Contemporary glacial lakes in the Peruvian Andes

Wood, J.L.[1], Harrison, S.[1], Wilson, R.[2], Emmer, A.[3], Yarleque, C.[4], Glasser, N.F.[5], Torres, J. C.[4], Caballero, A.[4], Araujo, J.[4], Bennett, G.L.[1], Diaz-Moreno, A.[6], Garay, D.[4], Jara, H.[4], Poma, C.[7], Reynolds, J.M[6], Riveros, C.A.[4,8], Romero, E.[4], Shannon, S.[9], Tinoco, T.[7], Turpo, E.[8] and Villafane, H.[4]

[1]University of Exeter, UK; [2]University of Huddersfield, UK; [3]University of Graz, AT; [4]INAIGEM, PE; [5]Aberystwyth University, UK; [6]Reynolds International Ltd, UK; [7]UNASAM, PE; [8]IBC, PE; [9]University of Bristol, UK; [10]UNALM, PE

Abstract

Glacier recession in response to climate warming has resulted in an increase in the size and number of glacial lakes. Glacial lakes are an important focus for research as they impact water resources, glacier mass balance, and some produce catastrophic glacial lake outburst floods (GLOFs). Glaciers in Peru have retreated and thinned in recent decades, prompting the need for monitoring of ice- and water-bodies across the cordilleras. These monitoring efforts have been greatly facilitated by the availability of satellite imagery. However, knowledge gaps remain, particularly in relation to the formation, temporal evolution, and catastrophic drainage of glacial lakes. In this paper we address this gap by producing the most current and detailed glacial lake inventory in Peru and provide a set of reproducible methods that can be applied consistently for different time periods, and for other mountainous regions.

The new lake inventory presented includes a total of 4,557 glacial lakes covering a total area of 328.85 km^2. In addition to detailing lake distribution and extent, the inventory includes other metrics, such as dam type and volume, which are important for GLOF hazard assessments. Analysis of these metrics showed that the majority of glacial lakes are detached from current glaciers (97%) and are classified as either embedded (i.e. bedrock dammed; ~64% of all lakes) or (moraine) dammed (~28% of all lakes) lakes. We also found that lake size varies with dam type; with dammed lakes tending to have larger areas than embedded lakes. The inventory presented provides an unparalleled view of the current state of glacial lakes in Peru and represents an important first step towards (1) improved understanding of glacial lakes and their topographic and morphological characteristics and (2) assessing risk associated with GLOFs.

1. Introduction

The recession of glaciers globally in response to climate warming has led to a dramatic increase in the size and number of supraglacial and proglacial lake systems (e.g. Rabatel et al., 2013; Haeberli et al., 2016; Shugar et al., 2020). In particular, post-Little Ice Age climatic warming has enhanced ice melt, leading to the development of a large number of glacial lakes behind ice dams, lateral and terminal moraines and within over-deepened de-glaciated valley bottoms (Quincey et al., 2007; Wilson et al., 2018). Glacial lakes are important globally and regionally as (1) they represent a considerable water resource (Loriaux and Casassa, 2013), (2) when in contact with or dammed by glaciers, they can have negative impacts on glacier mass balance (e.g. King et al., 2019), and (3) they are the source of glacial lake outburst floods (GLOFs; Richardson and Reynolds, 2000; Carrivick and Tweed, 2016; Harrison et al., 2018) which are considered to be the largest and most extensive glacial hazard in terms of disaster and damage potential (UNEP, 2007).

Glaciers in Peru have retreated and thinned considerably over recent decades (INAIGEM, 2018), prompting the need for greater monitoring of ice- and water-bodies contained within glacierised basins. Such efforts are important as they help inform local and national mitigation policies concerning the impacts of glacier retreat on water resources, mountain development, tourism and hazards. The availability of multi-temporal satellite imagery has greatly improved our understanding of glacier change and lake distribution in countries like Peru (e.g. Drenkhan et al., 2018), however, knowledge gaps remain, particularly in relation to the formation, temporal evolution, and drainage of glacial lakes.

This study aims to robustly identify, describe, and analyse the glacial lakes of Peru. In this paper we discuss the new lake inventory in detail and provide statistics regarding the different dam types, extent, and topographic setting of the glacial lakes of Peru. The methods used here are designed to be reproducible (allowing them to be applied to map glacial lakes in other glacierised regions) and will form the basis for assessing the evolution of lakes through time, as part of nationwide GLOF hazard assessments in Peru.

1.1 Glacial recession and GLOFs in the Peruvian Andes

The Peruvian Andes are home to 70% of the world's tropical glaciers covering an area of >1,600 km^2 (WGMS, n.d.). In line with other areas of the Andes (Masiokas et al., 2009; Davies & Glasser, 2012; Rabatel et al., 2013), glaciers in Peru have, in general, undergone a sustained period of retreat and thinning since reaching their Little Ice Age Maximums (LIAMs) (Vuille et al., 2008; Hanshaw & Bookhagen, 2014; UGRG, 2014); with some cordilleras becoming completely deglaciated since the 1970s (e.g. Cordillera Barroso; INAIGEM, 2018; Supplementary Information 4 Figure S1). Marked by lateral and terminal moraines, studies suggest that LIAM glacier positions in Peru extended some >2,000 m down valley of their 21st Century extent, with length varying according to localised topographic and climate settings (Drenkhan et al., 2018; Emmer et al., 2021; Supplementary Information 5 Table S1). The extent of glacial retreat across the Peruvian Andes over recent decades has led to concerns that the deglaciation discharge dividend in this region may have already peaked (Baraer et al., 2012); placing renewed emphasis on efforts to quantify both glacier health and lake distribution in the region.

Glacial lake outburst floods (GLOFs) are known for their extreme peak discharges (e.g. Clague et al., 2012) and are among the most important geomorphic agents in deglaciating mountain ranges across the globe; presenting a serious natural hazard (Reynolds, 1992; Carrivick and Tweed, 2016). As well as increasing the exposure of mountain societies to GLOF hazards, glacier retreat-induced formation and evolution of glacial lakes raises GLOF disaster risk concerns; especially in low-income countries of high Asia and South America (Emmer, 2018). While historical GLOF records, although incomplete, allow us to reveal general and regionally specific GLOF susceptibility indicators (e.g. lake, dam and surrounding geomorphic characteristics; see Kougkoulos et al., 2018), reliable evaluation of GLOF susceptibility still requires up to date lake inventory data with quantitative as well as qualitative lake characteristics.

1.2 Glacier lake inventories

Globally, there has been a recent increase in the number of available lake inventories; partly due to the impact that lakes have on continental carbon cycles, biogeochemical processes, water resources and GLOF hazards (Emmer et al., 2020; Verpoorter et al., 2014). From the use of early aerial images (e.g. Emmer et al., 2016, Viani et al., 2016) to the availability of long time-series global satellite data (such as the Landsat missions from 1972 onwards; Shugar et al., 2020), open source and “big data'' cloud computing has expedited the creation of lake inventories for individual basins (Mahdianpari et al., 2019; Kumar et al., 2020), wider regions (e.g. Mosquera et al., 2017; Wilson et al. 2018; Wang et al., 2020; Worni et al. 2013), globally (Verpoorter et al., 2014; Shugar et al., 2020), and their evolution through time. This has allowed researchers to identify changes in lake size (and therefore estimate changes to lake volume) in order to better constrain and model the hydrological response of glaciers to climate change (e.g. Shugar et al., 2020).

A number of historic sub-national lake inventories exist for the Peruvian cordilleras (e.g. Cordillera Blanca: Emmer et al., 2016, 2020, Vilímek et al., 2016; Vilcanota-Urubamba basin: Drenkhan et al., 2018), and provide some estimates of lake volume and lake area-depth-volume relationships (e.g. Cordillera Blanca: Munõz et al., 2020) and future lake growth potential (e.g. Colonia et al., 2017; Drenkhan et al., 2018). The inventory of the Instituto Nacional de Investigación en Glaciares y Ecosistemas de Montaña (INAIGEM, n.d.; henceforth the inventory is referred to as INAIGEM) is the only existing lake inventory to cover all 20 glaciated cordilleras. This inventory covers an observation period of 2016 and uses a variety of satellite sensors (Supplementary Information 5 Table S2). The Autoridad Nacional del Agua inventory (ANA, 2014; henceforth referred to as ANA) covers 19 cordilleras (with the exception of Cordillera Barroso) and an observation period of 2001-2010. This inventory also uses a variety of different sensors, which vary across the cordilleras (Supplementary Information 5 Table S3). The lake inventory by Emmer (2016, henceforth referred to as Emmer) covers the Cordillera Blanca and was generated using a variety of remote sensing techniques, covering the period 1948-2018. These inventories (ANA, INAIGEM and Emmer) include many important metrics for understanding lake evolution and GLOF potential (Table 1) and are presented in Supplementary Information 1 as they will be used as comparisons for this study.

Although these previous inventories are very valuable, for robust quantification of glacial lake changes through time it is important for there to be internal consistency in the methods and data sources used to derive the lake outlines. This is the gap that the current study aims to fill.

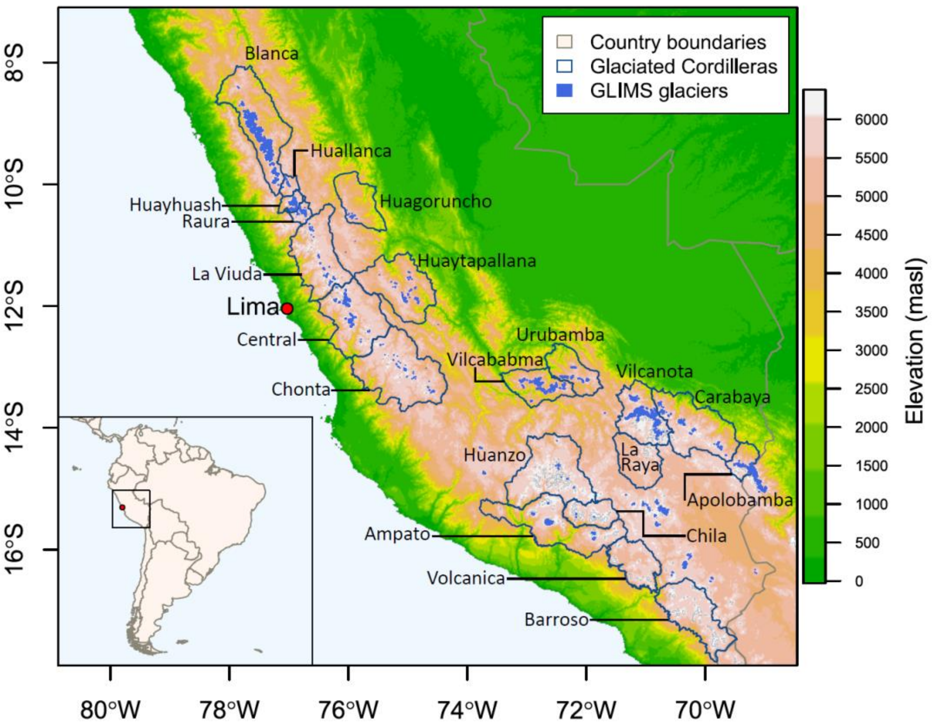

Figure 1. Of the 20 Peruvian cordilleras shown, only 17 are currently glaciarised (GLIMS, 2019; La Raya, Volcanica and Barroso are not currently glacierised). For this paper, lakes in the unknown regions (i.e. lakes that fall outside of the 20 named cordilleras; *n* = 309) have been removed from the presented analyses.

Table 1. Metrics that were recorded in ANA, INAIGEM, Emmer and the new projectGLOP (described in this study) lake inventories.

Metric	Details	Sample technique (projectGLOP)	project GLOP	Emmer	INAIGEM	ANA
ID	Unique ID for all mapped lakes	ID automatically generated in QGIS using the field calculator	yes	yes	yes	yes
lake_type	Type based on dam type. We give broad definitions of “dammed” (which includes moraine dammed lakes), “embedded” (bedrock dammed lakes) and “unclassified” for lakes dammed by landslides, as well as lakes in which the dam type cannot be identified.	Visual interpretation of remotely sensed imagery by the digitiser	yes	yes		
l_sub_type	In contact or not in contact with ice	Identified by contact with the GLIMS (2019) polygons	yes	yes		
Outflow	Lake outflow	Identified by the digitiser based on visible outflow on the ESRI, Google or Bing aerial images through QuickMapServices in QGIS	yes	yes		
Notes	Any notes deemed important	Identified by the digitiser	yes	yes		
Digitised	Username of person digitising lake	Signed by the digitiser	yes		yes	
Year	Year of image used	Year (date) of imagery used	yes		yes	yes
Sensor	e.g. Landsat, ASTER, Liss III, etc	Identified by the digitiser	yes		yes	yes
area_utm18	Lake area (in m^2; utm18)	Calculated using the Field Calculator in QGIS	yes	size category	yes	yes
area_utm19	Lake area (in m^2; utm19)	Calculated using the Field Calculator in QGIS	yes		yes	yes
area_utm	Lake area (in m^2; in either utm18 or utm19)	Compilation of the area_utm18 and area_utm19 data	yes		yes	yes
volume_m3	Lake volume from Guardamino and Drenkhan (2016) and ANA (2014) to define scaling relationships.	Calculated based on available data (see Supplementary Information 2.2.3)	Yes			yes
depth_m	Lake depth from Guardamino and Drenkhan (2016) and ANA (2014)	Manually input into the inventory	Yes			yes
ele30mean	Lake ele30mean	Calculated in QGIS using the raster Zonal Statistics tool	yes	yes	yes	yes

2. Methods

This paper principally presents a new glacier lake inventory for the Peruvian cordilleras (henceforth projectGLOP). A full account of the methods applied are summarised in Figure 2 and detailed in Supplementary Information 2.

Initially, glacier lakes were defined as all lakes within 3 km of existing glaciers (using the GLIMS/Randolph Glacier Inventory v6.0; GLIMS, 2019). This 3 km buffer represents an assumed LIAM, which was obtained through substantial review of relevant literature (Supplementary Information 5 Table S1; Supplementary Information 2.1.1). Lakes within the LIAM buffer were manually digitised using a combination of Landsat Thematic Mapper Tier 1 data (TMT1; Dykstra and Owen, 2017), derived Normalised Difference Water Index and Normalised Difference Snow Index data, as well as high resolution Quantum Geographic Information System QuickMapServices (NextGIS, 2015) satellite data (Figure 2; Supplementary Information 2.1.2 and 2.1.3). Where available, 2019 Landsat images with <10% cloud cover were used, in a number of limited cases data from 2018 and 2017 were used for lake digitisation (see Figure 2, Supplementary Information 2.1.2 and Supplementary Information 4 Figure S3). A lake digitisation uncertainty analysis was additionally performed to ascertain the repeatability of the methods described; this analysis involved comparison of lake outlines mapped by three separate users (see Supplementary Information 2.1.3). Important metrics (Table 1) were recorded in order that the inventory is applicable across a range of future analyses. In terms of the lake dam type, we use broad definitions of “dammed” (which includes moraine dammed lakes), “embedded” (bedrock dammed lakes) and “unclassified” for lakes dammed by landslides, as well as lakes in which the dam type cannot be identified. We also differentiate between lakes in contact with existing glaciers, and those which are detached.

Lake area was calculated for lakes in the projectGLOP inventory and uncertainty analyses performed to estimate (1) by how much lake area is over/under-estimated using 30m resolution Landsat TMT1 data, and (2) how many small lakes (<900 m^2) are excluded (Supplementary Information 2.2.1). Statistical analyses were performed to gain a picture of lake elevation across Peru (Supplementary Information 2.2.2). We calculated lake volume based on derived scaling relationships (Supplementary Information 2.2.3). Finally, we compared the projectGLOP lake inventory with a number of existing inventories for the Peruvian cordilleras (Supplementary Information 2.3).

