*Journal of Glaciology*



**Article**

**Cite this article:** Fernández A, Somos-Valenzuela M (2022). Revisiting glacier mass-balance sensitivity to surface air temperature using a data-driven regionalization. *Journal of Glaciology* **68**(272), 1041–1060. https://doi.org/10.1017/jog.2022.16

Received: 28 January 2021
Revised: 23 February 2022
Accepted: 23 February 2022
First published online: 11 April 2022

**Keywords:**
Climate change; glacier mass balance; ice and climate

**Author for correspondence:**
Alfonso Fernández,
E-mail: alfernandez@udec.cl

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

**cambridge.org/jog**

# Revisiting glacier mass-balance sensitivity to surface air temperature using a data-driven regionalization

Alfonso Fernández[1] and Marcelo Somos-Valenzuela[2]

[1]Department of Geography, Mountain Geoscience Group, Universidad de Concepción, Concepción, Chile and [2]Department of Forestry Sciences, Butamallin Research Center for Global Change, Universidad de La Frontera, Temuco, Chile

## Abstract

This study involves examination of glaciological mass-balance time series, glacier and climatic descriptors, the application of machine learning methods for glaciological clustering, and computation of mass-balance time series based upon the clustering and statistical analyses relative to gridded air temperature datasets. Our analysis revealed an increasingly coherent mass-balance trend but a latitudinal bias of monitoring programs. The glacier classification scheme delivered three clusters, suggesting these correspond to climate-based first-order regimes, as glacier morphometric characteristics weighed little in our multivariate analysis. We combined all available surface mass-balance data from in situ monitoring programs to study temperature sensitivity for each cluster. These aggregated mass-balance time series delivered spatially different statistical relationships to temperature. Results also showed that surface mass balance tends to have a temporal self-correlation of ∼20 years. Using this temporal window to analyze sensitivity since ∼ 1950, we found that in all cases temperature sensitivity, while generally negative, tended to fluctuate through time, with the largest absolute magnitudes occurring in the 1980s and becoming less negative in recent years, revealing that glacier sensitivity is non-stationary. These findings point to a scenario of a coherent signal of change no matter the glacier regime. This work provides new insights into glacier–climate relationships that can guide observational and analytical strategies.

## 1. Introduction

Glaciers are open dynamical systems in which input, storage and release of mass and energy fluctuate largely according to climatic stimuli. The surface mass balance of glaciers, or the sum of surface ablation and surface accumulation (Cogley and others, 2011) is primarily controlled by climate (Hanna and others, 2020). This climate–glacier relationship is expressed in the ice/snow mass transport from the highest to lowest elevations, ultimately releasing mass mainly as runoff. Major sources for mass buildup are snowfall and snowdrift, which directly relate precipitation to accumulation, while solar radiation, atmospheric radiation and turbulent fluxes control the amount of ice loss or ablation, and are often highly correlated to near-surface temperature. This makes glaciers natural recorders of environmental changes and thus robust evidence of the lasting effects of anthropogenic global warming (Roe and others, 2017).

Abundant prior research has documented a continuous generalized global glacier shrinkage (e.g. Meier, 1984; Mernild and others, 2013; Ciracì and others, 2020), while projections forecast further demise (Marzeion and others, 2012; Hock and others, 2019). The implications of these changes are manifold: sea level rise, changes in water availability, increase in natural hazards, release of previously locked pollutants and cultural changes, among others (Mark and Fernández, 2017). As one of the essential cryosphere variables that provide evidence to better understand climate evolution, surface mass balance delivers critical insights to quantify the impact of glacier changes downstream (Nussbaumer and others, 2017).

Thus, determining the sensitivity of the surface mass balance to climatic elements has proven to be fundamental in forecasting glacier volumetric changes. Surface mass balance sensitivity (henceforth 'sensitivity') can be defined as the change in mass balance due to a change in a climatic variable such as air temperature or precipitation, and is regularly denoted as a response to temperature in Kelvin or proportional to precipitation in percentages. Most studies to date have determined 'static' and 'dynamic' sensitivities (Cogley and others, 2011) by employing numerical modeling, typically for temperature. For example, Oerlemans and Fortuin (1992) determine a $-0.4\ \mathrm{m\ K^{-1}}$ from modeling a sample of 12 glaciers, while posterior studies find similar figures, with averages within $-0.2$ to $-0.4\ \mathrm{m\ K^{-1}}$ (Gregory and Oerlemans, 1998; Raper and Braithwaite, 2006; Meehl and others, 2007; Marzeion and others, 2014). However, some of the few calculations using observations deliver a slightly higher range, to $\sim -0.47\ \mathrm{m\ K^{-1}}$ (Meehl and others, 2007).

Decades of effort by the glaciological community gathering data seem to be paying off as a much larger and complete record of glaciers exists (Zemp and others, 2015). Figure 1 is a map featuring the total number of surface mass balance monitored years until 2017, with the Randolph Glacier Inventory (RGI; RGI Consortium, 2017) as background. Both databases

correspond to the most extensive and unified sources of information on the world's glaciers, and are regularly revised and expanded (WGMS, 2020; Pfeffer and others, 2014). The map shows 450 glaciers that are or have been monitored sometime since the 19th century. This sample covers the world's cryosphere more completely, compared to what was available just a decade ago or so (Dyurgerov, 2002; Radić and Hock, 2010). However, Figure 1 also reveals the disparity of surface mass balance monitoring between hemispheres, with the North concentrating 88.4% of all programs and 100% of all of those with more than 50 years of measurements. If one compares this proportion with the North-South glacier distribution according to the RGI there is a remarkable similarity, as the Southern Hemisphere contains 11.4% of the world's glaciers by number, and 20.4% by area. High Mountain Asia displays the majority of monitoring programs running 20 years or less, especially toward the southern slope. Glaciers with the longest time series are concentrated in Europe, followed by North America. Africa constitutes the other extreme, with just one monitored glacier, Lewis in Kenya. In the Southern Hemisphere, the time series of Echaurren Norte more than doubles the rest of the glaciers from the South, being the only included in the list of reference glaciers defined by the World Glacier Monitoring Service (WGMS, https://wgms.ch/products_ref_glaciers/). Notably, the last couple of decades have seen an increased monitoring of Tropical glaciers.

Initiatives leading to the World Glacier Inventory (WGI), launched during the 1970s, aimed at expanding the network of monitored glaciers to assess regional variability (WGMS, 1989). Today, more abundant information on glacier changes makes it more feasible to detect global signals, discriminate regional spread, as well as improve quantification of impacts (Zemp and others, 2019). The surface mass balance time series of reference glaciers represents an effort to minimize bias toward highly monitored regions when studying global mass balance from observational datasets. To build that curve, the WGMS computes regional averages of glaciers having, among other features, 30 or more years of ongoing mass-balance measurements (https://wgms.ch/products_ref_glaciers/). The core of this regional treatment of bias derives from the 19 glaciological zones of the RGI, based upon the work of Radić and Hock (2010). During our literature review for this research, we did not find published work explaining the rationale behind this zonation. Given today's wealth of information on surface mass balance and on the number of glaciers inventoried, we deemed this situation as an opportunity to revisit surface mass balance sensitivity to surface air temperature using a data-driven approach. We posit that there is room for extracting information about how temperature change impacts glaciers applying alternative approaches to computing the surface mass balance time series that maximize the use of all available data. In particular, we assert that using the RGI to derive a regionalization has to be tested in order to determine whether the current sample of surface mass balance adds value or changes some of the figures previously published. However, it is important to note that the development of glacier inventories, particularly the WGI, had first to deal with the lack of data instead of determining precise regions, and thus initial priority focus was establishing a network of collaborators in order to provide a basis for long-term monitoring of glacier behavior in different climatic regions (WGMS, 1989). Therefore, our main goal is to reanalyze global surface mass balance and its sensitivity to air temperature by applying machine learning methods of unsupervised classification to the RGI and related climatic variables. To our knowledge, this is the first study trying to explicitly employ machine learning to combine surface mass balance, glacier inventory descriptors and climatic data to analyze the global sensitivity of changes in glacier mass to temperature.

## 2. Data and methods

The flowchart in Figure 2 shows the stages followed. The study involved examination of available time series of glaciological mass-balance programs until the year 2017; the compilation of a geodatabase that included glacier and climatic descriptors; the application of machine learning methods to define glaciological regions using unsupervised clustering; the computation of different global mass-balance average time series based upon the clustering; and the correlations between these and different global land air temperature datasets.

