:1980–1990. doi: 10.1002/esp.3966

Haerberli W, Schaub Y, Huggel C (2016d) Increasing risks related to landslides from degrading permafrost into new lakes in de-glaciating mountain ranges. *Geomorphology*. doi: 10.1016/j.geomorph.2016.02.009

Heller V, Hager WH, Minor H-E (2009) Landslide generated impulse waves in reservoirs: Basics and computation. *Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie (VAW)*. ETH Zürich

Hinderer M (2001) Late Quaternary denudation of the Alps, valley and lake fillings and modern river loads. *Geodin Acta* 14:231–263. doi: 10.1080/09853111.2001.11432446

Hofflinger A, Somos-Valenzuela MA, Vallejos-Romero A (2019) Response time to flood events using a social vulnerability index (ReTSVI). *Nat Hazards Earth Syst Sci* 19:251–267. doi: 10.5194/nhess-19-251-2019

Hubbard B, Heald A, Reynolds JM, Quincey D, Richardson SD, Luyo MZ, Portilla NS, Hambrey MJ (2005) Impact of a rock avalanche on a moraine-dammed proglacial lake: Laguna Safuna Alta, Cordillera Blanca, Peru. *Earth Surf Process Landforms* 30:1251–1264. doi: 10.1002/esp.1198

Huggel C, Haerberli W, Kääb A, Bieri D, Richardson S (2004) An assessment procedure for glacial hazards in the Swiss Alps. *Can Geotech J* 41:1068–1083. doi: 10.1139/t04-053

Huggel C, Kääb A, Haerberli W, Teyssie P, Paul F (2002) Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps. *Can Geotech J* 39:316–330. doi: 10.1139/t01-099

INAIGEM (2018) *Inventario Nacional de Glaciares - Las Cordilleras Glaciares del Perú*, 1st edn. Huaraz, Peru

Juen I, Kaser G, Georges C (2007) Modelling observed and future runoff from a glacierized tropical catchment (Cordillera Blanca, Perú). *Glob Planet Change* 59:37–48. doi: 10.1016/j.gloplacha.2006.11.038

Jurt C, Burga MD, Vicuña L, Huggel C, Orlove B (2015) Local perceptions in climate change debates: insights from case studies in the Alps and the Andes. *Clim Change* 133:511–523. doi: 10.1007/s10584-015-1529-5

Kapitsa V, Shahgedanova M, Machguth H, Severskiy I, Medeu A (2017) Assessment of Evolution of Mountain Lakes and Risks of Glacier Lake Outbursts in the Djungarskiy (Jetyu) Alatau, Central Asia, using Landsat Imagery and Glacier Bed Topography Modelling. *Nat Hazards Earth Syst Sci Discuss* 1–54. doi: 10.5194/nhess-2017-134

Klimeš J, Benešová M, Vilímek V, Bouška P, Cochachin A (2014) The reconstruction of a glacial lake outburst flood using HEC-RAS and its significance for future hazard assessments: An example from Lake 513 in the Cordillera Blanca, Peru. *Nat Hazards* 71:1617–1638. doi: 10.1007/s11069-013-0968-4

Kougkoulos I, Cook SJ, Edwards LA, Clarke LJ, Symeonakis E, Dortch JM, Nesbitt K (2018) Modelling glacial lake outburst flood impacts in the Bolivian Andes. *Nat Hazards* 94:1415–1438. doi: 10.1007/s11069-018-3486-6

Mergili M, Emmer A, Juřicová A, Cochachin A, Fischer JT, Huggel C, Pudasaini SP (2018) How well can we simulate complex hydro-geomorphic process chains? The 2012 multi-lake outburst flood in