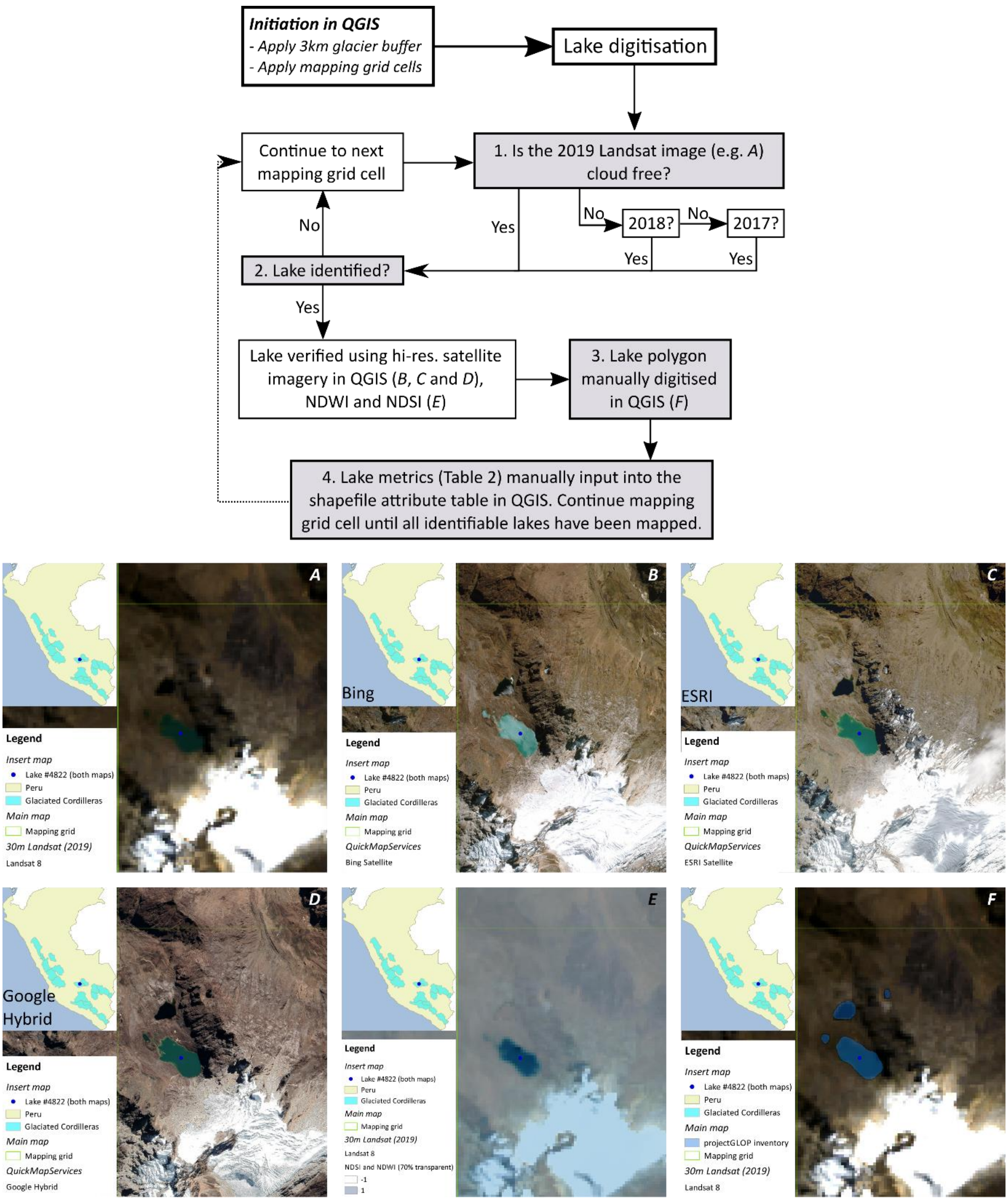

Figure 2. Flow chart showing the lake identification and manual digitisation. The bottom inserts (A-F) centre over lake #4822, Cordillera Vilcabamba, to provide an example of the Landsat data (A; Supplementary Information 2.1.2), NDSI and NDWI calculations (E; semi-transparent 70% over the Landsat data; Supplementary Information 2.1.2 Eq. 1 and 2 respectively). Secondary datasets used for lake identification include the QGIS QuickMapServices plugin (B, C and D) used during lake digitisation. Also shown is an example of the digitised lakes (F; Supplementary Information 2.1.3).

3. Results and Discussion

3.1 projectGLOP lake inventory

Our glacial lake inventory for the Peruvian Cordilleras, comprises a total of 4,557 glacial lakes within 3 km of existing glaciers (based on the GLIMS/Randolph Glacier Inventory v6.0; GLIMS, 2019; Raup et al., 2007). The majority of lakes are detached from current glaciers (97%; Table 2). If we consider all lakes, the majority of these are either embedded (~64%) or dammed (~28%) lakes, with the remaining 5% falling into the unclassified category (see Table 1). These unclassified lakes include 17 landslide dammed lakes (10 in the Cordillera Blanca, two in C. Central, and one in each of C. Apolobamba, C. Carabaya, C. Huallanca, C. La Viuda and C. Vilcanota), one ice dammed lake (in C. Vilcanota), with 222 lakes marked as unclassified as dam type was not identified using readily accessible (ESRI, Bing and Google) satellite imagery. Only 3% of lakes remain in contact with ice; 63% of these are embedded, 35% are dammed and the remaining 2% are unclassified (Table 2). The expansion of glacial lakes often occurs in response to glacial recession, when low gradient glacier termini retreat back into over-deepened basins. The fact that only 3% of lakes remain in contact with ice is of significance as it may limit the growth of current lakes into the future (see Wilson et al., 2018).

Table 2. Lake counts for each cordillera (please refer to Figure 1) by dam type and sub-type (not/in contact with ice). "Dammed" includes moraine dammed lakes, bedrock dammed lakes are "embedded". Lakes dammed by landslides are included in the "unclassified" category, as well as lakes in which the dam type cannot be identified due to unclear satellite images. Supraglacial lakes are not included in the inventory as they are influenced by seasonal variation in drainage.

	Cordillera	Lat/ Lon	Total	Lake type: Embedded		Lake type: Dammed		Lake type: Unclassified		Lake sub-type total	
				in contact with ice	not in contact with ice	in contact with ice	not in contact with ice	in contact with ice	not in contact with ice	in contact with ice	not in contact with ice
Cordilleras del norte	Blanca	77.87 W, 8.06 S to 76.78 W, 10.67 S	**803**	21	424	12	286	1	59	34	769
	Huallanca		**69**	3	55	0	9	0	2	3	66
	Huayhuash		**129**	5	46	2	72	0	4	7	122
	Raura		**245**	3	164	0	54	0	24	3	242
Cordilleras del centro	Huagoruncho	75.92 W, 9.80 S to 74.52 W, 13.62 S	**145**	1	74	0	65	0	5	1	144
	La Viuda		**442**	3	276	0	138	0	25	3	439
	Huaytapallana		**373**	6	244	3	115	0	5	9	364
	Central		**509**	18	340	2	98	0	51	20	489
	Chonta		**212**	0	188	0	22	0	2	0	212
Cordilleras del sur	Urubamba	72.53 W, 12.61 S to 69.40 W, 14.88 S	**139**	1	108	0	25	0	5	1	138
	Vilcabamba		**183**	7	107	0	59	0	10	7	176
	Vilcanota		**490**	13	233	28	199	2	15	43	447
	Carabaya		**590**	7	496	2	68	0	17	9	581
	Apolobamba		**142**	2	104	1	23	0	12	3	139
	Huanzo		**54**	0	42	0	9	0	3	0	54
	Chila		**13**	0	6	1	6	0	0	1	12
	Ampato		**19**	1	13	0	4	0	1	1	18
		Total	**4557**	**91**	**2920**	**51**	**1252**	**3**	**240**	**145**	**4412**

3.2 Lake inventory statistics

The following sections present an analysis of the projectGLOP lake inventory in terms of important characteristics and distributions; specifically, lake area (3.2.1), lake elevation (3.2.2), lake bathymetry (3.2.3), and finally we contextualise this new inventory alongside three existing inventories for Peru (3.3.4). Within each section we consider (1) variation across the Peruvian cordilleras; (2) lake connectivity to existing glaciers; and (3) variation relating to differences in dam type. We consider each of these to have important implications for both water resources and GLOF hazards.

3.2.1 projectGLOP lake area

Knowledge regarding the areal extent of existing glacial lakes in Peru is important, in respect to GLOFs, as it provides a basis for calculating the effective water volume of individual lakes (which can influence the magnitude and duration of GLOF events) and assessing potential likelihood of GLOF trigger and threshold parameters related to the lake dam and the surrounding geomorphic features (Reynolds, 2014; Kougkoulos et al., 2018). Overall, we found that the Cordilleras Vilcanota, Carabaya, La Viuda and Blanca account for >50% of glacial lake area across Peru (Table 3). In terms of dam type, embedded lakes cover the largest area (Table 3), but this varies across the cordilleras. In general, we found that cordilleras that contained a larger extent of dammed lakes than embedded lakes tended to be more glacierised (INAIGEM, 2018; Supplementary Information 4 Figure S1).

Between the cordilleras, the distribution of lake size varies (Figure 3A); pairwise Wilcox tests show that area distributions are statistically different between some cordilleras (e.g. Carabaya and Blanca), while others present similar distributions (e.g. Blanca and Central; Figure 3A; Table S5). There are a number of larger lakes that represent outliers in the distribution of lake areas across the majority of the cordilleras (Figure 3A). In total, there are 33 lakes which are greater than 1 km^2 in area. They lie in Vilcanota (n = 3), Raura (n = 3), La Viuda (n = 5), Huaytapallana (n = 1), Chonta (n = 2), Central (n = 7), Carabaya (n = 6), Blanca (n = 1) and Apolobamba (n = 4). Two of these lakes (one in Vilcanota and one in Apolobamba), are greater than 10 km^2, and are recorded here as they lie (at least partly) within the 3 km buffer (described in Supplementary Information 2.1.1). It is likely that these lakes are older than the LIAM, however, they are located in close proximity to existing glaciers (i.e. <3 km) and so have been included within the inventory.

Cordilleras which have seen the biggest loss of glacier extent (by >80% since the 1950's) include La Viuda, Chonta, Huanzo, Chila and La Raya (INAIGEM, 2018; Supplementary Information 4 Figure S1); Huanzo and Chonta are the only cordilleras in which all glacier lakes are detached from current glaciers (Table 3 and Figure 3B). For each cordillera, Kruskal-Wallis rank sum tests were performed to see if there is a significant difference between lake area and contact with existing glaciers. Where a significant difference was found, larger lakes were in contact with existing glaciers in cordilleras Vilcanota and Blanca, whilst lakes in La Viuda are larger when detached from glaciers (Figure 3B and Supplementary Information 5 Table S5); for all other cordilleras, no significant difference was found between the two groups, possibly due to low recorded lake numbers in contact with ice (Table 2).

Table 3. Calculated lake areas for each of the studied cordilleras in Peru and separated by dam type.

	Dammed (km²)	Embedded (km²)	Unclassified (km²)	Total (km²)
Blanca	19.64	15.41	3.68	38.72
Huallanca	0.77	0.82	0.10	1.68
Huagoruncho	6.53	5.09	0.70	12.33
Huayhuash	5.55	1.27	0.11	6.92
Raura	6.17	9.65	0.54	16.36
La Viuda	22.22	23.93	0.40	46.56
Huaytapallana	7.50	11.19	0.08	18.77
Central	14.40	20.50	3.02	37.92
Chonta	0.29	10.69	0.08	11.06
Urubamba	0.72	2.49	0.11	3.32
Vilcabamba	1.77	2.93	0.11	4.81
Vilcanota	13.07	5.59	30.67	49.33
Carabaya	12.63	23.90	10.33	46.86
Apolobamba	9.17	6.80	14.91	30.88
Huanzo	0.13	1.34	0.17	1.64
Chila	0.59	0.39	-	0.97
Ampato	0.02	0.71	0.01	0.73
Total	**121.15**	**142.68**	**65.03**	**328.86**

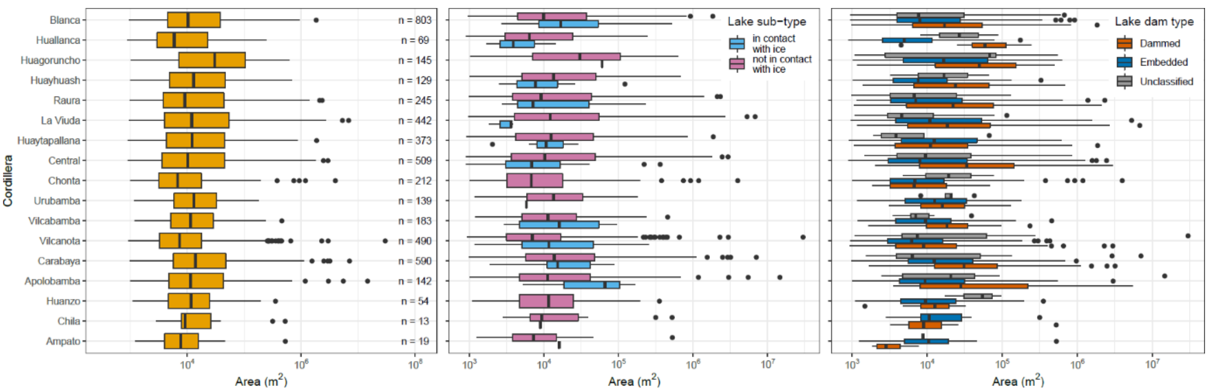

Figure 3. Lake area for the 17 glaciated cordilleras as recorded in the projectGLOP lake inventory for (A) all lakes (number of lakes included is detailed to the right of the figure); (B) lake contact with existing glaciers (total number of lakes is 145; Table 2 and Table S6); (C) lake dam type (see also Table S7). In all plots the length of the boxes (whiskers) encompass 50% (95%) of the data, points denote outliers.

Lake dam type was investigated to see if there was a significant difference between lake size recorded between different dam types (Figure 3C). Most of the significant differences in lake size were between embedded and dammed lakes (in 11 of the cordilleras; Supplementary Information 5 Table S7). In all cases, dammed lakes are significantly bigger than embedded lakes (for the 11 cordilleras where significant differences were found; Figure 3C and Table S7); this is an important finding for understanding future GLOF potential and hazard in these cordilleras.

3.2.2 projectGLOP lake elevation

Lake elevation was found to vary depending on the topography of the respective cordilleras. Overall, lakes were not found at the highest elevations of any cordillera due to limited topographic opportunity from the presence of both glaciers and steep slopes (Figure 4A). Instead, the majority of lakes were found to be constrained within a limited elevation range of between ≅ 4,500 m asl and ≅ 4,800 m asl (Figure 4A). The highest elevation recorded for any lake is an embedded lake in Ampato (5,660 m asl) whilst the lowest elevation lake is in Carabaya (at 3,686 m asl). Statistically, there is a significant difference between the lake elevation distribution and the random sample of points across the cordilleras (Kruskall-Wallace test; p-value < 0.01; Figure 4A)

In terms of lake sub-types, lakes in contact with ice were found to be at higher elevations in all cordilleras in which they are present. There is a significant difference in elevation between lakes in contact with glaciers and those which are glacier-detached in the cordilleras Blanca, Carabaya, Vilcanota, Central, Vilcabamba, Huayhuash, La Viuda, Raura and Huallanca (p-value < 0.01) and Apolobamba (p-value < 0.05) (Figure 4B and Table S8).

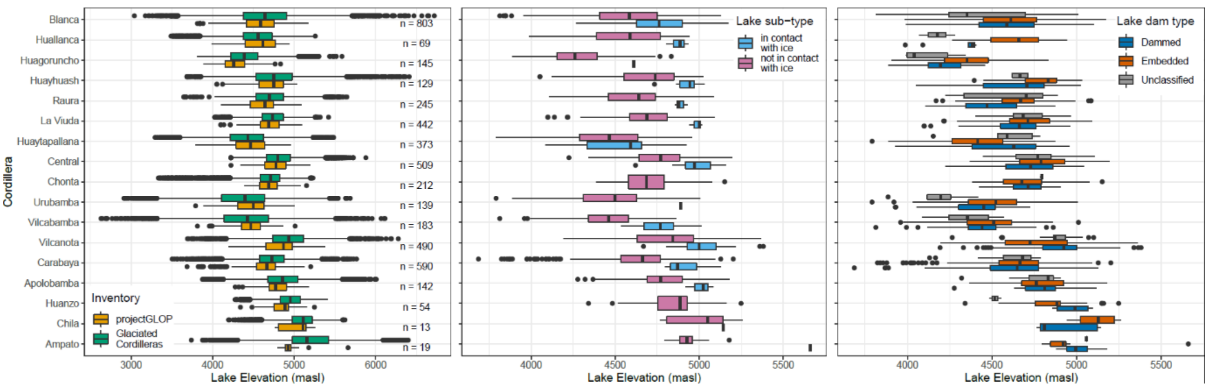

Figure 4. Lake elevation for lakes recorded in the projectGLOP inventory for (A) all lakes, with elevation estimates for each cordillera based on a random sample of *n* = 10,000 points within the 3 km glacier buffer; (B) connectivity to existing glaciers (see also Table 1 and Table S8); (C) lake dam type. The length of the boxes (whiskers) encompass 50% (95%) of the data, points denote outliers.