Annual surface mass balance time series are available for download from the WGMS. This database includes records of 450 glaciers that have been monitored since 1885, the first reference year for the Rhone glacier in Switzerland (WGMS, 2020). Our analysis is mostly focused on the periods 1920-2017 and 1950–2017 due to the facts that (1) before 1920 generally only records for one glacier exist (Rhone from 1885 to 1909 and Clariden for 1915–1918), and (2) that 1950 corresponds to the first record of reference glaciers according to WGMS (2020). In turn, the geodatabase is compiled from two sources, RGI version 6, and the WorldClim's global climate fields (Fick and Hijmans, 2017). The RGI comprises a geodatabase of glacier outlines that includes more than a dozen glacier attributes, from identification codes to quantitative descriptors such as the maximum altitude, and it is based upon satellite imagery spanning 1999 to the present (Pfeffer and others, 2014). The WorldClim database that we utilized is the version 2, 1 km gridded climatology that includes seven variables, developed at the monthly scale for the period 1970–2000 interpolating thousands of climatic records combined with several ancillary spatially-distributed variables as co-variates, such as elevations and satellite-derived land surface temperatures (Fick and Hijmans, 2017).

Exploratory analysis of surface mass balance time series included characterization of regional biases. First, we computed a surface mass balance global average composite for each measured year, with its corresponding 5% confidence interval, and compared it against the global mass change of reference glaciers available from the WGMS (https://wgms.ch/global-glacier-state/). We supplemented the analysis of the composite with the calculation of the number of monitored glaciers per year and their latitudinal distribution. Because the number of mass-balance programs has varied through time, we deemed it important to study whether their signal is consistent during the period. A linear correlation analysis and a spatial distribution index allowed us to test the spatio-temporal consistency between time series. The correlation analysis implemented corresponds to the Expressed Population Signal (EPS), an index regularly utilized in dendrochronology (Wigley and others, 1984) that is computed from the mean inter-series correlation ($\bar{r}$) and the sample size ($N$) within a given moving window:

$$\mathrm{EPS} = \frac{N\bar{r}}{1 + (N-1)\bar{r}} \tag{1}$$

EPS thus is a tool to study how similarly the time series covary overtime in the sense that provides an index to determine how much of the variance is shared by the individual time series. On the other hand, we developed a spatial coverage index (SC) to determine the distribution of mass-balance programs (SMB) relative to the global glacier distribution:

$$\mathrm{SC} = \frac{\sum \mid G \cap \mathrm{SMB} \mid}{\sum \mid G \mid} \times 100(\%) \tag{2}$$

The global glacier distribution used for this calculation ($G$) is a rasterized version of the RGI in which the number of glacier

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

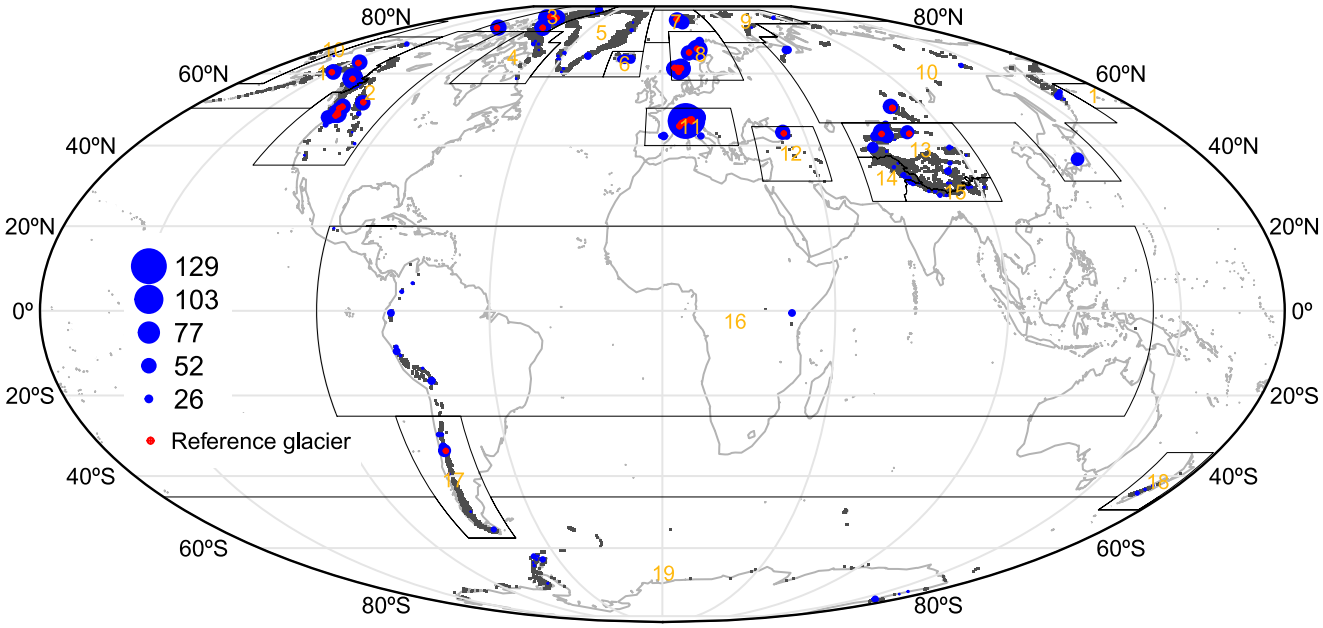



**Fig. 1.** World map that includes the global distribution of glaciers with dark gray dots and major glaciological zones with the corresponding number according to the RGI (RGI Consortium, 2017). Glaciological mass-balance programs from the WGMS database are also shown, with the size of blue dots representing the number of monitored years while the red dots identifying the WGMS reference glaciers.

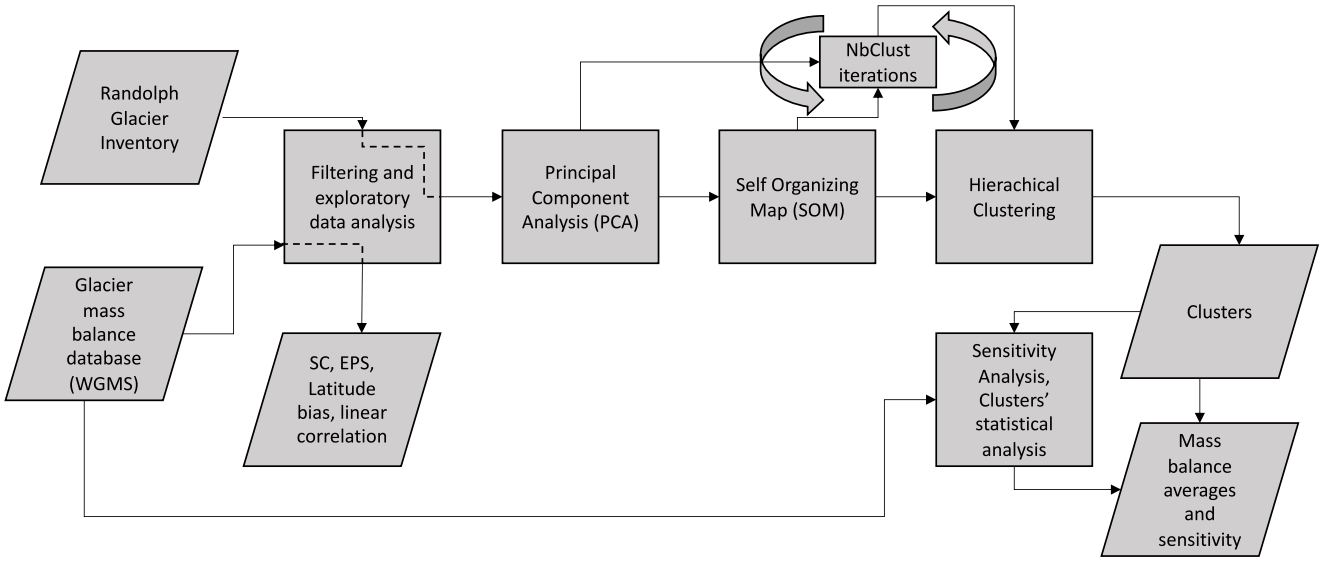



**Fig. 2.** Flowchart depicting the data processing and outputs. The process begins to the left with the inventory and surface mass balance data. After an initial exploratory data analysis, the inventory is used to produce an unsupervised classification of glaciers which in turn is utilized to evaluate temperature sensitivity per cluster.

polygons contained within a given 1° pixel is flagged as 'glacierized'. The spatial coverage is then the number of pixels that contain at least one surface mass balance program (computation represented by the intersection symbol, ∩, in Eqn (2)), divided by the total number of 'glacierized' pixels. Thus, a spatial coverage of 100% would mean that all 1° glacierized pixels contain at least one active mass-balance program; conversely a 0% the case that no region has an active surface mass balance program. Figure 3 is a scheme depicting the procedure to determine the number of pixels containing surface mass balance measurements. The EPS and spatial coverage were calculated over 20-year moving windows from 1920 to 2017. The 20-year moving window resulted from screening cross-correlations of surface mass balance data and global temperatures, as depicted in the end of this section.