- the Santa Cruz Valley (Cordillera Blanca, Perú). *Earth Surf Process Landforms* 43:1373–1389. doi: 10.1002/esp.4318
- Mergili M, Emmer A, Klimeš J, Cochachin A, Vilímek V (2016) 882 lakes of the Cordillera Blanca: An inventory, classification, evolution and assessment of susceptibility to outburst floods. *CATENA* 147:269–279. doi: 10.1016/j.catena.2016.07.032
- Milder JC (2008) ASTER processing method. Dep Nat Resour Cornell Univ
- Milner AM, Khamis K, Battin TJ, et al (2017) Glacier shrinkage driving global changes in downstream systems. *Proc Natl Acad Sci* 114:201619807. doi: 10.1073/pnas.1619807114
- O'Connor JE, Hardison III JH, Costa JE (2001) Debris Flows from Failures of Neoglacial- Age Moraine Dams in the Three Sisters and Mount Jefferson Wilderness Areas, Oregon
- Popov N (1991) Assessment of glacial debris flow hazard in the north Tien-Shan. In: *Proceedings of the Soviet-China-Japan Symposium and field workshop on natural disasters*. pp 384–391
- Portocarrero C (2014) *The Glacial Lake Handbook - Reducing Risk from Dangerous Glacial Lakes in the Cordillera Blanca, Peru*
- Schaub Y, Huggel C, Cochachin A (2016) Ice-avalanche scenario elaboration and uncertainty propagation in numerical simulation of rock-/ice-avalanche-induced impact waves at Mount Hualcán and Lake 513, Peru. *Landslides* 13:1445–1459. doi: 10.1007/s10346-015-0658-2
- Schauwecker S, Rohrer M, Acuña D, Cochachin A, Dávila L, Frey H, Giráldez C, Gómez J, Huggel C, Jacques-Coper M, Loarte E, Salzmann N, Vuille M (2014) Climate trends and glacier retreat in the Cordillera Blanca, Peru, revisited. *Glob Planet Change* 119:85–97. doi: 10.1016/j.gloplacha.2014.05.005
- Schauwecker S, Rohrer M, Huggel C, Endries J, Montoya N, Neukom R, Perry B, Salzmann N, Schwarb M, Suarez W (2017) The freezing level in the tropical Andes, Peru: An indicator for present and future glacier extents. *J Geophys Res Atmos* 122:5172–5189. doi: 10.1002/2016JD025943
- Schneider D, Huggel C, Cochachin A, Guillén S, García J (2014) Mapping hazards from glacier lake outburst floods based on modelling of process cascades at Lake 513, Carhuaz, Peru. *Adv Geosci* 35:145–155. doi: 10.5194/adgeo-35-145-2014
- Sivapalan M, Savenije HHG, Blöschl G (2012) Socio-hydrology: A new science of people and water. *Hydrol Process* 26:1270–1276. doi: 10.1002/hyp.8426
- Vuille M, Carey M, Huggel C, Buytaert W, Rabatel A, Jacobsen D, Soruco A, Villacis M, Yarleque C, Elison Timm O, Condom T, Salzmann N, Sicart JE (2018) Rapid decline of snow and ice in the tropical Andes – Impacts, uncertainties and challenges ahead. *Earth-Science Rev* 176:195–213. doi: 10.1016/j.earscirev.2017.09.019
- Wang X, Liu S, Ding Y, Guo W, Jiang Z, Lin J, Han Y (2012) An approach for estimating the breach probabilities of moraine-dammed lakes in the Chinese Himalayas using remote-sensing data. *Nat Hazards Earth Syst Sci* 12:3109–3122. doi: 10.5194/nhess-12-3109-2012
- Wegner S (2014) *Lo que el agua se llevó. Huaraz, Perú*
- Wens M, Johnson JM, Zagaria C, Veldkamp TIE (2019) Integrating human behavior dynamics into drought risk assessment—A sociohydrologic, agent-based approach. *Wiley Interdiscip Rev Water* e1345. doi: 10.1002/wat2.1345
- Westoby MJ, Glasser NF, Brasington J, Hambrey MJ, Quincey DJ, Reynolds JM (2014) Modelling outburst floods from moraine-dammed glacial lakes. *Earth-Science Rev* 134:137–159. doi:

10.1016/j.earscirev.2014.03.009

Yao X, Liu S, Sun M, Wei J, Guo W (2012) Volume calculation and analysis of the changes in moraine-dammed lakes in the north Himalaya: A case study of Longbasaba lake. *J Glaciol* 58:753–760. doi: 10.3189/2012JoG11J048

Accepted Article

Table 1. Available information in the bathymetry dataset, illustrated with the example of Palcacocha Lake. MxL: maximum length, MxW: maximum width, MxD: maximum depth, MD: mean depth, MFS: maximum front slope, AFS: average front slope, MBS: maximum back slope, ABS: average back slope.