In terms of dam type, results reveal a significant difference in lake elevation between embedded and dammed lakes in nine of the glaciated cordilleras (Figure 4C and Table S9). In Vilcanota and Huaytapallana, dammed lakes are found at significantly higher elevations than embedded lakes; whilst embedded lakes are found at higher elevations than dammed lakes in Huayhuash, La Viuda, Raura, Huallanca, Huagoruncho and Urubamba (Figure 4C and Table S9). Some of these cordilleras have shown a >70% reduction in glacial extent since the 1960s: La Viuda (>85% loss in glacial area), Huallanca (~75%), Huagoruncho and Urubamba (<70%); this rapid deglaciation is potentially associated with the difference in elevation between the dammed and embedded lakes observed

(Figure 4C) but the relationship between dam type and elevation is complex, with distributions varying between cordillera (Figure 4C and Table S9).

3.2.3 Bathymetry and lake volume

Bathymetry data for the Cordillera Blanca (Guardamino and Drenkhan, 2016; ANA, 2014) were used to explore existing scaling relationships (Table 4) between lake depth and width (Figure 5A) and lake area and volume (Figure 5B). Both area-depth (AD) and area-volume (AV) scaling relationships were calculated for all lakes through the application of log-linear models in R Statistical Software; scaling relationships were calculated for all lakes (Figures 5A and 5B), and were then a subset based on dam type (see Supplementary Information 4 Figure S4).

Relationships derived for the AD scaling tend to be weak (low R^2; Table 4), and the AD scaling exponents calculated for this study (for all lakes) is significantly different to those derived from similar studies (outside of the 95% confidence interval; Figure S5). When lakes are subset by dam type (Figure S4), the AD relationship (here judged by the R^2 value) improves for unclassified and embedded lakes (although this is largely a function of lower lake numbers in the case of embedded lakes).

The AV scaling relationship calculated for all lakes was slightly weaker than similar studies (R^2 = 0.825; Table 4 and Figure 5B), however the scaling exponents were similar to that of other similar studies (falling largely within the 95% confidence interval; Figure 5C). A recent study for the Cordillera Blanca (Munõz et al., 2020) found the AV approach less effective than deriving volume from area and a ratio between mean lake depth and width; as mean lake depth were not available for this study, we have relied on estimates of widely applied existing relationships (e.g. Shugar et al., 2020).

Table 4. A selection of scaling relationships used to estimate lake volume and the estimates for this study; a full discussion of the results and error can be found in Cook and Quincey (2015) and Munõz et al. (2020). Where D is the mean lake depth (in metres; for Munõz et al., 2020 see notes); A is the surface area of the lake (in m^2); V is lake volume (in m^3).

Study	Region	Estimation of lake depth (m)	Estimation of lake volume (m^3)	Notes
Evans (1986)	Canada	$D = 0.035\ A^{0.5}$	$V = 0.035\ A^{1.5}$	Cited in Munõz et al. (2020).
O'Connor et al. (2001)	British Columbia		$V = 3.114\ A + 0.0001685\ A^2$	Cited in McKillop and Clague (2007) and Cook and Quincey (2015).
Huggel et al. (2002)	Global	$D = 0.104\ A^{0.42}$	$V = 0.104\ A^{1.42}$	Huggel et al. (2002) show lake depth and area are correlated for a combination of ice dammed, moraine-dammed and thermokarst lakes. D/A R^2 = 0.916. Established relationship which has been applied directly (or modified) to estimate lake volume; but not over a range of lake dam types.
Wang et al. (2012)	Himalayas	$D = 0.087\ A^{0.434}$	$V = 0.0354\ A^{1.3724}$	Cited in Munõz et al. (2020). D/A R^2 = 0.503; V/A R^2 = 0.919
Loriaux and Casassa (2013)	Global	$D = 0.2933\ A^{0.3324}$	$V = 0.2933\ A^{1.3324}$	Cited in Munõz et al. (2020). V/A R^2 = 0.96.
Cook and Quincey (2015)	Global	$D = 0.1217\ A^{0.4129}$	$V = 0.1217\ A^{1.4129}$	Based on the re-plot of data presented in Huggel et al. (2002). D/A R^2 = 0.38; V/A R^2 = 0.91.
Kapitsa et al. (2017)	Kazakhstan	$D = 0.036\ A^{0.49}$	$V = 0.036\ A^{1.49}$	Cited in Munõz et al. (2020).
Munõz et al. (2020)	Cordillera Blanca	$d = 0.041 * W + 2$	$V = A * d$	Where d is the linear regression between mean lake depth and width (Md_Wi in Munõz et al. 2020); W is the lake width.
This study	Cordillera Blanca	$D = 0.38\ A^{0.394}$	$V = 0.126\ A^{1.412}$	All lakes: D/A R^2 = 0.374; V/A R^2 = 0.825.
		$D = 0.685\ A^{0.345}$	$V = 0.249\ A^{1.364}$	Dammed: D/A R^2 = 0.333; V/A R^2 = 0.848.
		$D = 0\ A^{3.047}$	$V = 0\ A^{4.65}$	Embedded: D/A R^2 = 0.964; V/A R^2 = 0.98.
		$D = 0.004\ A^{0.765}$	$V = 0.006\ A^{1.643}$	Unclassified: D/A R^2 = 0.886; V/A R^2 = 0.824.

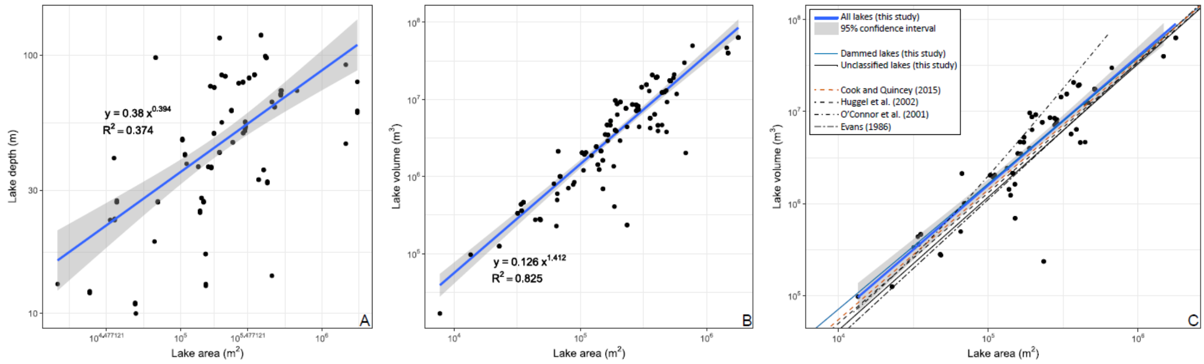

Figure 5. Scaling relationships for lakes were derived and applied across the projectGLOP inventory to estimate lake volume using the Guardamino and Drenkhan (2016) and ANA bathymetry dataset. In all figures grey bars represent 95% confidence intervals. (A) The relationship between lake depth and area (data were available for 31 lakes with a total of 117 measurements made through time). (B) Relationship between lake area and volume (data were available for 56 lakes with 170 measurements for different time periods). (C) Comparison of the scaling relationships between lake area and lake volume for this study and for other similar studies (for specific details of the relationships presented see Table 4).

The calculated AV scaling exponents (Table 4) were applied across the entire projectGLOP inventory (1) for all lakes irrespective of dam type (Figure 6), and (2) using the calculated exponents for dammed, embedded and unclassified lakes (Supplementary Information 4 Figure S5; Table 4). As with calculated lake area (3.2.1), there are a number of outliers in the data. From the calculated estimates of lake volume, lakes in the Cordillera Vilcanota contain the highest volume of water, with Apolobamba containing the second highest volume (Table 5); however, one large lake in Vilcanota (4.5 km^3) accounts for ~90% of the total water volume; in Apolobamba the largest lake (1.7 km^3) accounts for ~70% of the total; both lakes were possibly formed pre-LIAM but are located within 3 km of existing glaciers.

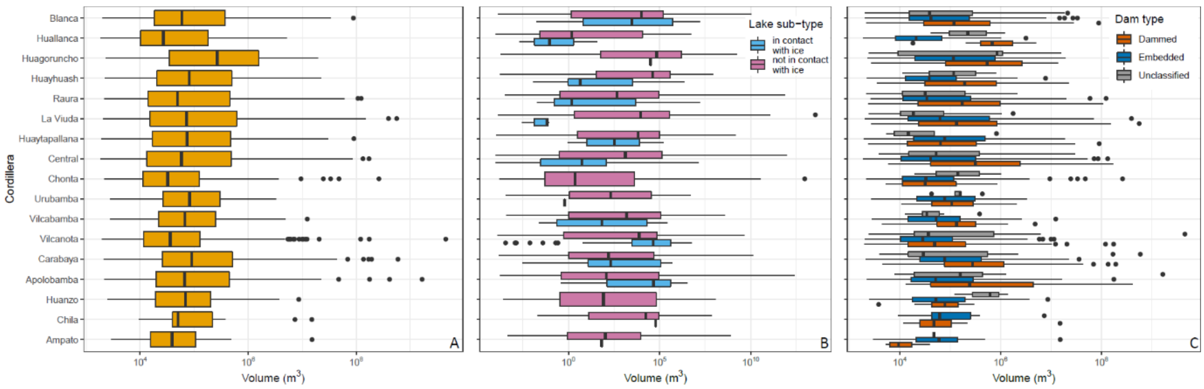

Figure 6. Lake volume for the 17 glaciated cordilleras calculated from the scaling exponents ($V = 0.126\,A^{1.412}$) derived for all lakes (Figure 5B and Table 4). Estimated lake volume is shown for (A) all lakes; (B) lakes depending on glacial connectivity; (C) lakes by dam type.

Table 5. Estimates of total lake volume for each cordillera based on the scaling exponents derived for this study (Figure 5). Presented here are the volume estimates derived from all lakes irrespective of dam type (*All lakes*), with these data also separated by dam type (*Dammed lakes*, *Embedded lakes* and *Unclassified lakes*).

Cordillera	**Volume (km³)**			
	All lakes	**Dammed lakes**	**Embedded lakes**	**Unclassified lakes**
Blanca	0.79	0.44	0.27	0.08
Huallanca	0.02	0.01	0.01	0.00
Huagoruncho	0.25	0.13	0.10	0.02
Huayhuash	0.14	0.12	0.02	0.00
Raura	0.49	0.19	0.29	0.01
La Viuda	1.87	1.03	0.84	0.00
Huaytapallana	0.40	0.20	0.19	0.00
Central	1.12	0.46	0.60	0.07
Chonta	0.42	0.00	0.42	0.00
Urubamba	0.04	0.01	0.03	0.00
Vilcabamba	0.06	0.02	0.04	0.00
Vilcanota	5.06	0.43	0.08	4.55
Carabaya	1.73	0.52	0.45	0.77
Apolobamba	2.42	0.52	0.24	1.66
Huanzo	0.02	0.00	0.02	0.00
Chila	0.02	0.02	0.01	
Ampato	0.02	0.00	0.02	0.00
Total	**14.87**	**4.09**	**3.62**	**7.17**

3.3 Lake inventory comparisons

The projectGLOP database provides a current (2019) picture of the nature of lakes across the Peruvian cordilleras, but there are also a number of other existing inventories for Peru. INAIGEM covers all 20 cordilleras, ANA covers 19, and Emmer provides an inventory for the Cordillera Blanca (Figure 1). The methods and satellite imagery (see Supplementary Information 5 Tables S1 and S2) used to collate these inventories differ, which has implications for the applicability of the inventories over different time periods as it will depend greatly on the availability of similar resolution satellite imagery to understand past lake fluctuations. Here, we use consistent methods (Supplementary Information 2.1) and data (Landsat missions) in order that the projectGLOP inventory is directly comparable through time (back to 1972). In terms of lake numbers, the INAIGEM dataset, which was digitised using high resolution satellite images (Supplementary Information 5 Table S2), has the highest number of recorded lakes (Table 6), whilst the ANA inventory consistently records the lowest lake numbers (Table 6) despite using a combination of both low (30m) and high (10m) resolution satellite images (Supplementary Information 5 Table S3).

Table 6. Lake counts for each inventory by cordillera. Data for the ANA, INAIGEM and Emmer inventories have been subset from the original data to include only lakes which occur within the 3 km buffer (Supplementary Information 2.1.1).

Cordillera	ANA	Emmer	INIAGEM	projectGLOP
Blanca	385	711	882	803
Huallanca	28	-	74	69
Huagoruncho	102	-	206	145
Huayhuash	72	-	172	129
Raura	133	-	513	245
La Viuda	212	-	583	442
Huaytapallana	192	-	614	373
Central	266	-	490	509
Chonta	96	-	127	212
Urubamba	81	-	241	139
Vilcabamba	100	-	269	183
Vilcanota	187	-	1250	490
Carabaya	366	-	661	590
Apolobamba	47	-	179	142
Huanzo	32	-	54	54
Chila	10	-	19	13
Ampato	7	-	39	19
Total	**2316**	**711**	**6373**	**4557**

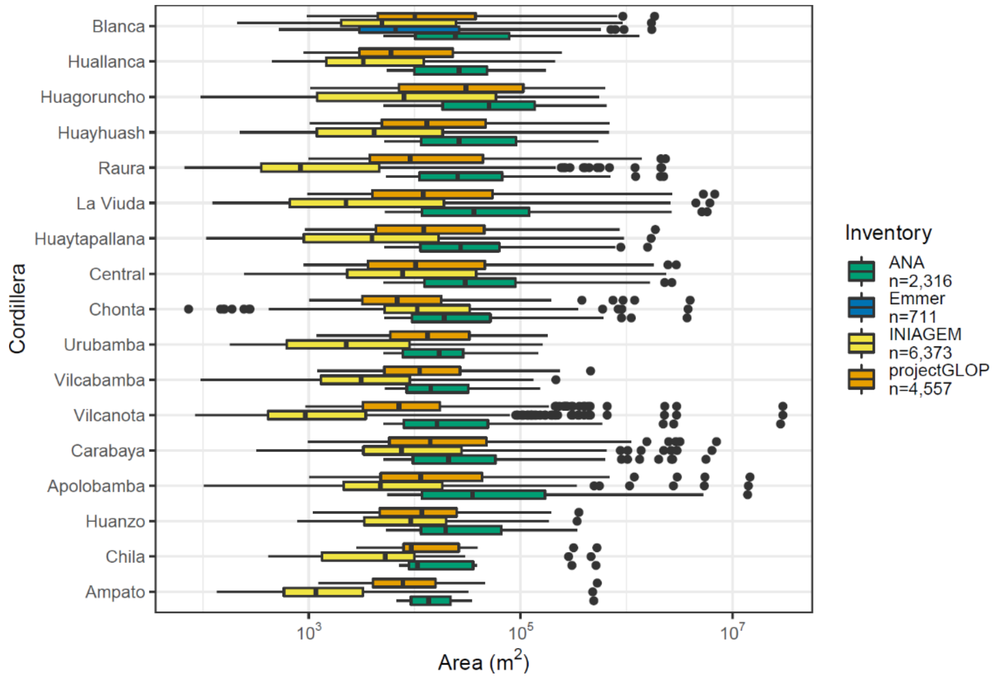

Figure 7. Across the different inventories, pairwise comparisons using Wilcoxon rank sum tests show a significant difference in lake areas recorded (p-value < 0.05 for all inventories), although this varies between cordilleras due to low numbers of lakes recorded (see Table 1 and Table 7).

The range of lake areas between each inventory is consistent with the different methods and satellite imagery used to compile each of the different lake inventories (Figure 7). Lakes in the INAIGEM inventory are consistently smaller than the ANA and projectGLOP inventories due to the higher resolution imagery being used for lake digitisation; which is matched by the higher number of lakes recorded (Table 6). Despite using a range of both high- and low-resolution satellite images, the lake areas recorded in the ANA inventory are consistently larger than the other inventories; further shown by the significant difference between the ANA inventory and most other inventories (Table 7).

Table 7. The majority of cordilleras in Peru show a significant difference (p-value < 0.05) in lake area between the different inventories (ANA, INAIGEM, Emmer and projectGLOP). There are only four cordilleras (Urubamba, Huanzo, Ampato and Chila) where there is no significant difference, possibly due to low lake numbers recorded (see also Table 6).

Urubamba	ANA	INAIGEM
INAIGEM	<0.01	-
projectGLOP	0.15	<0.01
Huanzo	ANA	INAIGEM
INAIGEM	<0.01	-
projectGLOP	<0.05	0.21
Ampato	ANA	INAIGEM
INAIGEM	<0.01	-
projectGLOP	0.19	<0.01
Chila	ANA	INAIGEM
INAIGEM	<0.05	-
projectGLOP	0.48	0.1

Lake elevation was available within the ANA and Emmer inventories. Lake elevation distributions for these and the projectGLOP inventory were compared using Wilcoxon rank sum tests; elevation was found to be significantly different between the three inventories. Reasons for this could include the different methods used to compile the inventories, differences in the number of lakes recorded (Table 6) or as a result of lakes having changed shape, size, emerged or been drained throughout the different time periods covered by each of the inventories.