During compilation of the geodatabase from the RGI, we checked for the presence of outliers in the fields and retained climatic and morphologic descriptors that were continuously recorded for all glaciers (Table 1). The attributes containing a category labeled as 'not assigned' were removed from further analysis in order to avoid misclassification, especially because some of those attributes have only been included for certain regions, such as connectivity for glaciers in Greenland (Pfeffer and others, 2014). On the other hand, glacier records containing anomalous attribute values, such as zero maximum elevation, were removed. To this database we added the hypsometric integral, computed using the equation presented in Pike and Wilson (1971) applied to the elevational distribution data provided within the RGI. Hypsometry has been used to analyze glacier response to climate change (Furbish and Andrews, 1984; Etzelmüller, 2000; Fernández and others, 2010), and thus the hypsometric integral may allow for detection of glacier regimes. Slope and aspect data from the RGI were transformed to m/m (rise over run)

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

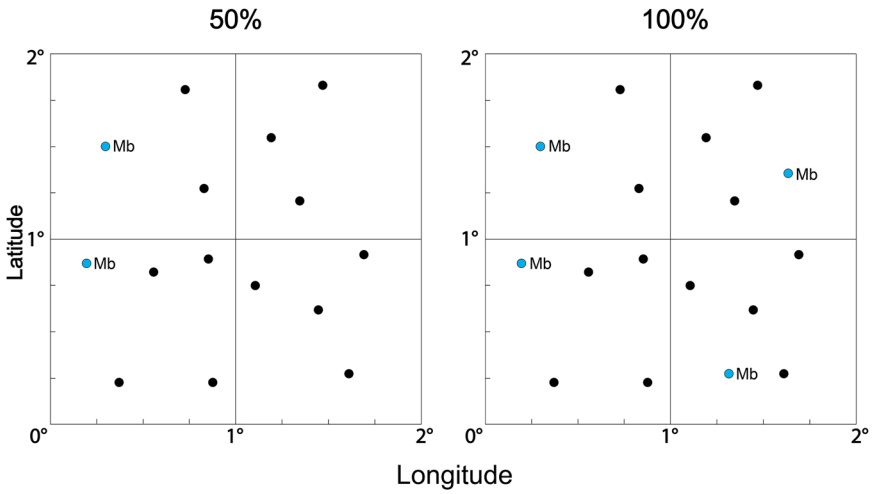



**Fig. 3.** Scheme describing how to determine the number of pixels containing surface mass balance measurements and its use in the final calculation of the spatial coverage; in this case, with an hypothetical scenario of a squared world spanning 2° of latitude and longitude, with only 15 glaciers (black and cyan dots). The quadrants represent the borders of the 1° pixels used to determine whether or not they are glacierized. Since in this case all quadrants contain at least one glacier, all are flagged as 'glacierized'. The 50% scenario is when two out of the four 1° glacierized pixels contain at least one glacier that is monitored for surface mass balance (cyan circles labeled as 'Mb'). The 100% scenario is then when all glacierized pixels contain at least one active surface mass balance program.

**Table 1.** Descriptors of the RGI and WorldClim databases

| Data from RGI | | |
|---|---|---|
| Attribute description | Abbreviation | Unit |
| Code for identification within the RGI | *RGIId* | No unit |
| Code for identification within the GLIMS | *GLIMSId* | No unit |
| Outline source date | *BgnDate* | Y-M-D |
| Outline source date (when several dates available) | *EndDate* | Y-M-D |
| Latitude of the glacier polygon centroid★ | *Lat* | Decimal degrees |
| Longitude of the glacier polygon centroid ★ | *Lon* | Decimal degrees |
| Code of glacier within its first order region | *01Region* | No unit |
| Code of glacier within its second order region | *02Region* | No unit |
| Minimum elevation★ | *Zmi* | m |
| Median elevation★ | *Zme* | m |
| Maximum elevation★ | *Zma* | m |
| Area★ | *Area* | $km^2$ |
| Slope★ | *Sl* | Degrees (m/m) |
| Aspect★ | *Asp* | Degrees (radians) |
| Maximum lenght★ | *Lm* | m |
| Outlines await subdivision or are nominal circles | *Status* | Categorical |
| Connectivity relative to an ice sheet | *Connect* | Categorical |
| Form classification | *Form* | Categorical |
| Terminus type | *TermType* | Categorical |
| Surging | *Surging* | Categorical |
| Database link to mass-balance measurement | *Linkages* | Categorical |
| Glacier name | *Name* | No unit |
| Hypsometric integral★ | *Hyp* | Ratio |
| **Data from WorldClim** | | |
| Attribute description | Abbreviation | Unit |
| Minimum temperature★ | *Tmin* | °C |
| Maximum temperature★ | *Tmax* | °C |
| Average temperature★ | *Tav* | °C |
| Temperature range★ | *Rn* | °C |
| Solar radiation★ | *Rad* | MJ $m^{-2}$ |
| Wind speed★ | *Wind* | m $s^{-1}$ |
| Vapor pressure★ | *Vap* | kPa |
| Precipitation★ | *Pp* | mm |

The star (★) indicates the ones used in this study.

and radians, respectively. The WorldClim attributes attached to each glacier in the geodatabase corresponded to the climatological value for the gridbox that intersects the centroid of each polygon. That climatological value for each gridbox was computed as arithmetic averages for the whole 1970–2000 period covered by the WorldClim product. From this database, we added the annual temperature range (*Rn*), i.e. the difference between maximum and minimum annual temperature. This resulted in a database containing 208 412 records, 96.7% of the RGI, with 18 attributes (Table 1). The exploratory analysis of this database included linear correlation analyses among all combination of variables.

Because the climatological descriptors were added to each glacier's centroid while the regular representation of the climatology contrasts with the diverse shapes of glacier polygons, two extreme cases may occur. The first is that the gridbox intersecting a centroid may fall within a large glacier polygon, which may result in insufficient sampling of the corresponding climatic descriptors. The second case occurs when the centroid of more than one glacier

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

polygon falls within one gridbox, leading to a possibly unrealistic identical climate description for contiguous glaciers. This problem will arise with any gridded climatology available at a spatial resolution smaller than 5°. While the first situation can be solved by sampling all gridboxes within a glacier, that method will not be applicable to the small glaciers of the second situation, preventing the elimination of possible biases. Assuming this uncertainty, and in order to minimize its effect on the clustering process, a sampling strategy was implemented, as described in the next paragraph.