Inventory code	Name	Catchment	Sub-catchment
1376962-7	Palcacocha	Santa	Quillcay

East (m)	North (m)	Altitude (m a.s.l.)	Date
238750	8960551	4562	Apr 09

Department	Province	District
Ancash	Huaraz	Independencia

Area (m ²)	MxL (m)	MxW (m)	MxD (m)	MD (m)	Volume (m ³)
518'426	1'592	397	73.0	33.6	17'325'207

Ratio MxW/MxL	MFS (m/m)	AFS (m/m)	MBS (m/m)	ABS (m/m)
0.25	0.6	0.07	0.32	0.14

Dam material	Dam type
Moraine	With infrastructure

Table 2. Classification of the bathymetry database

		Dam material			Total
		Moraine	Rock	Moraine/Rock	
Dam type	Without infrastructure	69	11	14	94
	With infrastructure	26	1	-	27
Total		95	12	14	121

Table 3. Previous studies on volume and mean depth estimation based on the lake area.

Reference	N°	Dam material	Region	Relationship	R ²	Acronym
Wang et al. (2012)	20	M	Himalayas	$V=0.0354 \cdot A^{1.3724}$ $D=0.087 \cdot A^{0.434}$	0.919 0.503	VW Md_W
Huggel et al. (2002)	15	I,T,M	Worldwide	$V=0.104 \cdot A^{1.42}$ $D=0.104 \cdot A^{0.42}$	NM 0.916	VH Md_H
Kapitsa et al. (2017)	32	R, M	Kazakhstan	$V=0.036 \cdot A^{1.49}$ $D=0.036 \cdot A^{0.49}$	NM NM	VK Md_K
Loriaux & Casassa (2013)	31	I, M	Worldwide	$V=0.2933 \cdot A^{1.3324}$ $D=0.2933 \cdot A^{0.3324}$	0.96 NM	VL Md_L
Evans (1986)	NM	I	Canada	$V=0.035 \cdot A^{1.5}$ $D=0.035 \cdot A^{0.5}$	NM NM	VE Md_E

N°: number of bathymetries used to develop the relationship. NM: not mentioned. M: moraine, I: ice, T: thermokarst, R: rock. R2: published coefficient of determination for the specific study. In italic, the derived relationship for mean depth from published volume relationship.

Table 4. Relations for peak discharge estimation selected from Huggel et al. (2002)

Relationship	Dam type	Comment	Source
$Q_{max} = \frac{2V}{t_w}$	Glacier, non-tunnel events, moraine breaching	Upper-bound statistics	Haeberli (1983)
$Q_{max} = 0.0048 \cdot V^{0.896}$	Moraine	regression statistics	Popov (1991)
$Q_{max} = 0.00077 \cdot V^{1.017}$	Moraine	regression statistics	Huggel et al. (2002)

Q_{max}: peak discharge; V: lake volume; t_w: time constant of 1000 seconds

Table 5. Statistics of the bathymetry dataset and glacial lake inventory dataset. In parenthesis are the name of the lakes (when known) or the lake inventory code (for lakes without known name, in that case see ANA (2014b))

Variable	Bathymetry dataset			Inventory dataset		
	Max	Min	Mean	Max	Min	Mean
Area (m ²)	1'799'207 (Pelagatos)	5'087 (Mellajirca)	242'524	1'799'207 (Pelagatos)	5'003 (4989966-3)	70'269
Width (m)	1'104 (Pelagatos)	44 (Huecrucocha)	352	1'400 (Conococha)	40 (1376498-13)	178
Length (m)	3'335 (Paron)	106 (Mellajirca)	827	3'335 (Paron)	82 (1376452-28)	355
Ratio width/length (m/m)	1.2 (1376936-7)	0.05 (Huecrucocha)	0.5	1.4 (1376935-1)	0.05 (Huecrucocha)	0.6
Max Depth (m)	124 (Purhuay)	3 (Contotora)	36			
Mean Depth (m)	62 (Auquiscocha)	1 (Contotora)	17			
Volume (m ³)	111'391'130 (Pelagatos)	16'722 (Huandoy)	6'374'737			

Table 6. Coefficients of determination (R^2) between mean depth and the different variables assessed in the four lake groups from the bathymetry dataset. Mean squared errors compare the measured and estimated mean depth. Bold and cursive indicate the higher coefficient of determination and the lowest mean square error for every lake group, for simple or linear relationships.