Comparisons between these inventories highlight the impact of differing methodologies on the mapping of glacial lakes in the same area. For the purposes of GLOF hazard assessments, it is important that temporal records of glacier lake changes are available, which are cross-comparable (e.g. Wilson et al., 2018). We would therefore recommend:

(1) A clearly defined glacial lake sampling strategy. This needs to be based on an appropriate understanding of the glacial history of the region (which we propose based on previous literature; Supplementary Information 5 Table S1). It was clear from comparing the inventories that the number of glacier lakes included depended on the different strategies used; this needs to be consistent to facilitate future glacial lake hazard investigations.

(2) Both the spatial and temporal resolution of the satellite data need to facilitate cross-comparability. Ideally, high resolution imagery should be used, however, the temporal and spatial coverage of these data are limited compared to low resolution imagery, such as Landsat. For adequate assessment of changes in lakes through time, individual inventories need to represent distinct time stamps. To compile the ANA inventory, for example, 10 years of mixed high- and low-resolution data were needed to map 19 of the cordilleras (Supplementary Information 5 Table S3). As such we feel Landsat offers the best option in terms of both resolution and longevity, and its applicability can be augmented with the use and consultation of high-resolution options available (e.g. within Google Earth Engine).

(3) Lake inventories in high-mountain regions should use a manual mapping approach. Our use of NDWI and NDSI highlighted the advantages of automated classification methods, as they allow for the rapid mapping of large areas. However, these often require extensive and time-consuming manual correction due to issues with (e.g.) cloud cover, shadow and snow/ice effects (Shugar et al., 2020), with the potential to miss lakes altogether. Due to these issues manual methods continue to represent the most accurate and cross-comparable mapping method for long-term lake monitoring. Additionally, important metrics (Table 1), such as dam type, can only be mapped manually.

(4) The number of digitisers operating on an inventory should be limited and mapping procedures clearly communicated. To address this issue, we analysed digitised lakes from a number of different expert users for intercomparison, and produced training sites in order to reduce errors prior to lake digitisation. Our analysis of this technique highlighted the need to reduce inconsistencies in mapping between users, which is why we recommend that the number of digitisers is limited (for further information see Supplementary Information 4 Figure S2).

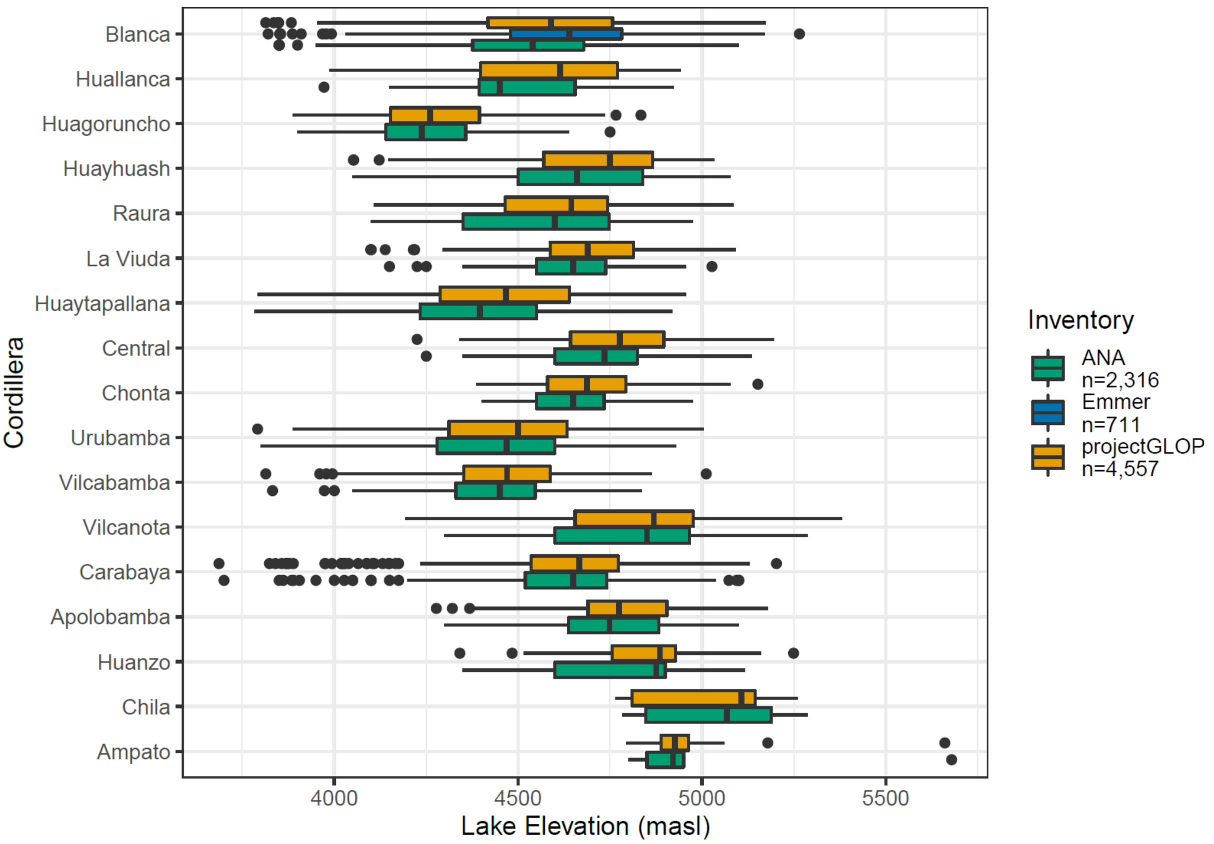

Figure 15. The ANA and Emmer inventories both include lake elevation. Lake elevation is significantly different between the different inventories (p-value <0.05).

3.4 Wider implications for glacial lake research

Our paper produces the most complete inventory to date of glacial lakes in the Peruvian glaciated mountains and we have provided a clear set of recommendations for the construction of similar glacier lake inventories (Section 3.3). This new inventory for Peru represents an important step towards a more complete understanding of the GLOF risk in Peru. While an assessment of the GLOF pattern here is beyond the scope of this paper, unpublished data (Emma et al., in preparation) show that GLOFs only affect a small number of glacial lakes in Peru (n = 150 out of 4,557 lakes). This is surprising given that Peru is seen as a global hotspot for GLOF events (e.g. Harrison et al., 2018), is in a region with a considerable record of damaging earthquakes and glacier detachment slides, and an area where sub-decadal climate events such as ENSO are common. Despite this, the vast majority of the lakes described and listed in this inventory have not produced GLOFs. There are several hypotheses that could be tested to explain this potential anomaly. First, this might be a consequence of the relatively small proportion of glacial lakes in the region dammed by unstable moraines (~28%). This might be a consequence of factors driven by climate or debris supply. Second, it may also reflect the small percentage (~3%) of lakes that are still attached to present glaciers. This might reflect the low latitude in which the glaciers have developed, and therefore the strong response of glaciers to recent climate change (e.g. Vuille et al., 2008, 2018; Jomelli et al., 2009). This means that such lakes may still be unstable and with a higher probability of failure than others; but given the small numbers of such lakes it may also suggest that the GLOF peak has passed. Third, it may be that glacier-lake systems have evolved in the region to be 'stress-hardened' to extremes such as ENSO and to earthquakes; what we are seeing now at the end of the present glacial-interglacial cycle is just the remaining lakes that have managed to survive in such an unstable environment.

Overall, the pattern of few lakes producing GLOFs may therefore highlight the likely stability of such systems to external and internal perturbations, and it calls into question the assertions from some researchers and policymakers that GLOFs will necessarily increase in frequency and magnitude in glacial mountains (see Harrison et al., 2018 for further discussion).

Finally, we now have the dataset to obtain an enhanced understanding of how glacier-lake systems evolve under conditions of climate change. While the use of such systems provides only an incomplete analogue for past deglaciation (the Peruvian Andes are tropical glacier systems which therefore provide only limited insight into other glaciated mountains which underwent deglaciation) our lake inventory will allow us to interrogate the patterns and timing of lake development.

4. Conclusions

In this paper we have presented a new glacial lake inventory for Peru which details lake distribution, extent and other important metrics, such as dam type. This dataset represents the most comprehensive inventory currently available for the Peruvian Andes. Covering an observation period between 2017 and 2019, the inventory includes 4,557 glacier lakes distributed across each of the glaciated Cordillera covering a total area of 328.86 km^2. Further analysis of these lakes revealed that the majority are now detached from current glaciers (97%) and are classified as either embedded (~67%) or (moraine) dammed (~28%) lakes, with the remaining 5% falling into the unclassified category. In terms of distribution, we found that the largest number of lakes exist in the Blanca and Carabaya cordillera (representing 18% and 13% of the total, respectively), whilst the Vilcanota and Carabaya cordillera contain the largest lake extent (representing 15% and 14% of the total, respectively). Overall, lake number, extent and type were found to vary significantly between each cordillera, which likely highlights differing topographic settings and glacier responses to recent climatic warming. Analysis of lake elevations revealed that the majority of lakes are found within a limited elevation range of between ≅ 4,500 m asl and ≅ 4,800 m, however, again this was shown to vary between the cordilleras. The information provided by this new inventory represents an important first step towards a better understanding of current glacial environments across the Peruvian cordilleras.

Comparisons of the new inventory presented here with existing lake inventories available for Peru reveal a number of inconsistencies related to differences in the source imagery used and the mapping methodology applied. Such differences represent a significant challenge when attempting to monitor lake changes through time using different data sources. To address this challenge, this paper presents a robust and easily reproducible mapping methodology that facilitates the consistent recording of glacier lakes for other locations and time periods using freely available satellite imagery (e.g. Landsat). The continual monitoring of glacial lakes in a standardised manner is of particular importance for the assessment of current and future risks associated with glacial hazards, such as GLOFs, which represent a significant socio-economic risk in Peru as well as in other mountainous regions globally.

Data availability

We have provided a .kml file of digitised lakes with this paper. Full details of all lakes in the inventory can be made available on request.

References

ArcGIS World Imagery (2020). World Imagery with Metadata. URL: https://www.arcgis.com/home/webmap/viewer.html?webmap=c1c2090ed8594e0193194b750d0d5f83 [Accessed 02/03/2021]

ANA (2014). Inventario Nacional de Glaciares y Lagunas. URL: https://hdl.handle.net/20.500.12543/199 [Accessed 15/11/2020]

Baraer, M., Mark, B. G., McKenzie, J. M., Condom, T., Bury, J., Huh, K.-I., Portocarrero, C., Gómez, J., and Rathay, S. (2012). Glacier recession and water resources in Peru's Cordillera Blanca. *Journal of Glaciology,* **58**, 134–150.

Brecher, H.H. and Thompson, L.G. (1993). Measurement of the retreat of Qori Kalis glacier in the tropical Andes of Peru by terrestrial photogrammetry. *Photogrammetric Engineering and Remote Sensing*, **59**, 1017-1017.

Burns, P. and Nolin, A. (2014). Using atmospherically-corrected Landsat imagery to measure glacier area change in the Cordillera Blanca, Peru from 1987 to 2010. *Remote Sensing of Environment*, **140**, 165–178.

Carrivick, J.L. and Tweed, F.S. (2016). A global assessment of the societal impacts of glacier outburst floods. *Global and Planetary Change*, **144**, 1-16.

Clague, J.J., Huggel, C., Korup, O. and MCguire, B. (2012). Climate change and hazardous processes in high mountains. *Revista de la Asociación Geológica Argentina*, **69(3)**, 328-338.

Cogley, G. (submitter); Kienholz, C., Miles, E., Sharp, M. and Wyatt, F. (analysts) (2015). GLIMS Glacier Database. Boulder, CO. National Snow and Ice Data Center. http://dx.doi.org/10.7265/N5V98602

Colonia, D., Torres, J., Haeberli, W., Schauwecker, S., Braendle, E., Giraldez, C. and 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(5)**, 336.

Cook, S.J. and Quincey, D.J. (2015). Estimating the volume of Alpine glacial lakes. *Earth Surface Dynamics Discussions*, **3**.

Davies, B.J. and Glasser, N.F. (2012). Accelerating shrinkage of Patagonian glaciers from the Little Ice Age (~AD 1870) to 2011. *Journal of glaciology,* **58 (212),** 1063–1084.

Drenkhan, F., Guardamino, L., Huggel, C. and Frey, H. (2018). Current and future glacier and lake assessment in the deglaciating Vilcanota-Urubamba basin, Peruvian Andes. *Global and Planetary Change*, **169**, 105-118.

Dykstra A. and Owen, L. (2017). Landsat Collections — What are Tiers? URL: https://www.usgs.gov/media/videos/landsat-collections-what-are-tiers [Accessed 20/08/2020]

Earth Engine Data Catalogue (2020). Landsat Collections. URL: https://developers.google.com/earth-engine/datasets/catalog/landsat [Accessed 21/05/2020]

Emmer, A (2018). GLOFs in the WOS: Bibliometrics, geographies and global trends of research on glacial lake outburst floods (Web of Science, 1979-2016). *Natural Hazards and Earth System Sciences*, **18(3)**, 813-827.

Emmer, A., Harrison, S., Mergili, M., Allen, S., Frey, H. and Huggel, C. (2020). 70 years of lake evolution and glacial lake outburst floods in the Cordillera Blanca (Peru) and implications for the future. *Geomorphology*, 107178.

Emmer, A., Klimeš, J., Mergili, M., Vilímek, V. and Cochachin, A. (2016). 882 lakes of the Cordillera Blanca: an inventory, classification, evolution and assessment of susceptibility to outburst floods. *Catena*, **147**, 269-279.

Emmer, A., Le Roy, M., Sattar, A., Veettil, B.K., Alcalá-Reygosa, J., Campos, N., Malecki, J. and Cochachin, A. (2021). Glacier retreat and associated processes since the Last Glacial Maximum in the Lejiamayu valley, Peruvian Andes. *Journal of South American Earth Sciences*, p.103254.

Evans, S.G. (1986). Landslide Damming in the Cordillera of Western Canada, Seattle, Washington, 111–130, 1986.

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., Ligtenberg, S. R. M., Bolch, T., Sharp, M. J., Hagen, J. O., van den Broeke, M. R., and Paul, F. (2013). A Reconciled Estimate of Glacier Contributions to Sea Level Rise: 2003 to 2009. *Science*, **340**, 852–857.

Gesch, D.B., Muller, J.-P. and Farrugia, T.J. (2006). The shuttle radar topography mission: data validation and applications. *Photogrammetric Engineering and Remote Sensing*, **72(1)**.

GLIMS (2019). GLIMS: Global Land Ice Measurements from Space - Monitoring the World's Changing Glaciers. URL: https://www.glims.org/ [Accessed 20/05/2020]

Guardamino, L. and Drenkhan, F. (2016). Evolution and potential threat of glacial lagoons in the Vilcabamba mountain range (Cusco and Apurímac, Peru) between 1991 and 2014. *Journal of Glaciers and Mountain Ecosystems*, **1**. https://doi.org/10.36580/rgem.i1.21-36

Haeberli, W., Linsbauer, A., Cochachin, A., Salazar, C. and Fischer, U.H. (2016). On the morphological characteristics of overdeepenings in high-mountain glacier beds. *Earth Surface Processes and Landforms*, **41(13)**, 1980-1990.

Hall, D.K., Riggs, G.A. and Salomonson, V.V. (1995). Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, **54(2)**, 127-140.

Hanshaw, M. N. and Bookhagen, B. (2014). Glacial areas, lake areas, and snow lines from 1975 to 2012: status of the Cordillera Vilcanota, including the Quelccaya Ice Cap, northern central Andes, Peru. The Cryosphere**, 8,** 359–376.

Harrison, S., Kargel, J.S., Huggel, C., Reynolds, J., Shugar, D.H., Betts, R.A., Emmer, A., Glasser, N., Haritashya, U.K., Klimeš, J. and Reinhardt, L. (2018). Climate change and the global pattern of moraine-dammed glacial lake outburst floods. *The Cryosphere*, **12(4)**, 1195-1209.

Hastenrath, S. and Ames, A. (1995). Recession of Yanamarey glacier in Cordillera Blanca, Peru, during the 20th century. *Journal of Glaciology*, **41(137)**, 191-196.