The clustering process of glaciers consisted of an unsupervised classification derived from dimensionality reduction using Machine Learning techniques. Three major procedures were applied to the compiled geodatabase. First, we applied a Principal Component Analysis (PCA) to reduce dimensionality from the 18 descriptors to 2–3 new uncorrelated variables, called principal components or PCs, which maximize the explained variance while reducing redundancy among similar variables from the original database (Demšar and others, 2013). Results applied to the descriptors both in their original units (Table 1) and on normalized versions (using the mean and standard deviation) were identical. In a second procedure, the first PCs were used as input for a Kohonen self-organizing map (SOM) algorithm (Kohonen, 1990) to force the input PCs to fit within a two-dimensional space. A SOM is a class of neural network that reveals the structure of a dataset by competitive learning. One of the tenets of this method is that when two objects are similar in the multidimensional space, they should be neighbors in the SOM two-dimensional space too. Thus, in a SOM the input vector is topologically organized as a web of nodes. This technique has been previously tested in hydroclimatic research, suggesting that can be powerful for non-linear problems (e.g. Reusch and others, 2005; Markonis and Strnad, 2020). The use of PCA to provide a normalized set of variables for initialization of the SOM is a now common procedure in data analysis that tends to support reproducible results (Akinduko and others, 2016). The number of training epochs of the network (1 152 000) and the size of the map (48 × 48 nodes) for computing the SOM was determined following recommendations by Kohonen (1990) and Vesanto (2000). Unsupervised clustering of the resultant SOM nodes was determined using a hierarchical algorithm using the Ward agglomerative approach (Ward, 1963), where the optimal number of clusters was selected by a majority analysis of more than a dozen available criteria implemented in the package NbClust of the R programming language (Charrad and others, 2014). NbClust was applied to the SOM using a range from 2 to 31 possible cluster numbers, starting from the simple idea of two possible groups, to testing a larger number of defined RGI regions (20) based upon the work of Radić and Hock (2010) and used in the calculation of the WGMS's global glacier mass-balance time series. The maximum number of proposed clusters comes from the regions defined by Meier (1984). As explained in the previous paragraph, the use of the polygon centroids implies a certain degree of uncertainty in the climatological representation of glaciers. Another issue to consider, that is derived from the RGI database, is that glacier outlines are extracted from different sources and (most critically) years (Pfeffer and others, 2014). This might deliver inhomogeneity in critical parameters from small glaciers that have been losing mass rapidly (e.g. Permana and others, 2019), so that the recorded minimum elevation of one glacier may not completely be comparable to other glaciers. In order to test whether these uncertainties may influence the definition of the optimal number clusters, we ran 100 NbClust iterations using a random sampling of 1000 records from the PCA output. The sampling size is computed from a 99% confidence level at 5% error margin and the samples are drawn without replacement. Unlike the application for the SOM, for the PCs the optimal number of clusters was determined by the most common winner of the majority analysis after tabulating the result of all 100 iterations.

New, alternative global and regional mass-balance averages can be constructed based upon the clustering process. We recalculated SBM from all glaciers available according to the regions determined through the clustering process. We evaluated these averages relative to the WGMS global mean using linear correlations, comparing linear trends and cumulative mass balances. The sensitivity evaluation included correlations and linear regressions between the time series generated in this work and the land air temperatures from three well-known global climatologies: NASA GISS (Lenssen and others, 2019, henceforth GISS), CRUTEM4 (Osborn and Jones, 2014, henceforth CRUTEM) and BerkeleyEarth (Rohde and others, 2013, henceforth Berkeley). We developed linear models since 1950 as well as a number of cross-correlations using moving windows in order to compute and analyze sensitivity, expressed in the slope or $\beta$ parameter of the linear regression.

## 3. Results

### 3.1. Exploratory analysis of mass balance data

We begin with an analysis of the composite mass-balance time series computed from all available data per year since 1920. Figure 4 shows the annual average of mass balance for all glaciers included in the WGMS database. High uncertainty in this mass-balance composite for the period prior to 1950 is evident as the 95% confidence intervals are remarkably wide, peaking with maxima/minima of $\sim -6$ and $+3$ m w.e. a$^{-1}$ during the 1920s (Fig. 4a). Mass-balance uncertainty diminished drastically since the early 1950s, with a positive maximum of $\sim$0.4 m w.e. a$^{-1}$ in the mid 1950s, and practically no year after 1980 recording a positive upper limit of the confidence interval (Fig. 4b). The reason for some years with high uncertainty and others with zero uncertainty in the annual average mass-balance series at the earlier sections of the series is obviously the extant number of data points during those first decades of continuous monitoring (Fig. 4c). Four periods concentrate the narrowest confidence intervals since 1950: 1962–1966, 1968–1970, 1996–2000 and 2008–2010. The narrow uncertainty detected in the two periods since 1996 are interesting for two reasons. First, those later periods record at least 25% more surface mass balance data, with 1996–2000 having an average of 120 while 1968–1970 registering 90 (Fig. 4c). Second, the last two periods are the only in which such a low uncertainty includes glaciers from the Southern Hemisphere as well as from the Tropics, suggesting a more globally consistent mass-balance response during that period. After 1950 the number of monitored glaciers began to increase significantly, triplicating the records from 1945, quadruplicating 1950's figure by the early 1960s, and then increasing to more than 150 by the late 2000s. Comparing panels a and c in Figure 4, there seems to be a moderate inverse relationship between uncertainty and the number of mass-balance data per year since 1960, as more mass-balance data become available. However, the opposite seems to occur as there is a moderate statistically significant positive linear correlation between the number of monitored glaciers for the whole period 1920–2017 and the standard deviation of mass balance computed for the corresponding year (0.65 with $p < 0.001$) which lowers after 1950 (0.52 with $p < 0.001$) whereas the correlation before 1950 is 0.63 ($p < 0.001$). What in fact seems to happen is that before 1950 the low number of monitored glaciers produces extreme fluctuations in the signal to noise ratio, leading to an overall positive relationship. After 1950 the uncertainty varies less, leading to a more stable signal to noise ratio.

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

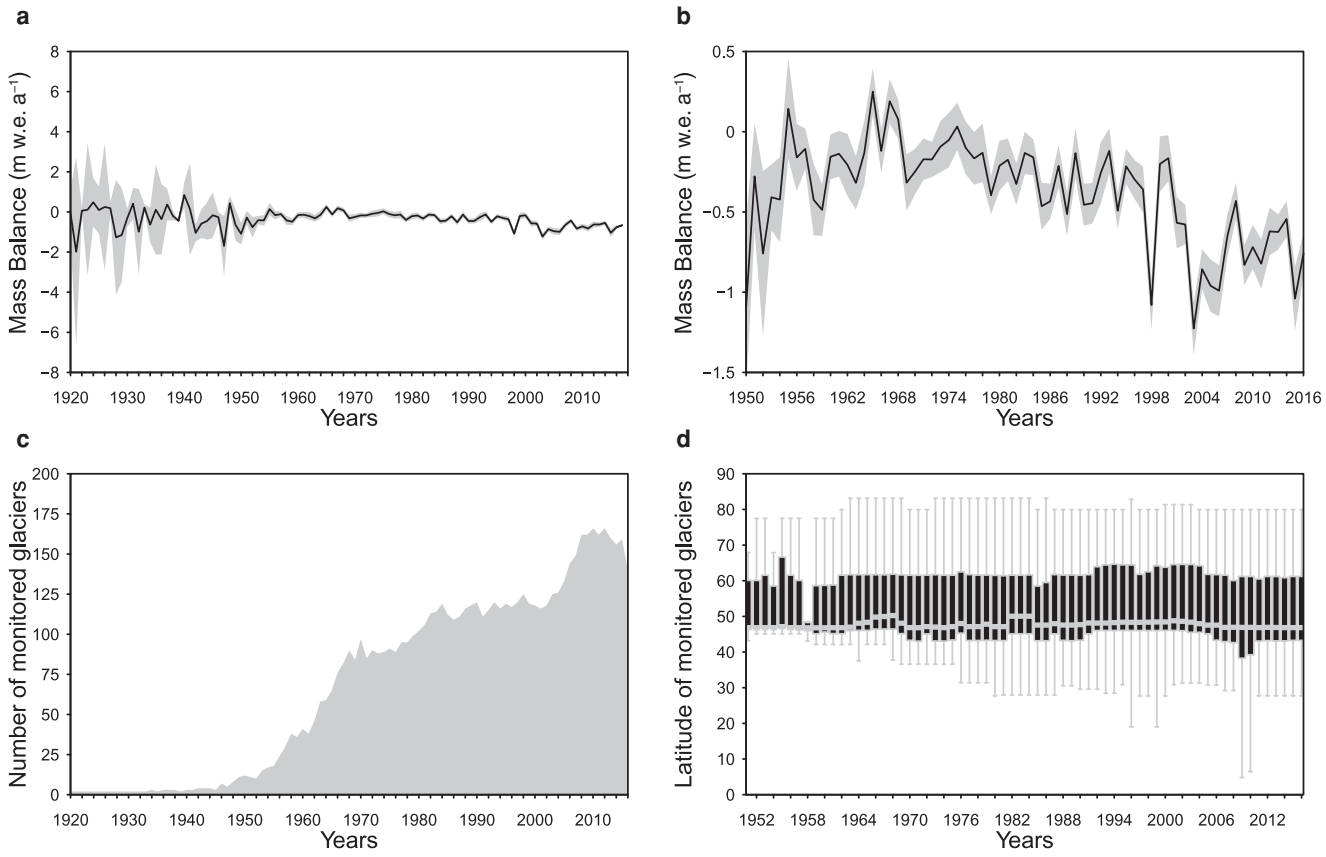



**Fig. 4.** Different views of glacier mass-balance data. (a) Annual average of mass balance for all glaciers included in the WGMS database since 1920; (b) the same data but zoomed in for the period 1950–2017; (c) the total number of monitored glaciers per year since 1920; and (d) the annual latitude distribution of monitored years since 1950 shown as a boxplot per year. In (a) and (b) the gray envelope indicates the 95% confidence interval. In (d), positive latitudes indicate the Northern Hemisphere while outliers, mostly glacier locations of the Southern Hemisphere, are not shown.