Relationship	EBDS		RBDS		MBDS		MBDS10	
	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
Simple linear relationship								
Mean depth vs Area	0.26	112	0.61	124	0.34	93	0.45	63
Mean depth vs Length	0.18	125	0.29	226	0.24	107	0.39	70
Mean depth vs Width	0.36	98	0.73	86	0.34	92	0.48	59
Mean depth vs Ratio width/length	0.00	152	0.12	281	0.01	139	0.00	115
Mean depth vs AFS	0.00	131	0.22	271	0.01	125	0.00	114
Mean depth vs MFS	0.06	143	0.29	225	0.01	138	0.01	114
Mean depth vs ABS	0.11	136	0.17	263	0.12	123	0.00	115
Mean depth vs MBS	0.07	142	0.25	238	0.02	138	0.00	115
Multiple linear relationships								
Mean depth vs Area and Width	0.35	98	0.71	450	0.35	248	0.47	196
Mean depth vs Area and Length	0.26	111	0.63	449	0.34	212	0.46	176
Mean depth vs Width and length	0.35	98	0.73	444	0.37	248	0.51	196
Mean depth vs Area, Width and Length	0.34	98	0.67	458	0.37	248	0.51	195

Table 7. Coefficients of determination R^2 and total error (TE%) between the measured and estimated mean depth from bathymetry dataset. Bold and italic fonts highlight the higher R^2 and lower TE for every lake group.

Relationship	Acronym	EBDS		RBDS		MBDS		MBDS10	
		R^2	TE%	R^2	TE%	R^2	TE%	R^2	TE%
Wang et al. (2012)	Md_W	0.32	45	0.60	58	0.34	47	0.45	38
Huggel et al. (2002)	Md_H	0.32	45	0.60	58	0.33	41	0.45	37
Kapitsa et al. (2017)	Md_K	0.32	56	0.61	71	0.34	55	0.45	58
Loriaux & Casassa (2013)	Md_L	0.33	47	0.59	64	0.33	48	0.44	38
Evans (1986)	Md_E	0.32	50	0.61	63	0.34	51	0.45	49
$Md = a \cdot W + b$	Md_Wi	0.36	44	0.73	40	0.35	50	0.48	24

Where Md is mean depth, W width, and a and b are coefficients, different for every lake group: EBDS (a=0.041; b=2); RBDS (a=0.069; b=-5.43); MBDS (a=0.037; b=1.562); MBDS10 (a=0.039; b=6.042)

Table 8. Coefficients of determination R^2 and total error (TE%) between the bathymetrically derived and estimated volume with different relationships for every lake group. Bold and italic fonts highlight the higher R^2 and lower TE for every lake group.

Relationship	Acronym	EBDS		RBDS		MBDS		MBDS10	
		R^2	TE%	R^2	TE%	R^2	TE%	R^2	TE%
Wang et al. (2012)	VW	0.81	442	0.91	528	0.85	370	0.91	559
Huggel et al (2002)	VH	0.81	45	0.93	69	0.85	44	0.92	39
Kapitsa et al (2017)	VK	0.81	58	0.93	87	0.85	51	0.92	60
Loriaux & Casassa (2013)	VL	0.81	47	0.92	74	0.85	46	0.91	41
Evans (1986)	VE	0.81	51	0.93	78	0.85	46	0.92	50
$V = A \cdot Md_Wi$	V_Md_Wi	0.87	45	0.97	52	0.85	46	0.92	26

V: volume, A: area, Md_Wi: mean depth estimated from linear relationship with lake width as independent variable.

Table 9. Coefficients of determination (R^2) and total error (TE %) between measured and estimated lake mean depth and volume for every lake group according to the distance of lakes to the current glacier margins. Mean depth was estimated by linear relationship and volume as lake area multiplied with estimated mean depth.