Huggel, C., Kääb, A., Haeberli, W., Teysseire, P., Paul, F. (2002). Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps. *Canadian Geotechnical Journal*, **39**, 316–330. http://dx.doi.org/10.1139/t01-099

INAIGEM (2017). Methodological Manual of the National Glacier Inventory. URL: https://cdn.www.gob.pe/uploads/document/file/913161/Manual_Metodologico_de_Inventario_Nacional_de_Glaciares.pdf [Accessed 20/11/2020]

INAIGEM (2018). The national inventory of glaciers: the glacial mountain ranges of Peru. URL: https://hdl.handle.net/20.500.12543/2623 [Accessed 11/11/2020]

INAIGEM (n.d.). Glacial lake inventory for Peru. Unpublished.

Jomelli, V., Favier, V., Rabatel, A., Brunstein, D., Hoffmann, G. and Francou, B. (2009). Fluctuations of glaciers in the tropical Andes over the last millenium and palaeoclimatic implications: A review. *Palaeogeography, Palaeoclimatology, Palaeoecology,* **281,** 269-282.

Jomelli, V., Grancher, D., Brunstein, D. and Solomina, O. (2008). Recalibration of the yellow Rhizocarpon growth curve in the Cordillera Blanca (Peru) and implications for LIA chronology. *Geomorphology*, **93(3-4)**, 201-212.

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 (Jetysu) Alatau, central Asia, using Landsat imagery and glacier bed topography modelling. *Natural Hazards and Earth System Sciences Discussion*, **1–54**. https://doi.org/10.5194/nhess-2017-134.

Kaser, G., Ames, A. and Zamora, M. (1990). Glacier fluctuations and climate in the Cordillera Blanca. *Annals of Glaciology*, **14**, 136-140.

King, O., Bhattacharya, A., Bhambri, R. and Bolch, T. (2019). Glacial lakes exacerbate Himalayan glacier mass loss. *Scientific Reports*, **9(1)**, 1-9.

Kochtitzky, W. (submitter); Kochtitzky, William (analyst) (2017). GLIMS Glacier Database. Boulder, CO. National Snow and Ice Data Center. http://dx.doi.org/10.7265/N5V98602

Kougkoulos, I., Cook, S.J., Jomelli, V., Clarke, L., Symeonakis, E., Dortch, J.M., Edwards, L.A. and Merad, M. (2018). Use of multi-criteria decision analysis to identify potentially dangerous glacial lakes. *Science of the Total Environment*, **621**, 1453-1466.

Kumar, R., Bahuguna, I.M., Ali, S.N. and Singh, R. (2020). Lake inventory and evolution of glacial lakes in the Nubra-Shyok basin of Karakoram Range. *Earth Systems and Environment*, **4(1)**, 57-70.

López-Moreno, J. I., Navarro, F., Izagirre, E., Alonso, E., Rico, I., Zabalza, J. and Revuelto, J. (2020). Glacier and climate evolution in the Pariacacá Mountains, Peru. *Geographical Research Letters*, **46**, 127-139.

Loriaux, T. and Casassa, G. (2013). Evolution of glacial lakes from the Northern Patagonia Icefield and terrestrial water storage in a sea-level rise context. *Global and Planetary Change*, **102**, 33-40.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S. and Gill, E. (2019). The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform. *Remote Sensing*, **11(1)**, 43. https://doi.org/10.3390/rs11010043

Márquez, A. and Francou, B. (1995). Cordillera Blanca: glaciers in history. *Bulletin de l'Institut Français d'Etudes Andines*, **24(1)**, 37-64.

Masiokas, M., Rivera, A., Espizua, L.E., Villalba, R., Delgado, S. and Aravena, J.C. (2009). Glacier fluctuations in extratropical South America during the past 1000 years. *Palaeogeography Palaeoclimatology, Palaeoecology,* **281**, 242–268.

McFeeters, S.K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, **17(7)**, 1425-1432, DOI: 10.1080/01431169608948714

McKillop, R.J. and Clague, J.J. (2007). A procedure for making objective preliminary assessments of outburst flood hazard from moraine-dammed lakes in southwestern British Columbia. *Natural Hazards*, **41**, 131–157.

Mercer, J.H. and Palacios M, O. (1977). Radiocarbon dating of the last glaciation in Peru. *Geology*, **5(10)**, 600-604.

Mernild, S. H., Liston, G. E., Hiemstra, C., and Wilson, R. (2017) The Andes Cordillera. Part III: glacier surface mass balance and contribution to sea level rise (1979–2014). International Journal of Climatology, **37**, 3154– 3174.

Mosquera, P.V., Hampel, H., Vázquez, R.F., Alonso, M. and Catalan, J. (2017). Abundance and morphometry changes across the high-mountain lake-size gradient in the tropical Andes of Southern Ecuador. *Water Resources Research*, **53(8)**, 7269-7280.

Motschmann, A., Huggel, C., Carey, M., Moulton, H., Walker-Crawford, N. and Muñoz, R. (2020). Losses and damages connected to glacier retreat in the Cordillera Blanca, Peru. *Climatic Change*, **162**, 837-858.

Muñoz, R., Huggel, C., Frey, H., Cochachin, A. and Haeberli, W. (2020). Glacial lake depth and volume estimation based on a large bathymetric dataset from the Cordillera Blanca, Peru. *Earth Surface Processes and Landforms*. https://doi.org/10.1002/esp.4826

NASA (2020). Landsat Science: Landsat 4. URL: https://landsat.gsfc.nasa.gov/landsat-4-2/ [Accessed 21/05/2020]

NextGIS (2015). QuickMapServices: easy basemaps in QGIS. URL: https://nextgis.com/blog/quickmapservices/ [Accessed 20/05/2020]

O'Connor, J.E., Hardison III, J.H. and Costa, J.E. (2001). Debris Flows from Failures of Neoglacial-Age Moraine Dams in the Three Sisters and Mount Jeerson Wilderness Areas, Oregon. US Geological Survey Professional Paper 1606, Reston, Virginia, p. 105.

Oliver-Smith, A. (1979). The Yungay avalanche of 1970: Anthropological perspectives on disaster and social change. *Disasters*, ***3*(1)**, 95-101.

Quincey, D.J., Richardson, S.D., Luckman, A., Lucas, R.M., Reynolds, J.M., Hambrey, M.J. and Glasser, N.F. (2007). Early recognition of glacial lake hazards in the Himalaya using remote sensing datasets. *Global and Planetary Change*, **56(1-2)**, 137-152.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Rabatel, A., Francou, B., Soruco, Á., Gomez, J., Cáceres, B., Ceballos, J.L., Basantes, R., Vuille, M., Sicart, J.E., Huggel, C. and Scheel, M. (2013). Current state of glaciers in the tropical Andes: a multi-century perspective on glacier evolution and climate change. *The Cryosphere*, **7(1)**, 81-102.

Racoviteanu, A. (submitter); Racoviteanu, A. (analyst) (2005, 2007). GLIMS Glacier Database. Boulder, CO. National Snow and Ice Data Center. http://dx.doi.org/10.7265/N5V98602

Raup, B.H., Racoviteanu, A., Khalsa, S.J.S., Helm, C., Armstrong, R and Arnaud, Y. (2007). The GLIMS Geospatial Glacier Database: A New Tool for Studying Glacier Change. *Global and Planetary Change*, **56**, 101-110. DOI: 10.1016/j.gloplacha.2006.07.018.

Reynolds, J.M. (1992). The identification and mitigation of glacier-related hazards: examples from the Cordillera Blanca, Peru. In: McCall, G.J.H., Laming, D.C.J. and Scott, S. (eds), *Geohazards*, London, Chapman & Hall, pp. 143-157.

Reynolds, J.M. (2014). Assessing glacial hazards for hydro development in the Himalayas, Hindu Kush and Karakorum. *International Journal of Hydropower and Dams*, **2**, 60-65.

Richardson, S.D. and Reynolds, J.M. (2000). An overview of glacial hazards in the Himalayas. *Quaternary International*, **65**, 31-47.

Salzmann, N., Huggel, C., Rohrer, M., Silverio, W., Mark, B. G., Burns, P., and Portocarrero, C. (2013). Glacier changes and climate trends derived from multiple sources in the data scarce Cordillera Vilcanota region, southern Peruvian Andes. *The Cryosphere*, **7**, 103–118.

Schauwecker, S., Rohrer, M., Huggel, C., Endries, J., Montoya, N., Neukom, R., Perry, B., Salzmann, N., Schwarb, M. and Suarez, W. (2017). The freezing level in the tropical Andes, Peru: An indicator for present and future glacier extents. *Journal of Geophysical Research: Atmospheres*, **122(10)**, 5172-5189.

Seehaus, T., Malz, P., Sommer, C., Lippl, S., Cochachin, A. and Braun, M. (2019). Changes of the tropical glaciers throughout Peru between 2000 and 2016 - mass balance and area fluctuations. *The Cryosphere*, **13**, 2537-2556.

Seimon, T.A., Seimon, A., Daszak, P., Halloy, S.R., Schloegel, L.M., Aguilar, C.A., Sowell, P., Hyatt, A.D., Konecky, B. and Simmons, J.E. (2007). Upward range extension of Andean anurans and chytridiomycosis to extreme elevations in response to tropical deglaciation. *Global Change Biology*, **13(1)**, 288-299.

Shugar, D., Burr, A., Haritashya, U., Kargel, J., Watson, S., Bevington, A., Steiner, N., Betts, R., Harrison, S., Strattman, K. and Kennedy, M. (2019). Where are the world's glacial lakes and how big are they? In *Geophysical Research Abstracts*, **21**.

Shugar, D.H., Burr, A., Haritashya, U.K., Kargel, J.S., Watson, C.S., Kennedy, M.C., Bevington, A.R., Betts, R.A., Harrison, S. and Strattman, K. (2020). Rapid worldwide growth of glacial lakes since 1990. *Nature Climate Change*. https://doi.org/10.1038/s41558-020-0855-4.

Silverio, W. (2018). Impact of Climate Change on Mount Coropuna (Cordillera Ampato, Arequipa, Peru) and on Water Resources. *Revista de Glaciares y Ecosistemas de Montaña*, **4**, 43-56. URL: https://revista.inaigem.gob.pe/index.php/RGEM/article/view/33/33 [Accessed 11/11/2020]

Silverio, W. and Jaquet, J. M. (2017). Evaluating glacier fluctuations in Cordillera Blanca (Peru)**,** *Archives des Sciences***, 69**, 145-162.

Solomina, O., Jomelli, V., Kaser, G., Ames, A., Berger, B. and Pouyaud, B. (2007). Lichenometry in the Cordillera Blanca, Peru: “Little Ice Age” moraine chronology. *Global and Planetary Change***, 59,** 225-235.

Thompson, L.G., Mosley-Thompson, E., Brecher, H., Davis, M., León, B., Les, D., Lin, P.N., Mashiotta, T. and Mountain, K. (2006). Abrupt tropical climate change: *Past and present. Proceedings of the National Academy of Sciences*, **103(28)**, 10536-10543.

Thompson, L.G., Mosley-Thompson, E., Dansgaard, W., Grootes, P.M., (1986). The Little Ice Age as recorded in the stratigraphy of the tropical Quelccaya Ice Cap. *Science*, **234**, 361–364.

UGRH: Inventario de glaciares del Peru. (2014). URL: http://groundwater.sdsu.edu/INVENTARIO_GLACIARES_ANA.pdf [Accessed 02/11/2020]

UNEP (2007). Global Outlook for Ice and Snow, UNEP, pp. 235, 2007. URL: http://hdl.handle.net/20.500.11822/7792 [Accessed 15/11/2020]

Verpoorter, C., Kutser, T., Seekell, D.A. and Tranvik, L.J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, **41(18)**. https://doi.org/10.1002/2014GL060641

Viani, C., Giardino, M., Huggel, C., Perotti, L., Mortara, G. (2016). An overview of glacier lakes in the Western Italian Alps from 1927 to 2014 based on multiple data sources (historical maps, orthophotos and reports of the glaciological surveys). *Geografia Fisica e Dinamica Quaternaria*, **39(2)**, 203-214.

Vilímek, V., Klimeš, J. and Červená, L. (2016). Glacier-related landforms and glacial lakes in Huascarán National Park, Peru. *Journal of Maps*, **12(1)**, 193-202.

Vuille, M., Francou, B., Wagnon, P., Juen, I., Kaser, G., Mark, B.G. and Bradley, R.S. (2008). Climate change and tropical Andean glaciers: Past, present and future. *Earth-science reviews*, **89(3-4)**, 79-96.

Vuille, M., Carey, M., Huggel, C., Buytaert, W., Rabatel, A., Jacobsen, D., Soruco, A., Villacis, M., Yarleque, C, Timm, O. E., Condom, T., Salzmann, N. and Sicart, J-E. (2018). Rapid decline of snow and ice in the tropical Andes - Impacts, uncertainties and challenges ahead. *Earth-Science Reviews*, **176**, 195-213.

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. *Natural Hazards and Earth System Sciences***, 12**, 3109–3122. https://doi.org/10.5194/nhess-12-3109-2012

Wang, X., Guo, X., Yang, C., Liu, Q., Wei, J., Zhang, Y., Liu, S., Zhang, Y., Jiang, Z. and Tang, Z. (2020). Glacial lake inventory of High Mountain Asia (1990–2018) derived from Landsat images. *Earth System Science Data Discussions*, 1-23.

Wegner, S. (2014). LO QUE EL AGUA SE LLEVÓ: Consecuencias y Lecciones del Aluvión de Huaraz de 1941. Proyecto IMACC - Ministerio del Ambiente, Perú. URL: https://archive.org/details/NotaTecnica7 [Accessed 20/11/2020]

WGMS (2020). Fluctuations of Glaciers Database. World Glacier Monitoring Service, Zurich, Switzerland. DOI:10.5904/wgms-fog-2020-08. URL: http://dx.doi.org/10.5904/wgms-fog-2020-08

WGMS (n.d.). Glacier Monitoring: Peru. URL: https://wgms.ch/downloads/cp/cp_Peru.pdf [Accessed 15/11/2020]

Wilson, R., Glasser, N.F., Reynolds, J.M., Harrison, S., Anacona, P.I., Schaefer, M. and Shannon, S. (2018). Glacial lakes of the Central and Patagonian Andes. *Global and Planetary Change*, **162**, 275-291.

Worni, R., Huggel, C. and Stoffel, M. (2013). Glacial lakes in the Indian Himalayas—From an area-wide glacial lake inventory to on-site and modeling based risk assessment of critical glacial lakes. *Science of the Total Environment*, **468**, S71-S84.

Supplementary Information

SI 1: Existing lake inventories in Peru

The INAIGEM lake inventory was compiled following the guidelines established in the Methodological Manual of the National Glacier Inventory (INAIGEM, 2017). A combination of Sentinel-2, Landsat, Google Earth, BingMaps and SAS Planet satellite images from 2016 were used to digitise lakes. The dataset have been subject to the 3 km buffer used for the projectGLOP inventory with a total of 8,577 lakes included across 18 of the cordilleras (Supplementary Information 5 Table S2).

Lakes in the ANA lake inventory were digitised using SPOT 4, SPOT 5, ASTER, LISS III, and Landsat 5 TM satellite data (Supplementary Information 5 Table S3). These lakes were subset from the original dataset using the 3 km glacier buffer with a total of 8,355 lakes being included.

The Emmer inventory includes a total of 822 lakes and was created using historic aerial images (1948, 1962 and 1970), Landsat images (1990, 2000) and high-resolution optical images available through Google Earth Pro (2010 and 2018). This inventory was also subsetted using the 3 km buffer to include a total of 711 lakes.

Whilst these inventories are important resources, we argue here that varying methods used to generate these inventories result in limitations in using these across a variety of applications; importantly for understanding GLOF hazards across the Peruvian cordilleras.

SI 2: Methods

The following sections detail the methods used to create a new glacial lake inventory for Peru (projectGLOP). In order that the inventory is useful across a range of applications, we suggest that a number of metrics are consistently recorded, and that the selection of appropriate satellite imagery should consider the period of time covered by the sensor. The following methodology has been applied across Peru but can be translated to other montaine environments which are (or have been) glaciated.