As presented in Figure 1, it seems that today there is better global coverage of surface mass balance in the world. Figure 4d provides an alternative spatial view indicating the existence of a biased distribution. Most of the signal recorded in Figure 4b may be actually a product of a majority of glaciers sampled since 1950 being located within a relatively narrow latitudinal band, between 48°N to 62°N more precisely. By comparison, the average latitude of the glaciers included in the RGI is 37° 56'N± 33°23' (1 standard deviation). The narrow band where most glaciers are currently monitored leaves several important regions weakly represented, as for example the High Mountain Asia, having its largest concentration of mountain glaciers in the band 26°N to 42°N. The bias is also evident when inspecting the situation of Southern Hemisphere glaciers. Although the distribution of mass-balance programs has become increasingly even, with a peak in the late 2000s, especially toward low latitudes, all monitored glaciers in the Southern Hemisphere still remain as outliers in each annual boxplot shown in Figure 4d; all boxplot ranges extend along Northern Hemisphere latitudes only. The mean latitude of all monitored glaciers classified as outliers is 32°32'S± 32°31' (1 standard deviation). In the RGI, the mean latitude of outliers is 42°16′S± 13°12' (1 standard deviation), evidencing that most of the tropical and mid-latitude glaciers in the Southern Hemisphere are underrepresented in the current global network of mass-balance programs. The distribution according to latitude of the WGMS reference glaciers is more biased since currently only one glacier of the Southern Hemisphere is included (Fig. 1). Furthermore, the set that contains all monitored glaciers represents more features of the RGI than the reference glaciers, particularly for minimum and maximum elevation, and area (see Supplementary Fig. S1).

When analyzing the EPS and spatial coverage among time series (Fig. 5), the EPS applied to 20-year moving windows indicates this index never goes below 0.65, suggesting that the variance is consistently shared by the individual time series. An abrupt descent occurs at 1965 that does not fully recover until late 1990s, with the most monotonic increase beginning in 1987 (Fig. 5a). Not all glaciers fulfill the requirements of having a continuous 20-year record within a given moving window, and in fact <50% of the reported programs in 2017 account for the signal. The 1965's drop occurs despite the number of glaciers in the calculation almost doubling, from 3 to 5 (Fig. 5b), with a non-significant change in the latitudinal distribution of glaciers relative to the previous years (Fig. 5d). EPS descent toward the minimum in 1987 is punctuated by short periods of increases as the number of monitored glaciers climbs without interruptions to 47 (Fig. 5b), more than nine times in 30 years. After 1987, the monotonic EPS increment follows the number of monitored glaciers (although the maximum of the period 1987–1989 is only surpassed by 2003, Fig. 5b), and the noticeable expansion of the lower whisker in the latitudinal distribution (Fig. 4d). From 2003, the EPS reaches a stable value above 0.9. A $-0.3 (p > 0.005)$ linear correlation between EPS and the number of monitored glaciers per 20-year window since 1920 further indicates that more glaciers do not necessarily mean a stronger common signal. Conversely, it seems that if the growing number of monitored glaciers is accompanied by a better spatial distribution, in this case expressed in the latitude, the signal may become more robust. Since 1987, the linear correlation is 0.69 ($p < 0.001$), even though the average number of available time series only barely doubles during the period 1965–1987, from 22 to 54. Figure 5c reinforces the finding that the signal benefits from a more diverse sample of glaciers.

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

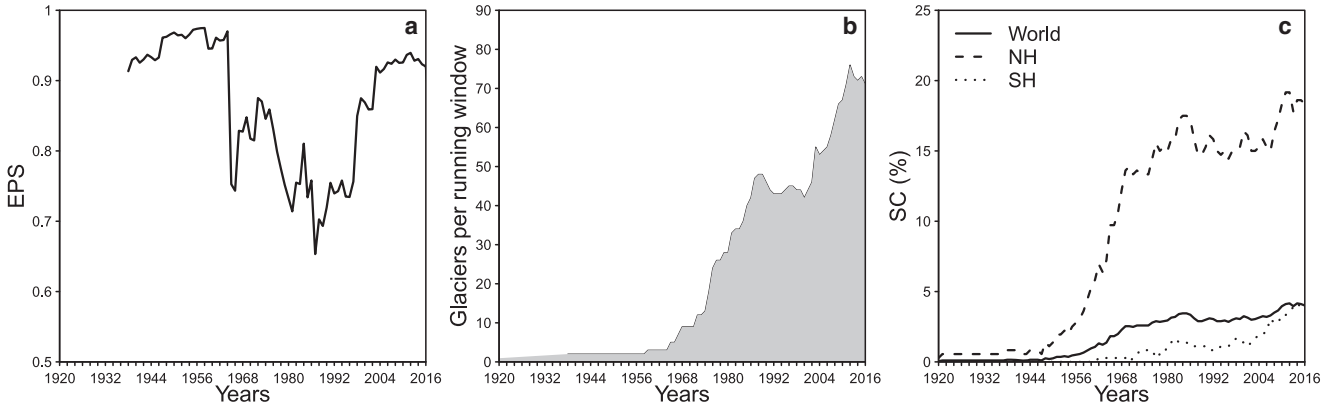



**Fig. 5.** Time series of EPS (a) and spatial coverage (c) considering 20-year running windows, with (b) indicating the number of glaciers used for the calculation on each 20-year window.

Though the spatial coverage curve for the Northern Hemisphere appears to be the one that more closely follows the number of monitored glaciers, showing a correlation of 0.81 since 1920 and a consistent coverage above 15% since the late 1970s, the Southern Hemisphere correlates higher (0.87) while showing a similar shape as the global spatial coverage curve, with a maximum of almost 5% by the end of the studied period. Even more, after 1987, the Southern Hemisphere displays the only positive correlation with EPS that accounts for more than 50% of the variance, with 0.77 ($p < 0.001$), while the world's and Northern Hemisphere spatial coverage are 0.68 and 0.58 (both with $p < 0.001$), respectively. These correlations corroborate the fact that more continuous mass-balance programs are becoming broadly distributed around the world. However, the finding that most programs are within a relatively narrow latitudinal range suggests that certain zones of the RGI sectorization that are used for the calculation of the reference glaciers surface mass balance curve pertain to overrepresented latitudes (notably zones 2, 8, 10 and 11) and therefore the current averaging might be improved by representing the evolving nature of the surface mass balance coverage through an alternative regionalization scheme.