Distance threshold, up to:	Lake group	N	Mean depth		Volume	
			R^2	TE%	R^2	TE%
1 km	EBDS	53	0.44	34	0.71	36
	MBDS	21	0.66	34	0.96	35
	RBDS	9	0.41	40	0.32	51
	MBDS10 †	13	0.71	17	0.97	21
5 km	EBDS	93	0.45	40	0.84	43
	MBDS	47	0.42	45	0.89	50
	RBDS	11	0.73	43	0.97	54
	MBDS10	28	0.61	17	0.92	21
10 km	EBDS †	108	0.40	42	0.89	44
	MBDS	59	0.39	49	0.90	51
	MBDS10 †	36	0.56	22	0.93	23
Without threshold	EBDS	121	0.36	44	0.87	45
	MBDS	69	0.73	40	0.97	52
	RBDS	12	0.35	50	0.85	46
	MBDS10	42	0.48	24	0.92	26

N: number of lakes assessed in the threshold. Bold and cursive formats indicate where both R^2 and TE% values improve at the same time compared to values obtained without applying distance thresholds (the latter come from Table 7 and 8). †: indicates where both mean depth and volume estimations improved at the same time.

Table 10. Relationships applied to estimate mean depth, volume and peak discharge for the inventory dataset. Lake width and area are measured values.

Estimation	Relationship	Source
Mean depth	$Md_Wi = 0.041 \cdot \text{Width} + 2$	This study
Volume	$V = \text{area} \cdot Md_Wi$	This study
Peak discharge	$Q_{max} = \frac{2V}{t_w}$	Haeberli (1983)
	$Q_{max} = 0.0048 \cdot V^{0.896}$	Popov (1991)
	$Q_{max} = 0.00077 \cdot V^{1.017}$	Huggel et al. (2002)

Table 11. Estimated mean depth, volume and peak discharge for the inventory dataset.

	Mean depth (m)	Volume (million m ³)	Peak discharge (m ³ /s)		
			Haeberli (1983)	Popov (1991)	Huggel et al. (2002)
Max	60	99.6	199,383	70,476	104,985
Min	4	0.02	39	33	18
Mean	9	1.5	2,998	1,357	1,521

Figure 1. An overview map of the Cordillera Blanca with glacier and lake locations from the AEGL latest inventory.

Figure 2. Diagram with the lake geometric data available in the bathymetry dataset.

MxL: maximum length, MxW: maximum width, MxD: maximum depth, MD: mean depth, MFS: maximum front slope, AFS: average front slope, MBS: maximum back slope, ABS: average back slope.

Figure 3. Lake groups from the bathymetry dataset to assess the volume estimation. Selected lakes represent examples of every lake group. EBDS: Entire bathymetry dataset, MBDS: moraine dam bathymetry dataset without infrastructure, RBDS: rock dam bathymetry dataset, MBDS10: moraine dam bathymetry dataset with mean lake depths exceeding 10 m.

Figure 4. Stages for the mean depth, volume and peak discharge estimation used in the present paper.

Figure 5. Box-plot for the lake variables including outlier values (bullets). The length of the whisker (box) groups 95% (50%) of the data points, while every point outside this region is considered an outlier

Figure 6. Histograms of variables for the inventory and bathymetry datasets. Y-axis in logarithmic scale.

Figure 7. Plots between mean depth and other variables from for the entire bathymetry dataset (EBDS). AFS: average front slope, MFS: maximum front slope, ABS: average back slope, MBS: maximum back slope (for correlation values cf. Table 6).

Figure 8. Correlation between mean depth and length or width, for every lake group.

Figure 9. Box-plot comparison between measured and estimated mean depth for every lake group. For the meaning of acronyms, see Table 7.

Figure 10. Box-plot comparison between bathymetrically derived and estimated volume for every lake group including outliers (bullets).

Figure 11. Peak discharge (plotted as log-log) estimates using bathymetrically derived volumes from the bathymetry dataset.

Figure 12. Box-plots and histograms of the estimated lake mean depth (A), volume (B) and peak discharge (C) for the 860 lakes of the inventory dataset in the Cordillera Blanca, Peru. For lake volume and peak discharge the box-plot on the left includes the outlier points (bullets) while the right one does not to facilitate the visualization.