2.1 projectGLOP lake inventory

2.1.1 Delineation of study site

There are a total of 20 cordilleras in Peru; only 17 remain glaciated according to the GLIMS/Randolph Glacier Inventory (v6.0; GLIMS, 2019; Raup et al., 2007; or 18 indicated by INAIGEM, 2018); outside of these there are also a number of smaller regions where glaciers still persist (Figure 1). The Little Ice Age Maximum (LIAM) represents the most recent significant extension of modern glaciers, and there are a wealth of reviews showing the impact of the subsequent glacial recession on geomorphological and hydrological hazards (Oliver-Smith, 1979; Reynolds, 1992; Clague et al., 2012; Verpoorter et al., 2014; Wegner 2014; Carrivick and Tweed, 2016; Emmer et al., 2016; Harrison et al., 2018; Kougkoulos et al., 2018; Emmer et al., 2020), and implications for water resources (Baraer et al., 2012). It has been shown that significant hazards, such as GLOFs, can occur on timescales ranging from a few years to a few centuries following deglaciation (depending on the topography, dam type and the climate; Harrison et al., 2018); we therefore delineate areas which are important for GLOF hazards by buffering modern-day glaciers based on LIAM limits (Supplementary Information 5 Table S1). Using the GLIMS/Randolph Glacier Inventory (v6.0; GLIMS, 2019; Raup et al., 2007), we applied a ≅3 km buffer (WGS 84 0.03° Latitude/Longitude buffer) to existing glaciers (Cogley, 2015; Kochtitzky, 2017; Racoviteanu, 2005, 2007) using the Fixed Distance Buffer in Quantum Geographic Information System (QGIS); 3 km was selected as the literature shows that the regional LIAM did not extend beyond this limit (Supplementary Information 5 Table S1) and so for mapping, all LIAM lakes would be included by this buffer. Differences in the extent of the 3 km buffer may be present due to the varying topography of the cordilleras, but the impact of this on lake sampling is mitigated by the increased extent of the buffer (from the estimated glacial retreat of ≈2,400 m; Supplementary Information 5 Table S1).

The Glacier Little Ice Age Maximum (LIAM) represents the most recent significant glacial advance, and glaciers have receded by around ≈2,400m since this time (Supplementary Information 5 Table S1). Glacial recession leaves behind steep slope, unstable terrain, and lakes can form behind lateral and terminal moraines, and in over steepened valley bottoms. We therefore argue that lakes which have formed since the LIAM are of great import to understanding GLOF hazards, and so focus mapping efforts in these regions. We omit supraglacial lakes as these are affected by seasonal changes and can represent temporary phases in lake evolution.

2.1.2 Satellite data

Landsat Thematic Mapper Tier 1 (TMT1; see https://www.usgs.gov/media/videos/landsat-collections-what-are-tiers [Accessed: 20 Aug, 2020]) data are available at 30m resolution from 1982 (Landsat 4) to present (Landsat 8; NASA, 2020), providing a 38-year time series. In addition, Landsat TMT1 Panchromatic data are also available at 15m resolution from 1999 (Landsat 7; Earth Engine Data Catalogue, 2020) to facilitate mapping. This consistency in resolution (30m) and validation (i.e. by using TMT1 data) provides the end-user assurances with respect to accuracy when considering lake evolution through time.

Landsat TMT1 data were used to remotely identify and map lakes across the Peruvian cordilleras (Figure 1). Google Earth Engine (GEE) was used to identify and download cloud-free (<5% or <10% depending on availability) Landsat 8 images for 2019 (during the dry season from May to

September; Supplementary Information 3 and Supplementary Information 4 Figure S3a). Where cloud-free images were not available, full year (2019) images were initially sought, then 2018 (Supplementary Information 4 Figure S3b) or 2017 (Supplementary Information 4 Figure S3c) data were used. 30 m RGB and 15 m Panchromatic Landsat 8 TMT1 data were compiled as median composite images (from May to September, and also full year composites) in GEE to reduce the effect of seasonal fluctuations in lake extent (Supplementary Information 3).

NDWI (Eq. 1; McFeeters, 1996) and NDSI (Eq. 2; Hall et al., 1995) were calculated in GEE using available cloud free Landsat images to facilitate lake digitisation and identification (Supplementary Information 3; Figure 2E). Shuttle Radar Topography Mission 30 m resolution (SRTM30) data were also downloaded for the whole of Peru in order to obtain lake elevation.

$$NDWI = \frac{TM_{Band\ 3} - TM_{Band\ 5}}{TM_{Band\ 3} + TM_{Band\ 5}} \quad \text{[Eq. 1]}$$

where $TM_{Band\ 3}$ is green and $TM_{Band\ 5}$ is near infrared (McFeeters, 1996).

$$NDSI = \frac{TM_{Band\ 3} - TM_{Band\ 6}}{TM_{Band\ 3} + TM_{Band\ 6}} \quad \text{[Eq. 2]}$$

where $TM_{Band\ 3}$ is green and $TM_{Band\ 6}$ is shortwave infrared (Hall et al., 1995).

2.1.3 Lake digitisation and metrics

The Landsat 8 TMT1 median composite images (RGB, Panchromatic, NDWI and NDSI; Supplementary Information 2.1.2) were downloaded and imported into QGIS for lake digitisation. A 0.05° latitude/longitude grid (Figure 2) was superimposed over the glaciated cordilleras and used to facilitate systematic lake mapping. Lakes were identified in QGIS using the high resolution ESRI Satellite, Bing Satellite and Google Hybrid images obtained through the QuickMapServices plugin (NextGIS, 2015; Figure 2B-D), and mapped at 30m resolution using the Landsat 8 TMT1 datasets (RGB, Figure 2A; Panchromatic; NDWI and NDSI, Figure 2E). Layer transparency was adjusted to help with lake identification (e.g. Figure 2E) and lakes were only mapped where they could be clearly identified in the Landsat images (Figure 2F).

The projectGLOP inventory is stored as a polygon shapefile, with important metrics being manually input into the attribute table for each lake in turn; these include lake type[1] (dammed, embedded or unclassified), lake sub type (in contact or not in contact with ice), outflow, year, digitiser and other notes (Table 1). A unique ID was assigned to each lake in QGIS.

[1] Terms used to define lake type are: "dammed" where the lake is dammed behind a moraine, landslide or debris; "embedded" for lakes which have formed behind bedrock, irrespective of process; "unclassified" is used for the few instances where lake dam type cannot be definitively identified.

For each lake in the projectGLOP inventory, mean elevation was calculated from the SRTM30 data in QGIS using the raster Zonal Statistics plugin to sample lake elevation (Table 1; ele30mean). As the lakes were mapped in the WGS 84 projection (EPSG: 4326; in line with the SRTM and Landsat datasets), lakes were divided and then reprojected (from EPSG: 4326), based on location, to either UTM zone 18S (EPSG: 5387) or UTM zone 19S (EPSG: 5389). Lake area (area_utm) was calculated in meters for projectGLOP lakes using the QGIS Field Calculator (Table 1), and subsequently reprojected back to EPSG: 4326.

All inventories (ANA, Emmer, INAIGEM and projectGLOP) were also combined into a single unified lake inventory (as a .shp file), with all columns from all datasets being included (from the source .dbf files). This was to facilitate the analyses detailed in Supplementary Information 2.2.

2.1.4 Mapping resolution

Mapping resolution is an important consideration for water resources and hazard research for determining (e.g.) lake size and volume. We selected Landsat TMT1 data for this study as this will allow for a consistent comparison through time over a long time period (as the TMT1 data are available from 1982); we felt that this was important for longer term studies into lake evolution and stability. However, Landsat TMT1 data are relatively low-resolution (30 m), and so we potentially propagate errors in mapping due to this low resolution, throughout the inventory. We quantified the impact of this by comparing high-resolution (~0.5 m) satellite imagery with the Landsat data used in the projectGLOP inventory (Supplementary Information 2.1.3). For the high-resolution data we selected ESRI Satellite data provided by Maxar between 13th April 2019 and 4th July 2020 at 0.31 m - 0.50 m resolution (ArcGIS World Imagery, 2020). Lakes in the Cordillera Ampato were digitised, and these high-resolution digitised lakes were then compared with the same lakes recorded in the projectGLOP inventory. We calculated both the proportional error and Root Mean Square Error (RMSE) for all lakes in the Cordillera Ampato to understand any bias associated with mapping resolution.

From this analysis, a one sample t-test showed that lake size is generally overestimated by an average of ~68% when using low- compared with high-resolution imagery ($t = 2.91$, $df = 18$, p-value < 0.01); but this varies with lake area, with larger lakes being overestimated by ~19% compared with the smallest lakes by ~116% (represented by the lower and upper 95% confidence intervals). This proportional error translates to a Root Mean Squared Error (RMSE) of 2637 m^2 ($t = 5.3695$, $df = 18$, p-value < 0.01); but this again varies, with most lakes being overestimated by between ~1,605 m^2 (~2 Landsat pixels) and 3,669 m^2 (~4 Landsat pixels; Supplementary Information 5 Figure S2). Whilst these errors are seemingly large, the value of this dataset is in identifying all lakes within proximity of existing glaciers and in providing a first pass assessment of important lake features and metrics across Peru.

As we are using 30m Landsat data it is important to understand the proportion of lakes potentially missed by using this low-resolution satellite imagery. In order to test this, we additionally investigated at the ANA and INAIGEM inventories (both of which use higher resolution satellite imagery for lake digitisation; Supplementary Information 5 Tables S2 and S3) to quantify the influence of mapping resolution on recorded lake size (Supplementary Information 2.1.3). To do this we calculated the number of lakes in each inventory which was above (and below) a 900 m^2 area threshold both in terms of lakes numbers and the area that these account for in the inventories.

From this we found that the smallest lake recorded in the ANA inventory is 1,505.9 m^2, so despite the higher mapping resolution, small lakes (<900 m^2) are not recorded. Within the INAIGEM inventory, 1,560 lakes are smaller than 900 m^2, accounting for <25% of the total number of lakes. However, in terms of total lake area, lakes smaller than 900 m^2 made up only 0.25% of the total lake area recorded across the Peruvian cordilleras (Supplementary Information 5 Table S4). Whilst the number of small lakes contributes a significant portion of the total inventory, the area covered by these small lakes, and thus the volume, makes up a very small portion. We consider mapping resolution to be less important than obtaining longer and more consistent time series data for understanding GLOF hazards.

Although consistent methods were used to digitise the projectGLOP lake inventory, it is important to quantify errors that may propagate due to the individual bias during lake digitisation. For this analysis, we asked two additional experts and one non-expert user to digitise all lakes in the Cordillera Ampato using the same 30m Landsat data for 2019. The distribution of lake areas were then compared to understand individual effects on lake mapping; both the proportional error and the RMSE were calculated to quantify mapping biases.

We found that the average standard deviation of lake areas between the digitised lake outlines to be ~4,609 m^2 (~4.5 Landsat pixels; Supplementary Information 5 Figure S2). We additionally found the mean proportional error of the different lake outlines to vary between ~5% (generally for the larger lakes) up to ~45% (for the smallest lakes).

2.2 Lake inventory statistics and comparisons

The following analyses look at lake area (2.2.1), lake elevation (2.2.2) and finally lake bathymetry for the projectGLOP lake inventory, including estimates of total water content across Peru (2.2.3). As there are a number of existing lake inventories for the whole of Peru and for the Cordillera Blanca (Supplementary Information 1), we wanted to understand similarities between the datasets, how the different methods of inventory compilation (such as mapping resolution) affected the different lakes distributions across the cordilleras, and so we also include a comparison of the datasets (3.3).

2.2.1 Lake area

Lake area for each of the glaciated cordilleras was compared to ascertain any variation in lake size; boxplots were created for each cordillera in R-Statistical software (henceforth R; R Core Team, 2019). Lake area distributions were then compared using Wilcoxon rank sum tests to see if there was a significant difference in lake area between the different cordilleras.

Understanding lake size in relation to connectivity to glaciers (whether they were in contact with active ice, or not) was also investigated for the projectGLOP inventory. Kruskall-Wallace tests were conducted in R to show whether there was a significant difference in lake area between the lakes in contact with ice and those which were disconnected. Lakes were additionally subdivided by dam type, and pairwise comparisons using Wilcoxon rank sum tests were performed in R to see whether there was a significant difference in lake area depending on the dam type.

2.2.2 Lake elevation

SRTM30 data were sampled into the lake inventory in QGIS. The errors associated with SRTM30 data for South America (compared with ground-truth data) are 9 m (horizontally) and 6.2 m (absolute vertical, and 5.5 m relative vertical) (Gesch et al., 2006). Lake elevation for each of the glaciated cordilleras was compared to ascertain regional variation in lake elevation; boxplots were created for each cordillera in R. A random sample of points (n = 10,000) was taken from within the 3 km glacier buffer to compare regional elevation with lake elevation. Kruskall-Wallace tests were applied to see whether lakes occur at the same elevation distribution as the cordillera in which they were mapped.

We also looked at the difference in elevation between lakes connected (or not) to existing glaciers; Kruskall-Wallace tests were used to show whether there was a significant difference in lake elevation between the two groups. Lakes were then subdivided by dam type (either embedded, dammed or unclassified), and pairwise comparisons using Wilcoxon rank sum tests were performed to see whether there was a significant difference in lake elevation depending on the dam type.

2.2.3 Bathymetry and lake volume

Scaling relationships between lake area and depth (AD), and area and volume (AV), have been derived and used to estimate glacier lake volume for a number of regions (e.g. Cook and Quincey, 2015; Munõz et al., 2020) and globally (e.g. Shugar et al., 2020). Empirical AD and AV relationships are derived by fitting power-law functions to the data and have been performed on a number of geographic areas. Available bathymetry data for the Cordillera Blanca (Guardamino and Drenkhan, 2016; ANA, 2014) were used to define and compare these empirical relationships; lake depth and area were available for 31 lakes (with a total of 117 measurements for different time periods); 56 lakes (with a total of 170 measurements for different time periods) were available to approximate the volume and area scaling relationship.

Log-linear models were applied, to the AD and AV data, to estimate the scaling exponents (Cook and Quincey, 2015; Munõz et al., 2020); firstly, for all data, and then depending on dam type (embedded, dammed or unclassified). The scaling relationship between lake volume and lake area was then compared with other published lake bathymetry studies (Evans; 1986; O'Connor et al., 2001; Huggel et al., 2002; Cook and Quincey, 2015). Finally, the derived scaling exponents for the projectGLOP inventory were applied to all lakes to estimate lake volume across the glaciated cordilleras.

2.2.4 Lake inventory comparisons

The ANA inventory includes estimates of lake area, while the methods used to derive lake area for the projectGLOP inventory (Supplementary Information 2.1.3) were applied to both the INAIGEM and Emmer inventories; the calculated (or existing) area data were then compared using Wilcoxon rank sum tests to test for significant differences between areas calculated/recorded across the inventories.

Lake elevation is given in both the ANA and Emmer inventories. Elevation data for the projectGLOP inventory were compared with elevation across the ANA and Emmer lake inventories. Pairwise comparisons using Wilcoxon rank sum tests were used to assess differences in elevation between the inventories.

SI 3: Google Earth Engine code used for selecting and downloading Landsat 8 images and for calculating NDSI and NDWI.

```
// Define image collections
var l8p = ee.ImageCollection('LANDSAT/LC08/C01/T1_TOA') // panchromatic
 .select(['B8']);
var l8 = ee.ImageCollection('LANDSAT/LC08/C01/T1_SR') // RGB
 .select(['B4', 'B3', 'B2']);
var n8 = ee.ImageCollection('LANDSAT/LC08/C01/T1_SR'); // NDWI and NDSI

// panchromatic
var l8pfiltered = l8p.filter(ee.Filter.calendarRange(2019,2019,'year'))
 .filter(ee.Filter.calendarRange(5,9,'month'))
 .filterMetadata('CLOUD_COVER_LAND','less_than',5);

// RGB
var l8filtered = l8.filter(ee.Filter.calendarRange(2019,2019,'year'))
 .filter(ee.Filter.calendarRange(5,9,'month'))
 .filterMetadata('CLOUD_COVER_LAND','less_than',5);

// NDWI and NDSI
var n8filtered = n8.filter(ee.Filter.calendarRange(2019,2019,'year'))
 .filter(ee.Filter.calendarRange(5,9,'month'))
 .filterMetadata('CLOUD_COVER_LAND','less_than',5);

// Median composites
var l8pmedian = l8pfiltered.median(); // panchromatic
var l8median = l8filtered.median(); // RGB
var n8median = n8filtered.median(); // NDWI and NDSI

// Compute the Normalized Difference WATER Index (NDWI)
var nir = n8median.select('B5');
var grn = n8median.select('B3');
var ndwi = grn.subtract(nir).divide(grn.add(nir)).rename('NDWI');

// Compute the Normalized Difference SNOW Index (NDSI)
var sir = n8median.select('B6');
var ndsi = grn.subtract(sir).divide(grn.add(sir)).rename('NDSI');
```

SI 4: Figures

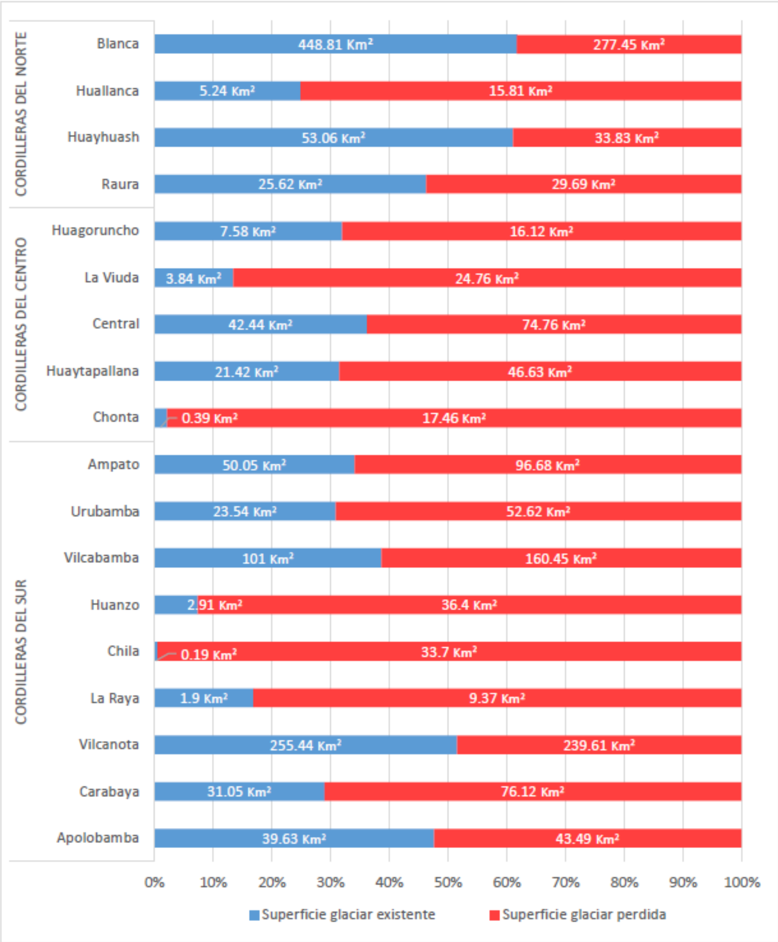

Figure S1. Showing the glacial recession in the Peruvian cordilleras (INAIGEM, 2018, p. 55).