### 3.2. Analysis of the glacier inventory and clustering of the world's glaciers

#### 3.2.1. Correlations among inventory fields

Correlations among pairs of variables from the geodatabase reveal some differences in the relationships at the global scale and between hemispheres (Table 2). The only glacier descriptors showing some consistent moderate to strong correlations with climatic variables are *Zmi*, *Zme* and *Zma*. The high mutual correlation among *Zmi*, *Zme* and *Zma* is a predictable outcome as generally *Zme* is derived from *Zmi* and *Zma*. *Rn* is highly correlated with *Zmi*, *Zme* and *Zma*, with the figures consistently higher for the Southern Hemisphere. The lack of statistical relationship of slope and area with the other glacier descriptors and climatic variables suggests that these morphometric properties play a discriminant role only within regional to local scales, and that most of the glacier distribution is explained by how glacier altitudinal descriptors relate to climate variables. A remarkable correlation pattern occurs with temperatures, since only *Tmin* correlates with *Zmi*, *Zme* and *Zma*, with a moderate negative figure. This negative correlation is stronger for *Zmi* and *Zme* in the Southern Hemisphere relative to the Northern, and this likely reflects a pattern within the Andes mountains where freezing temperatures become more common on glaciers at the highest elevations of the Tropics and Subtropics. This is also detectable for vapor pressure, as the relatively moderate negative correlation with *Zmi*, *Zme* and *Zma* is higher in the Southern Hemisphere, which is concomitant with the strong positive correlation between *Tmin* and *Vap*, also suggesting that glaciers at high elevations are characterized by dry environments that allow energy loses to the atmosphere, leading to lower temperatures. While mean temperatures (*Tav*) do not correlate strongly with any of the glacier descriptors, *Tmax* presents a slight relationship with *Zma* in the Southern Hemisphere. The highest linear correlation between glacier and climate parameters is in *Rad*, but in this case with the Northern Hemisphere reaching the highest values. The high correlation of *Rad* with *Zmi*, *Zme* and *Zma* reaffirms the fact that radiation is the most important component of the energy balance (Schaefer and others, 2020), as it at least explains 69% of the variance in glacier descriptors. The case with precipitation is noteworthy, as global correlations are the same for *Zmi*, *Zme* and *Zma*, −0.45. That variable correlates higher in the Northern Hemisphere, probably reflecting a stronger continentality control on precipitation. In fact, the correlation of precipitation with longitude is of opposite sign depending on the Hemisphere, negative in the Northern reflecting a tendency to decrease on glaciers located more inland. On the other hand, the increase in the Southern Hemisphere is probably related to the fact that glaciers in the Subtropical Andes and the Southern Alps are at a more similar distance to the ocean and that there is a significant rain shadow effect given the height of these mountains (Viale and Nuñez, 2011; Caloiero, 2014; Schumacher and others, 2020). This is also seen in the correlations between *Zmi*, *Zme* and *Zma* relative to latitude and longitude. At the planetary scale, latitude does not correlate with *Zmi*, *Zme* and *Zma* unlike longitude. In the Northern Hemisphere, both coordinates correlate significantly with *Zmi*, *Zme* and *Zma* while only latitude does it in the Southern Hemisphere. A final noteworthy finding is about wind speed (*Wind*), a climate parameter that presents an identical moderate negative correlation with *Zmi*, *Zme* and *Zma* globally, but that behaves quite differently depending on the hemisphere. While in the Northern Hemisphere Wind practically does not correlate with glacier parameters, in the Southern Hemisphere correlations are almost as high as *Tmin* relative to *Zmi*, *Zme* and *Zma*, following correlations associated with *Rad*.

#### 3.2.2. Multivariate analysis of the inventory using PCA

From the previous exploratory analysis, it becomes clear that the variables related to glacier morphometry, namely *Area*, Slope (*Sl*), Aspect (*Asp*), Length (*Lm*) and the hypsometric integral (*Hyp*), do not correlate with the climate parameters included in the geodatabase, neither at global nor at hemispheric scales. This finding is also reinforced with the PCA that includes all variables indicated in Table 1. The fact that six components are needed to

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

**Table 2.** Linear correlations among the variables of the compiled database for the world, northern (NH) and southern (SH) hemispheres

| **World** | *Lat* | *Area* | *Zma* | *Zme* | *Zmi* | *Sl* | *Asp* | *Lm* | *Hyp* | *Rn* | *Tmin* | *Tmax* | *Tav* | *Wind* | *Vap* | *Rad* | *Pp* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Lon* | − 0.16 | − 0.03 | 0.69 | 0.68 | 0.68 | 0.15 | − 0.01 | − 0.07 | − 0.01 | 0.51 | − 0.37 | − 0.13 | − 0.26 | − 0.12 | − 0.3 | 0.68 | − 0.3 |
| *Lat* | | 0 | − 0.19 | − 0.18 | − 0.16 | − 0.14 | 0.04 | 0.1 | 0.02 | − 0.26 | − 0.44 | − 0.55 | − 0.51 | − 0.52 | − 0.38 | − 0.29 | − 0.3 |
| *Area* | | | − 0.08 | − 0.04 | − 0.06 | − 0.09 | 0 | 0.58 | 0.01 | − 0.05 | − 0.05 | − 0.08 | − 0.07 | 0.05 | − 0.03 | − 0.06 | − 0.01 |
| *Zma* | | | | 0.98 | 0.99 | 0.24 | − 0.01 | − 0.18 | − 0.02 | 0.81 | − 0.33 | 0.04 | − 0.15 | − 0.32 | − 0.42 | 0.92 | − 0.45 |
| *Zme* | | | | | 0.99 | 0.24 | − 0.01 | − 0.06 | − 0.02 | 0.79 | − 0.35 | 0.01 | − 0.17 | − 0.32 | − 0.43 | 0.91 | − 0.45 |
| *Zmi* | | | | | | 0.24 | − 0.01 | − 0.12 | 0 | 0.8 | − 0.35 | 0.02 | − 0.17 | − 0.32 | − 0.43 | 0.92 | − 0.45 |
| *Sl* | | | | | | | 0.02 | − 0.25 | 0 | 0.16 | 0.11 | 0.18 | 0.15 | − 0.05 | 0.03 | 0.23 | 0.05 |
| *Asp* | | | | | | | | 0 | 0.02 | − 0.02 | − 0.01 | − 0.01 | − 0.01 | − 0.03 | 0 | − 0.02 | 0.01 |
| *Lm* | | | | | | | | | 0.02 | − 0.14 | − 0.14 | − 0.21 | − 0.18 | 0.05 | − 0.09 | − 0.15 | − 0.05 |
| *Hyp* | | | | | | | | | | − 0.05 | − 0.09 | − 0.11 | − 0.1 | 0.06 | − 0.04 | − 0.06 | − 0.02 |
| *Rn* | | | | | | | | | | | − 0.22 | 0.24 | 0.02 | − 0.27 | − 0.21 | 0.83 | − 0.33 |
| *Tmin* | | | | | | | | | | | | 0.89 | 0.97 | 0.21 | 0.84 | − 0.19 | 0.72 |
| *Tmax* | | | | | | | | | | | | | 0.97 | 0.08 | 0.73 | 0.19 | 0.57 |
| *Tav* | | | | | | | | | | | | | | 0.15 | 0.81 | 0 | 0.66 |
| *Wind* | | | | | | | | | | | | | | | 0.24 | − 0.27 | 0.25 |
| *Vap* | | | | | | | | | | | | | | | | − 0.27 | 0.74 |
| *Rad* | | | | | | | | | | | | | | | | | − 0.31 |
| **NH** | | | | | | | | | | | | | | | | | |
| *Lon* | − 0.67 | − 0.03 | 0.76 | 0.76 | 0.76 | 0.15 | − 0.02 | − 0.08 | − 0.01 | 0.64 | − 0.4 | − 0.09 | − 0.26 | − 0.04 | − 0.35 | 0.75 | − 0.54 |
| *Lat* | | 0.09 | − 0.93 | − 0.93 | − 0.93 | − 0.28 | 0.02 | 0.18 | 0.08 | − 0.82 | − 0.02 | − 0.41 | − 0.22 | 0.28 | 0.09 | − 0.95 | 0.21 |
| *Area* | | | − 0.09 | − 0.04 | − 0.06 | − 0.09 | 0 | 0.64 | 0.02 | − 0.06 | − 0.06 | − 0.09 | − 0.08 | 0.04 | − 0.02 | -0.07 | − 0.01 |
| *Zma* | | | | 0.98 | 0.99 | 0.27 | − 0.03 | − 0.19 | − 0.02 | 0.83 | − 0.26 | 0.13 | − 0.06 | − 0.23 | − 0.37 | 0.93 | − 0.45 |
| *Zme* | | | | | 0.99 | 0.27 | − 0.02 | − 0.07 | − 0.02 | 0.81 | − 0.27 | 0.11 | − 0.08 | − 0.23 | − 0.38 | 0.92 | − 0.45 |
| *Zmi* | | | | | | 0.27 | − 0.02 | − 0.13 | 0 | 0.82 | − 0.28 | 0.12 | − 0.08 | − 0.23 | − 0.38 | 0.93 | − 0.46 |
| *Sl* | | | | | | | 0.03 | − 0.25 | 0 | 0.19 | 0.09 | 0.18 | 0.14 | − 0.11 | − 0.01 | 0.26 | − 0.02 |
| *Asp* | | | | | | | | 0 | 0.03 | − 0.03 | 0.01 | 0 | 0 | 0 | 0.02 | − 0.03 | 0.02 |
| *Lm* | | | | | | | | | 0.03 | − 0.16 | − 0.14 | − 0.21 | − 0.18 | 0.09 | − 0.07 | − 0.16 | − 0.03 |
| *Hyp* | | | | | | | | | | − 0.05 | − 0.14 | − 0.16 | − 0.15 | 0.07 | − 0.09 | − 0.05 | − 0.05 |
| *Rn* | | | | | | | | | | | − 0.2 | 0.28 | 0.04 | − 0.24 | − 0.19 | 0.84 | − 0.38 |
| *Tmin* | | | | | | | | | | | | 0.89 | 0.97 | − 0.19 | 0.8 | − 0.14 | 0.72 |
| *Tmax* | | | | | | | | | | | | | 0.97 | − 0.31 | 0.69 | 0.26 | 0.53 |
| *Tav* | | | | | | | | | | | | | | − 0.26 | 0.77 | 0.06 | 0.65 |
| *Wind* | | | | | | | | | | | | | | | − 0.08 | − 0.23 | − 0.01 |
| *Vap* | | | | | | | | | | | | | | | | − 0.21 | 0.74 |
| *Rad* | | | | | | | | | | | | | | | | | − 0.37 |
| **SH** | | | | | | | | | | | | | | | | | |
| *Lon* | 0.05 | − 0.01 | − 0.06 | − 0.08 | − 0.03 | 0.3 | 0.09 | − 0.07 | − 0.02 | − 0.1 | 0.31 | 0.23 | 0.31 | − 0.05 | 0.24 | 0.21 | 0.78 |
| *Lat* | | 0.11 | − 0.91 | − 0.9 | − 0.92 | − 0.14 | − 0.09 | 0.14 | − 0.01 | − 0.86 | 0.25 | − 0.56 | − 0.16 | 0.64 | 0.16 | − 0.75 | 0.19 |
| *Area* | | | − 0.07 | − 0.04 | − 0.05 | − 0.11 | 0.01 | 0.6 | − 0.03 | − 0.03 | − 0.15 | − 0.19 | − 0.19 | 0.06 | − 0.09 | − 0.03 | − 0.04 |
| *Zma* | | | | 0.97 | 0.99 | 0.07 | 0.07 | − 0.14 | − 0.02 | 0.88 | − 0.52 | 0.3 | − 0.14 | − 0.51 | − 0.48 | 0.88 | − 0.35 |
| *Zme* | | | | | 0.99 | 0.07 | 0.08 | 0 | -0.04 | 0.87 | − 0.55 | 0.25 | − 0.19 | − 0.5 | − 0.5 | 0.86 | − 0.36 |
| *Zmi* | | | | | | 0.07 | 0.08 | − 0.07 | − 0.01 | 0.88 | − 0.54 | 0.28 | − 0.16 | − 0.51 | − 0.49 | 0.87 | − 0.36 |
| *Sl* | | | | | | | − 0.02 | − 0.24 | 0.03 | − 0.01 | 0.24 | 0.25 | 0.27 | − 0.07 | 0.22 | 0.07 | 0.28 |
| *Asp* | | | | | | | | 0.03 | − 0.02 | 0.09 | − 0.03 | 0.05 | 0.01 | − 0.09 | 0.03 | 0.07 | 0.08 |
| *Lm* | | | | | | | | | − 0.06 | − 0.07 | − 0.21 | − 0.29 | − 0.28 | 0.09 | − 0.14 | − 0.08 | − 0.09 |
| *Hyp* | | | | | | | | | | − 0.04 | 0.12 | 0.09 | 0.12 | 0.05 | 0.11 | − 0.08 | 0.03 |
| *Rn* | | | | | | | | | | | − 0.53 | 0.4 | − 0.09 | − 0.62 | − 0.4 | 0.79 | − 0.3 |
| *Tmin* | | | | | | | | | | | | 0.56 | 0.89 | 0.09 | 0.84 | − 0.49 | 0.59 |
| *Tmax* | | | | | | | | | | | | | 0.87 | − 0.51 | 0.52 | 0.24 | 0.34 |
| *Tav* | | | | | | | | | | | | | | − 0.23 | 0.78 | − 0.16 | 0.53 |
| *Wind* | | | | | | | | | | | | | | | − 0.07 | − 0.51 | 0.02 |
| *Vap* | | | | | | | | | | | | | | | | − 0.51 | 0.56 |
| *Rad* | | | | | | | | | | | | | | | | | − 0.13 |