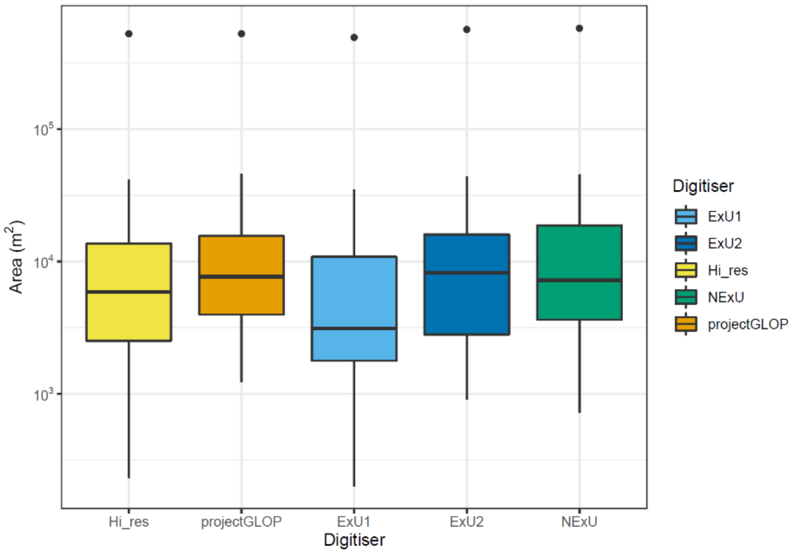

Figure S2. Lakes were digitised for the Cordillera Ampato using high-resolution satellite imagery (Hi_res) for comparison with the 30 m Landsat digitised projectGLOP lake inventory. Also, two experts (ExU1, ExU2) and a non-expert (NExU) digitised the same lakes to quantify biases in the projectGLOP lake inventory. The length of the boxes (whiskers) encompass 50% (95%) of the data, points denote outliers.

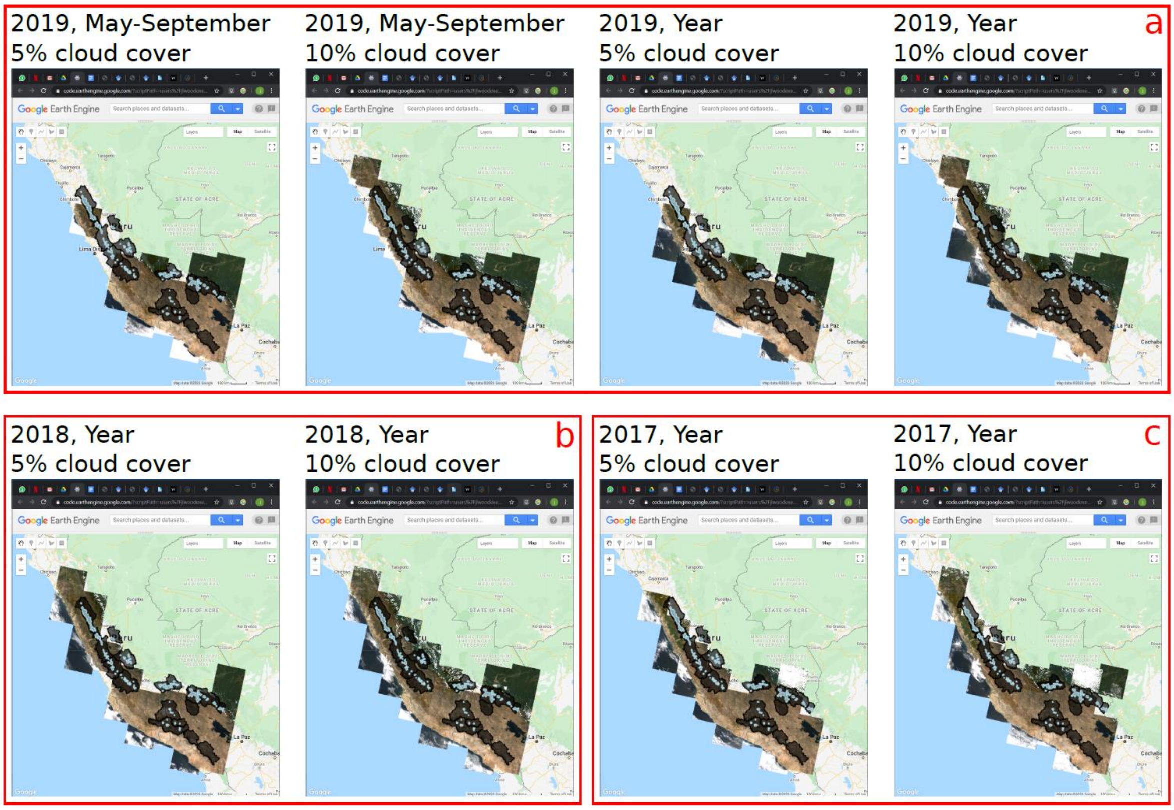

Figure S3. Showing the coverage of Landsat 8 data for Peru (a) 2019, b) 2018, c) 2017). A total of 4,042 lakes were mapped using the 2019 data, 389 were mapped using 2018 data (30 in Huaytapallana, seven in Apolobamba and, two in Carabaya, and 30 in Huaytapallana), 187 were mapped using 2017 data (six in Vilcabamba and 11 in Urubamba, six in Vilcabamba and one in Vilcabamba).

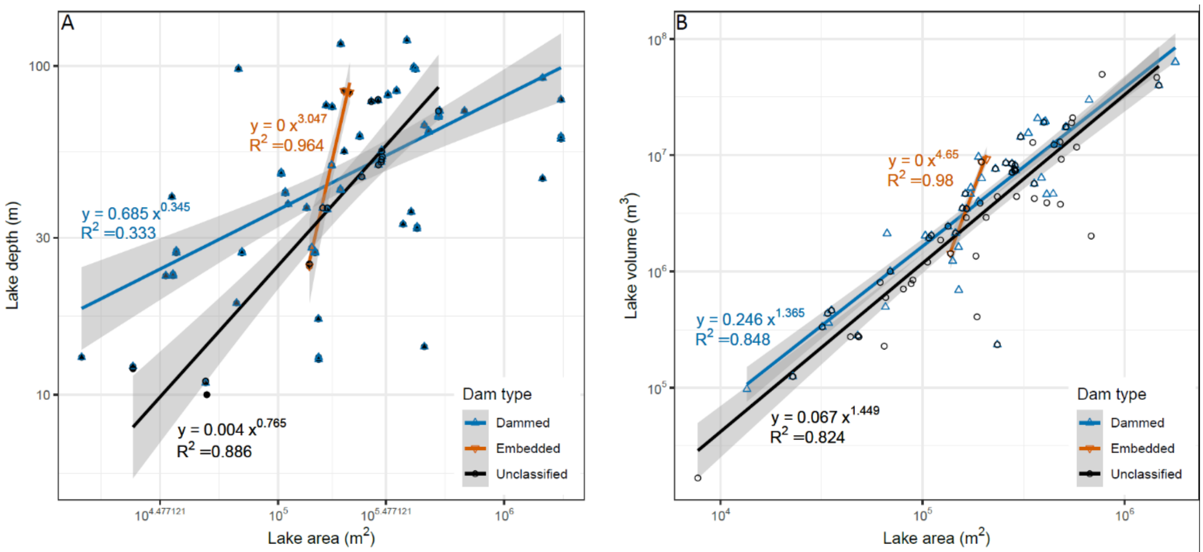

Figure S4. Scaling relationships derived from the Guardamino and Drenkhan (2016) and ANA bathymetry dataset for the Cordillera Blanca. (A) Area-Depth (data were available for 31 lakes, with a total of 117 measurements through time) and (B) Area-Volume (data were available for 31 lakes, with a total of 120 measurements through time). Lakes have been partitioned by dam type. Scaling exponents and R^2 values are given. Grey bars represent 95% confidence intervals.

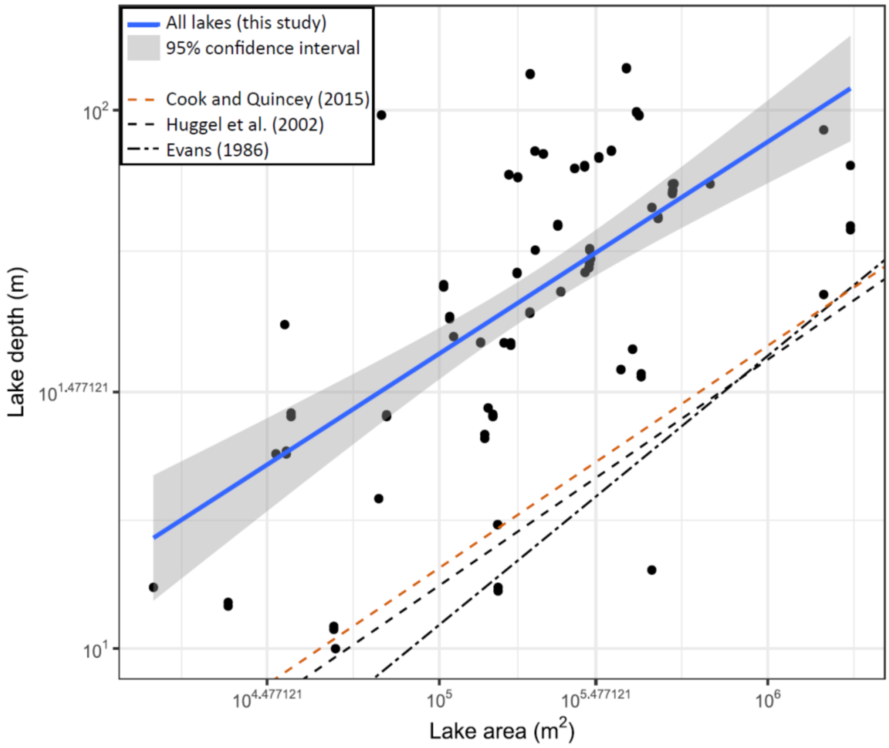

Figure S5. Comparison of the scaling relationships between lake area and lake depth or this study and for other similar studies. For specific details of the relationships presented, see Table 4.

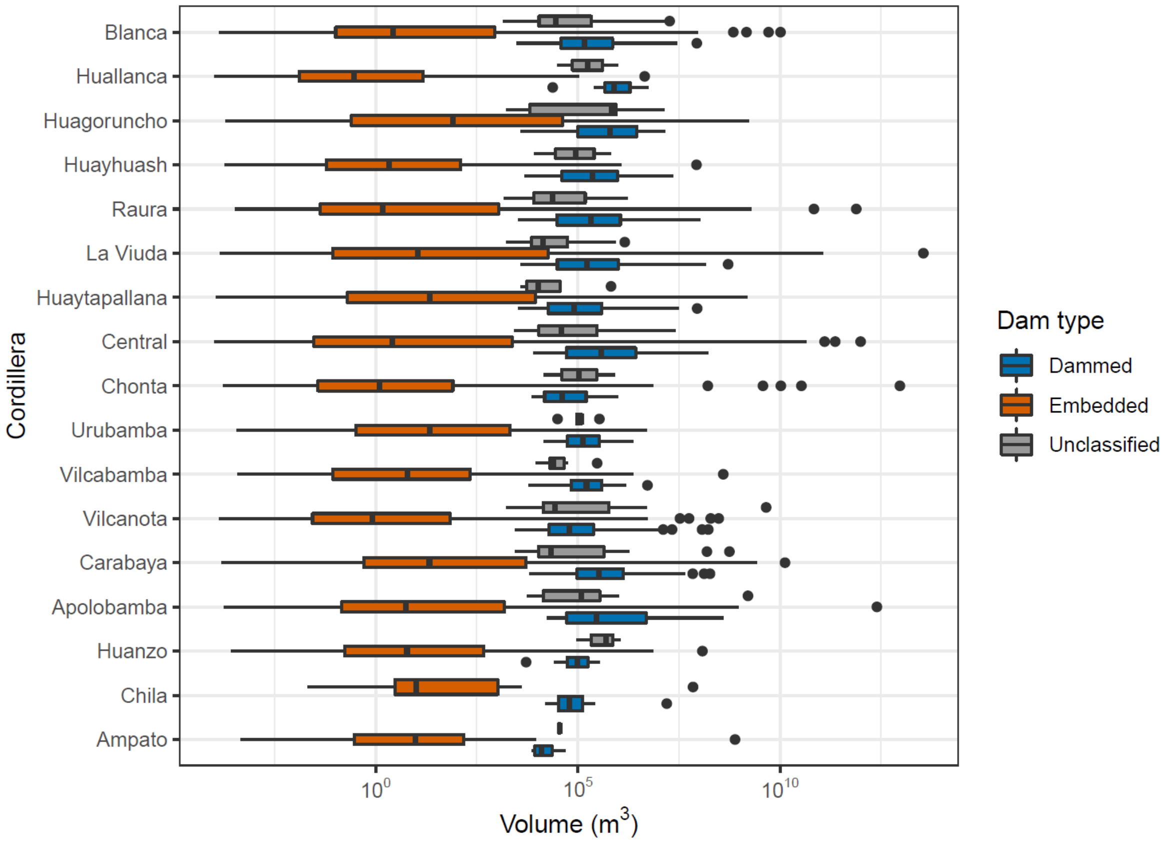

Figure S6. Lake volume for the 17 glaciated cordilleras calculated from the scaling exponents derived for dammed, embedded and unclassified lakes (Table 4). A degree of caution is required when interpreting this, as the exponents calculated for dammed lakes were derived from a limited number of points (e.g. Figure S4).

SI 5: Tables

Table S1. A review of glacier length change since the LIAM. Glaciers have retreated in all areas of Peru since the LIAM. The distance that glaciers have retreated is presented in several papers and summarised here.

Cordillera	Glacier	Total glacier length retreat (m)	Retreat estimate
Blanca		≅ 1,000 m (LIAM to beginning of 20th century) (Jomelli et al., 2008); < 1,400m (1930 - 2010) (Vuille et al., 2008)	**≅ 2,400 m across the Cordillera Blanca**
	Artesonraju	700 m (1948 - 1963) and 300m (1963 - 2000) (Jomelli et al., 2008)	
	Broggi	<600 m (1948 - 1988) (Kaser et al., 1990)	> 1,079 m for the 20th Century
		289.6 m (1932 - 1948); 720 m (1948-1993); ≅ 1009.6 m (1932-1993) (Marquez and Francou, 1995)	
		1,079 m (1932 - 1994) (Vuille et al., 2008)	
	Gajap	<300 m (1948 - 1988) (Kaser et al., 1990)	
	Pucaranra	690 m (1936 - 1994) (Vuille et al., 2008)	
	Uruashraju	<500 m (1939 - 1988) (Kaser et al., 1990)	675 m (1936 - 1994)
		495 m (1948 - 1993) (Marquez and Francou, 1995)	
		675 m (1936 - 1994) (Vuille et al., 2008)	
	Yanamarey	< 500 m (1939 - 1988) (Kaser et al., 1990)	Estimates from Yanamarey up to ≅ 2,270 m (from LIAM extent to 2019)
		≅ 600 m (1939 - 1988); ≅ 1,650 m (max extent to 1988) (Hastenrath and Ames, 1995)	
		405 m (1948 - 1993) (Marquez and Francou, 1995)	
		350 m (1948 - 1988); 552 m (1932 - 1994); Rate ≅ 20 m yr^{-1} (average 1977–2003) (Vuille et al., 2008)	
		≅ 950 m (MGE - 1938) (Jomelli et al., 2009)	
Vilcanota	Qori Kalis	73.5 m (mean retreat; 1963 - 1978); 41.6 m (mean retreat; 1978 - 1983); 112.9 m (mean retreat; 1983 - 1991) (Brecher and Thompson, 1993)	> 228 m (1963 - 1991) ≅ 870 m (1963 - 2005)
		6 m yr^{-1} (1963 - 1968; 30 m); 60 yr^{-1} (1991 - 2005; 840 m) (Thompson et al., 2006)	
	Sibinacocha	Estimate not explicit, but provided via large scale map (glaciers above lake Sibinacocha) (Seimon et al., 2007)	< 2,000 m
	Upismayo	Moraine 1,200 m in front of glacier terminus; ≅ 600 m downvalley is a second moraine dated as 630 ± 65 yr (Mercer et al., 1977)	< 1,800 m

Table S2. Satellite imagery used for the INAIGEM lake inventory (INAIGEM, 2018, p. 48).

Grupo o Cordillera	Fecha	Código de imagen / fuente	Resolución espacial (m)	Nivel	Área utilizada (%)
Blanca	28/07/2016	S2A_OPER_MSI_L1C_TL_MTI__20160728T215553_A005742_T17	10 y 20	L1C	2.52
	15/11/2016	S2A_OPER_MSI_L1C_TL_SGS__20161115T202633_A007315_T17	10 y 20	L1C	2.41
	28/07/2016	S2A_OPER_MSI_L1C_TL_MTI__20160728T215553_A005742_T18	10 y 20	L1C	28.86
	15/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161115T215249_A007315_T18	10 y 20	L1C	60.61
		Google Earth y Bing Maps	-	-	5.59
Huallanca	18/06/2016	S2A_OPER_MSI_L1C_TL_MTI__20160618T215428_A005170_T18	10 y 20	L1C	0.22
	15/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161115T215249_A007315_T18	10 y 20	L1C	99.41
		Google Earth	-	-	0.37
Huayhuash	17/08/2016	S2A_OPER_MSI_L1C_TL_MTI__20160728T215553_A005742_T18	10 y 20	L1C	33.72
	15/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161115T215249_A007315_T18	10 y 20	L1C	56.15
	28/07/2016	S2A_OPER_MSK_NODATA_MTI__20160817T215454_A006028_T18	10 y 20	L1C	6.26
		Google Earth	-	-	3.87
Raura	13/09/2016	S2A_OPER_MSI_L1C_TL_MTI__20160913T214558_A006414_T18	10 y 20	L1C	2.12
	12/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161112T214417_A007272_T18	10 y 20	L1C	38.98
	15/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161115T215249_A007315_T18	10 y 20	L1C	57.11
		Google Earth	-	-	1.79
Huagoruncho	13/09/2016	S2A_OPER_MSI_L1C_TL_MTI__20160913T214558_A006414_T18	10 y 20	L1C	60.03
	12/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161112T214417_A007272_T18	10 y 20	L1C	39.71
		Google Earth	-	-	0.26
La Viuda	13/09/2016	S2A_OPER_MSI_L1C_TL_MTI__20160913T214558_A006414_T18	10 y 20	L1C	35.44
	12/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161112T214417_A007272_T18	10 y 20	L1C	56.40
		Google Earth	-	-	8.16
Huaytapallana	13/09/2016	S2A_OPER_MSI_L1C_TL_MTI__20160913T214558_A006414_T18	10 y 20	L1C	88.13
	12/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161112T214417_A007272_T18	10 y 20	L1C	11.61
		Google Earth	-	-	0.26
Central	13/09/2016	S2A_OPER_MSI_L1C_TL_MTI__20160913T214558_A006414_T18	10 y 20	L1C	0.86
	12/11/2016	S2A_OPER_MSI_L1C_TL_MTI__20161112T214417_A007272_T18	10 y 20	L1C	94.20
		Google Earth y Bing Maps	-	-	4.94
Chonta	13/09/2016	S2A_OPER_MSI_L1C_TL_MTI__20160913T214558_A006414_T18	10 y 20	L1C	100.00
Vilcabamba	14/01/2016	S2A_OPER_MSI_L1C_TL_MTI__20160114T195530_A002939_T18	10 y 20	L1C	25.07
	12/06/2016	S2A_OPER_MSI_L1C_TL_MTI__20160612T213332_A005084_T18	10 y 20	L1C	18.62
	2/07/2016	S2A_OPER_MSI_L1C_TL_MTI__20160702T213420_A005370_T18	10 y 20	L1C	22.92
	31/08/2016	S2A_OPER_MSI_L1C_TL_MTI__20160831T213213_A006228_T18	10 y 20	L1C	23.65
	29/07/2016	S2A_OPER_MSI_L1C_TL_SGS__20160729T200057_A005756_T18	10 y 20	L1C	4.38
		Google Earth y Bing Maps	-	-	5.36
Urubamba	3/05/2016	S2A_OPER_PVI_L1C_TL_SGS__20160503T214932_A004512_T18	10 y 20	L1C	18.28
	10/05/2016	S2A_OPER_PVI_L1C_TL_SGS__20160510T200611_A004612_T18	10 y 20	L1C	14.77
	30/05/2016	S2A_OPER_PVI_L1C_TL_SGS__20160530T195530_A004898_T18	10 y 20	L1C	48.31
	29/07/2016	S2A_OPER_PVI_L1C_TL_SGS__20160729T200057_A005756_T18	10 y 20	L1C	5.37
	17/09/2016	S2A_OPER_PVI_L1C_TL_SGS__20160917T195316_A006471_T18	10 y 20	L1C	10.35
		Google Earth	-	-	2.91

Table S3. Satellite images used for the ANA lake inventory (ANA, 2014, p. 8).

Cordilleras	Año de imagen	Imagen	Resolución espacial m
Blanca	2001,2002,2003, 2006	Spot 4 y Aster	10,15
Huallanca	2007	Spot 5 y Aster	10,15
Huayhuash	2007	Spot 5 y Aster	10,15
Raura	2007	Spot 5 y Aster	10,15
La Viuda	2005, 2007	Spot 5 y Aster	10,15
Central	2007,2008	Aster	15
Huagoruncho	2009	Landsat	30
Huaytapallana	2009	Landsat	30
Chonta	2009	Landsat	30
Ampato	2010	Aster, LisIII	15,23
Vilcabamba	2009,2010	Aster, Landsat	15,30
Urubamba	2009,2010	Aster, Landsat	15,30
Huanzo	2010	Aster, LisIII	15,23
Chila	2010	Aster	15
La Raya	2009, 2010	Spot 4, LissIII	10, 24
Vilcanota	2009,2010	Spot 4, Spot 5 y LissIII	10,20,24
Carabaya	2009,2010	Spot 4, Spot 5, LissIII y Landsat 5	10,20,24,30
Apolobamba	2010	Landsat	30
Volcánica	2009	Aster	15

Table S4. The INAIGEM lake inventory uses both Google and Bing satellite imagery to map lakes across Peru at 10-20m spatial resolution (Table S2). Whilst the number of lakes smaller than 900 m^2 accounts for up to ≈50% of the total lake numbers (varying by region), this accounts for a fraction of the total lake area (with <2% of the total lake area being made up of these small lakes).

Cordillera	Total *n* lakes	*n* lakes ≤900 m^2	% *n* lakes ≤900 m^2	Total lake area (km^2)	Lake area ≤900 m^2 (km^2)	% lake area ≤900 m^2
Ampato	39	16	41.03	0.6	0.007	1.15
Apolobamba	179	22	12.29	28.7	0.010	0.03
Blanca	882	67	7.60	32.7	0.039	0.12
Carabaya	661	8	1.21	40.0	0.005	0.01
Central	490	16	3.27	30.8	0.011	0.03
Chila	19	3	15.79	0.9	0.002	0.23
Chonta	127	13	10.24	9.3	0.004	0.05
Huagoruncho	206	40	19.42	11.3	0.020	0.17
Huallanca	74	10	13.51	1.3	0.007	0.52
Huanzo	54	3	5.56	1.5	0.002	0.16
Huayhuash	172	34	19.77	6.0	0.019	0.32
Huaytapallana	614	155	25.24	18.3	0.072	0.39
La Viuda	583	185	31.73	38.3	0.089	0.23
Raura	513	261	50.88	14.6	0.100	0.69
Urubamba	241	84	34.85	2.4	0.041	1.71
Vilcabamba	269	36	13.38	3.1	0.023	0.76
Vilcanota	1250	607	48.56	47.7	0.268	0.56
Peru	**6373**	**1560**	**24.48**	**287.5**	**0.719**	**0.25**

Table S5. Kruskal-Wallis rank sum tests results for Figure 3A (lake area for the 17 glaciated cordilleras). Comparison between regions to see if there is a significant difference between recorded lake area. Presented here are the p-values for the pairwise Wilcox tests. Cells coded in green are those with a significant difference between lake areas recorded across each region.

	Ampato	Apolobamba	Blanca	Carabaya	Central	Chila	Chonta	Huagoruncho	Huallanca	Huanzo	Huayhuash	Huaytapallana	La Viuda	Raura	Urubamba	Vilcabamba
Apolobamba	0.32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Blanca	0.41	0.62	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Carabaya	0.11	0.50	<0.05	-	-	-	-	-	-	-	-	-	-	-	-	-
Central	0.49	0.57	0.97	<0.05	-	-	-	-	-	-	-	-	-	-	-	-
Chila	0.54	0.97	0.95	0.80	0.90	-	-	-	-	-	-	-	-	-	-	-
Chonta	0.95	<0.05	<0.05	<0.05	<0.05	0.32	-	-	-	-	-	-	-	-	-	-
Huagoruncho	<0.05	<0.05	<0.05	<0.05	<0.05	0.33	<0.05	-	-	-	-	-	-	-	-	-
Huallanca	0.90	<0.05	<0.05	<0.05	0.05	0.30	0.90	<0.05	-	-	-	-	-	-	-	-
Huanzo	0.50	0.73	0.95	0.30	0.93	1.00	0.15	<0.05	0.22	-	-	-	-	-	-	-
Huayhuash	0.30	0.97	0.54	0.64	0.57	0.99	<0.05	<0.05	<0.05	0.62	-	-	-	-	-	-
Huaytapallana	0.32	0.90	0.63	0.18	0.66	1.00	<0.05	<0.05	<0.05	0.74	0.88	-	-	-	-	-
La Viuda	0.31	1.00	0.34	0.42	0.37	0.99	<0.05	<0.05	<0.05	0.63	1.00	0.74	-	-	-	-
Raura	0.64	0.34	0.54	<0.05	0.64	0.78	0.05	<0.05	0.17	0.95	0.32	0.34	0.19	-	-	-
Urubamba	0.18	1.00	0.62	0.38	0.66	0.90	<0.05	<0.05	<0.05	0.73	0.90	1.00	0.90	0.37	-	-
Vilcabamba	0.34	0.66	1.00	0.05	1.00	1.00	<0.05	<0.05	<0.05	0.97	0.54	0.64	0.43	0.66	0.63	-
Vilcanota	1.00	<0.05	<0.05	<0.05	<0.05	0.43	0.74	<0.05	0.66	0.21	<0.05	<0.05	<0.05	0.07	<0.05	<0.05

Table S6. Kruskal-Wallis rank sum tests results for Figure 3B (ice contact and lake size). Cells coded in green are those with a significant difference between lake areas for lakes in contact and not in contact with ice for each of the cordilleras.

	Kruskal-Wallace χ^2	df	p-value
Blanca	4.36	1	<0.05
Chila	0.07	1	0.79
Ampato	0.83	1	0.36
Carabaya	0.01	1	0.94
Vilcanota	5.32	1	<0.05
Central	1.40	1	0.24
Vilcabamba	0.24	1	0.63
Apolobamba	1.00	1	0.32
Huayhuash	0.69	1	0.41
La Viuda	3.88	1	0.05
Raura	0.02	1	0.90
Huallanca	0.54	1	0.46
Huaytapallana	0.21	1	0.65
Huanzo	all observations are in the same group		
Chonta	all observations are in the same group		
Huagoruncho	0.16	1	0.68
Urubamba	0.76	1	0.38

Table S7. Pairwise Wilcox tests results for Figure 3C (dam type and lake size). Cells coded in green are those with a significant difference between lake areas for the different dam types in each of the cordilleras.

		Dammed	Embedded
Blanca	Embedded	<0.01	-
	Unclassified	<0.05	0.70
Chila	Embedded	0.73	-
Ampato	Embedded	0.18	-
	Unclassified	0.60	1.00
Carabaya	Embedded	<0.01	-
	Unclassified	0.09	0.41
Vilcanota	Embedded	<0.05	-
	Unclassified	0.77	0.32
Central	Embedded	<0.01	-
	Unclassified	<0.01	0.25
Vilcabamba	Embedded	<0.01	-
	Unclassified	0.01	0.70
Apolobamba	Embedded	0.01	-
	Unclassified	0.34	0.34
Huayhuash	Embedded	0.01	-
	Unclassified	0.71	0.59
La Viuda	Embedded	0.05	-
	Unclassified	<0.01	0.01
Raura	Embedded	0.01	-
	Unclassified	<0.05	0.63
Huallanca	Embedded	<0.01	-
	Unclassified	0.58	0.18
Huaytapallana	Embedded	0.34	-
	Unclassified	0.34	0.34
Huanzo	Embedded	0.82	-
	Unclassified	0.14	0.14
Chonta	Embedded	0.93	-
	Unclassified	0.71	0.71
Huagoruncho	Embedded	0.01	-
	Unclassified	0.75	0.75
Urubamba	Embedded	0.45	-
	Unclassified	0.45	0.45

Table S8. Kruskal-Wallis rank sum tests results for Figure 4B (ice contact and lake elevation). Cells coded in green are those with a significant difference between lake elevation for lakes in contact and not in contact with ice for each of the cordilleras.

	Kruskal-Wallace χ^2	df	p-value
Blanca	15.38	1	<0.01
Chila	0.64	1	0.4227
Ampato	2.70	1	0.1003
Carabaya	17.04	1	<0.01
Vilcanota	35.85	1	<0.01
Central	25.67	1	<0.01
Vilcabamba	12.81	1	<0.01
Apolobamba	4.74	1	<0.05
Huayhuash	9.86	1	<0.01
La Viuda	7.92	1	<0.01
Raura	7.36	1	<0.01
Huallanca	7.17	1	<0.01
Huaytapallana	0.47	1	0.4911
Huanzo	all observations are in the same group		
Chonta	all observations are in the same group		
Huagoruncho	2.41	1	0.1204
Urubamba	2.79	1	0.09496

Table S9. Pairwise Wilcox tests results for Figure 4C (dam type and lake elevation). Cells coded in green are those with a significant difference between lake areas for the different dam types in each of the cordilleras.

		Dammed	Embedded
Blanca	Embedded	0.15	-
	Unclassified	<0.01	<0.01
Chila	Embedded	0.07	-
Ampato	Embedded	0.38	-
	Unclassified	0.80	0.40
Carabaya	Embedded	0.86	-
	Unclassified	0.86	0.86
Vilcanota	Embedded	<0.01	-
	Unclassified	0.23	0.17
Central	Embedded	<0.01	-
	Unclassified	0.47	0.20
Vilcabamba	Embedded	0.17	-
	Unclassified	0.22	0.12
Apolobamba	Embedded	0.96	-
	Unclassified	0.96	0.96
Huayhuash	Embedded	<0.01	-
	Unclassified	0.90	0.09
La Viuda	Embedded	<0.01	-
	Unclassified	0.15	0.87
Raura	Embedded	<0.01	-
	Unclassified	0.12	0.76
Huallanca	Embedded	<0.01	-
	Unclassified	0.22	<0.05
Huaytapallana	Embedded	<0.01	-
	Unclassified	0.89	0.18
Huanzo	Embedded	0.07	-
	Unclassified	0.01	<0.01
Chonta	Embedded	0.53	-
	Unclassified	0.39	0.39
Huagoruncho	Embedded	<0.01	-
	Unclassified	0.15	<0.05
Urubamba	Embedded	0.05	-
	Unclassified	0.05	0.01