Correlations for latitude and longitude were calculated using absolute values. Meaning of abbreviations can be found in Table 1.

capture more than 70% of the variability, with the first three components summing up only 46.1%, suggests that some of the global signal in glacier distribution is hidden by local morphometric effects (Fig. 6a). This can be seen when comparing the variables with the highest absolute weight in each PC. While the variables with the highest correlations relative to *Zmin*, *Zme* and *Zma* appear important in PC1 and PC2 (e.g. *Rad*), morphometric variables have minimal influence, with *Lm* and *Area* showing up strongly in PC3 (with positive moderate mutual correlation in all cases), Hyp and *Asp* in PC4, and *Sl* in PC6. This result allows us to infer that, in order to maximize explained variance of the glacier distribution, the dimensionality reduction should only include *Zmin*, *Zme*, *Zma* and climatic parameters.

Results of the second PCA now indicate that 71.57% of the variance is explained with only three PCs (Fig. 6b). The first PC (32.9%) of this subset clearly shows a direct relationship between glacier parameters and radiation while inverse with respect to *Tmin*, *Vap* and *Pp*. Radiation's direct relationship is logical when interpreted in the light of *Vap* and *Pp* inverse relationship with glacier parameters, indicating the fact that elevations tend to correlate with drier, clear skies. Conversely, the opposite relationship with temperatures, and especially with *Tmin*, can be interpreted as the general decrease of temperatures according to the increase in elevation and corroborates the finding described in the previous section. The PC2 (25.84%), on the other hand, highlights the relationships among climatic variables, especially for temperatures. The spatial distribution of the PC scores of every glacier within the geodatabase evidences certain patterns worth highlighting. In the PC1 (Fig. 7a), there seem to be well differenced two groups, with most positives above 0.4 spreading South America between 15°S to 38°S and High Mountain Asia, and negatives below − 0.4 mainly occurring in

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

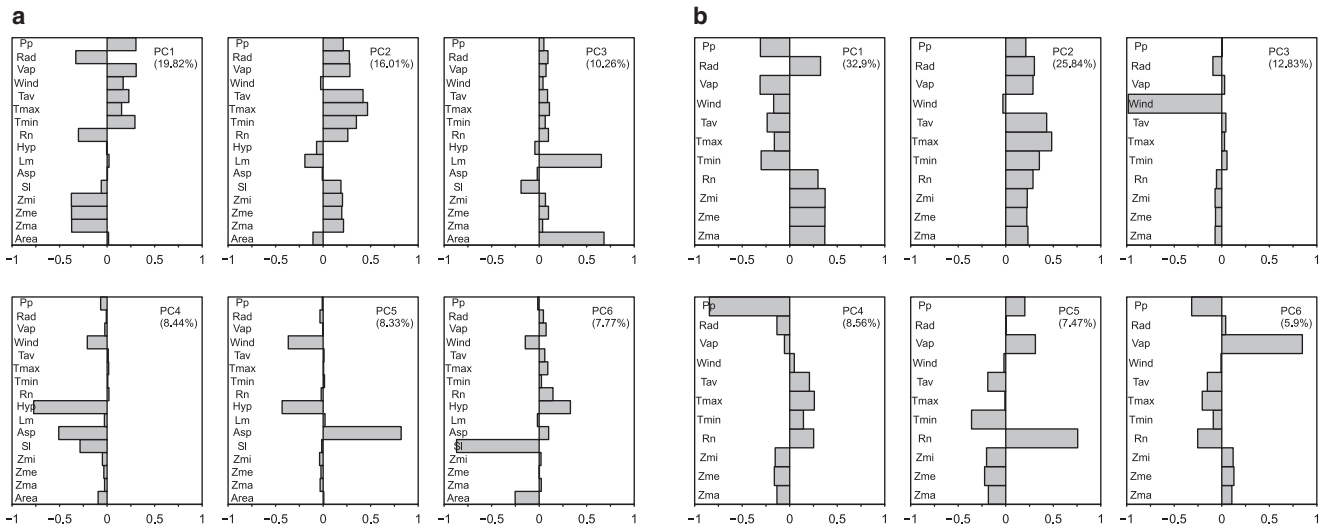



**Fig. 6.** Results of the two PCA attempts. Panels PC1 to PC6 located on the (a) side of the figure show the results including morphometric descriptors. Panels PC1 to PC6 on the (b) side correspond to the second attempt, output utilized in the rest of the analysis for this study.

North America, the South Coast of Greenland, Iceland, Scandinavia, Svalbard, Indonesia, Patagonia, New Zealand and the Antarctic Peninsula. The rest of the glaciers are generally between −0.2 to 0.2. In turn, the PC2 clearly discriminates the polar regions from the rest of the world, as almost all glacier polygons poleward 70°N/S show values negative below −0.6, while almost all other regions are between 0 and 0.4 (Fig. 7b). Extreme negative numbers in the PC3 are overwhelmingly located in the continuum from the Andes to the Antarctic Peninsula, followed by a secondary cluster in Iceland (Fig. 7c). In the PC3 too, there is a relatively well-defined gradient from positive (West) to negative (East) in the Pacific coast of North America and in High Mountain Asia.

### 3.2.3. From self-organizing maps to glacier clusters

Two attempts to determine the optimal number of clusters indicated similar results. When using the first three PCs (71.57%) or the first four PCs (80.13%), NbClust found three clusters to be the optimal number in 99 and 98% of the iterations, respectively. With the Kohonen SOM, 11 (48%) of the NbClust's indices also find that the most suitable number of clusters is 3, followed by 5 (2.2%). The hierarchical clustering delivers a set of three groups with very few glaciers located within regions where they seem to not fit (Fig. 8). Our approach finds that at the global scale High Mountain Asia glaciers can be classified together with the southernmost section of the Rocky Mountains, Zagros Mountains, most Tropical regions and the arid subtropical Andes north from ~36°S. However, it is worth mentioning that a number of High Mountain Asia glaciers pertain to a different group.The Pacific coast of the American Continent and the West Antarctic Peninsula contain the largest proportion of glaciers in the cluster 2, a group that also includes all New Zealand glaciers, almost all glaciers located in continental Europe, Iceland and the southern third of Greenland. Glaciers located south/north of the ~70th parallel in both hemispheres configure a group where Antarctica's glaciers appear similar to those in Arctic Canada, Central and North Greenland, and most Arctic archipelagos. Within the Antarctic peninsula there is a clear separation between groups 3 to the East and 2 to the West. These results also reveal discrepancies relative to the RGI zonation, as for example in the first-order region 2 (Western Canada and USA) our method suggests only two subgroups compared to 5 from the global inventory, with the section containing the southern part of the Rocky Mountains roughly coinciding with the second-order 02–05. In Greenland (region 5), our results firmly split the region in two sections while currently these glaciers are considered as one unit. High Mountain Asia is other remarkable case with all three regions there pertaining to the same cluster. In the Southern Hemisphere, there is a clear mismatch between the limits that our method finds and those utilized in the RGI, in particular the limit between first-order regions 16 and 17, since our clustering suggests the boundary is around 36°S. The clustering also shows that in very few cases glaciers are located in large regions where a majority pertains to a different cluster. For example, nearly two dozen of glaciers in Alaska that are classified as from cluster 1 show up in the middle of a large area of cluster 2, a few Ecuadorian glaciers classified in cluster 2 are within this large region of cluster 1, and there are certain mixes in Kamchatka Peninsula. However, there is almost no case where a glacier of cluster 1 lies within a region of cluster 3 or vice versa (more detailed view of the clustering can be observed in Figs S2–S20).

It is evident the inverse relationship between the total surface area of each cluster and the number of glaciers contained (Fig. 9), where glaciers of the cluster 3 concentrate the largest area with the smallest number of associated polygons. But, what does our procedure tell about glacier regimes that characterize the regions presented in Figure 8? In order to attempt an explanation on the reasons, we need to dig into the variables utilized in the dimensionality reduction and clustering processes (Fig. 9). Although in no case the variables presented as boxplots in Figure 9 suggest significant statistical differences in the distribution of variables according to each cluster, it is possible to identify distinct patterns. In cluster 1, *Zme*, *Zmi* and *Zma* tend to be on average 3000 m higher than the other groups, with larger temperature range and radiation overall. In turn, salient features in cluster 2 are the generally higher temperatures, slightly more remarkable for *Tmin*, and also higher *Vap* and precipitation. Somewhat of a mixture in certain variables occurs for cluster 3, a group that while evidencing differences on the three temperature descriptors relative to the other clusters, coincides with cluster 2 regarding glacier elevations, and in *Vap* and precipitation with cluster 1. Therefore, cluster 1 seems to represent a group of generally high elevation glaciers located in relatively cold, dry areas with large radiation input, also characterized by significant temperature range. Subsequently, cluster 2 features suggest a regime of low elevation glaciers with more frequency of temperatures above zero on relatively humid climates. Finally, cluster 3 represents glaciers located very close to sea level, in cold and dry areas with the lowest average radiation input.

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

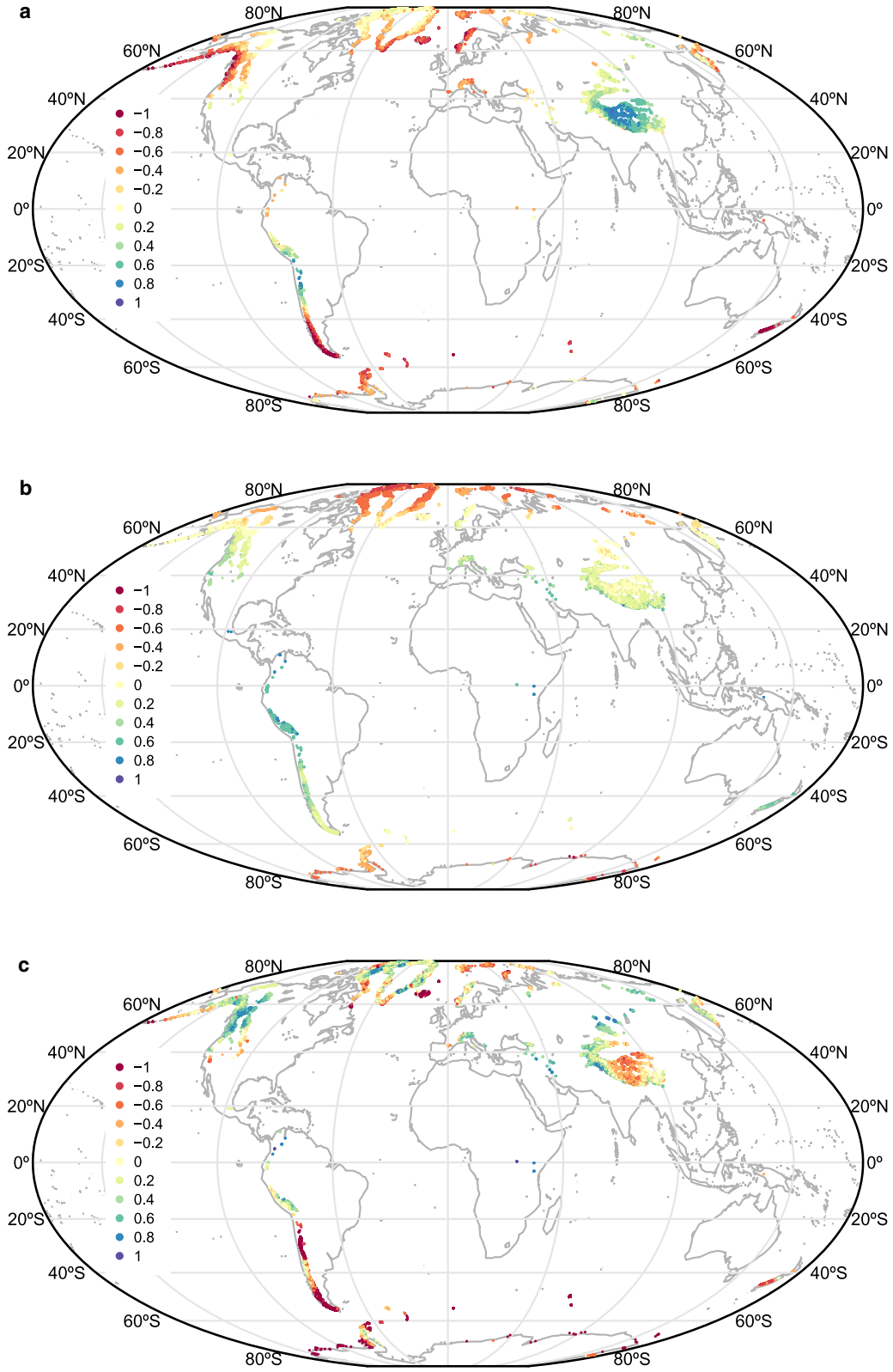



**Fig. 7.** Global maps displaying the spatial distribution of glaciers, color-coded according to their respective score in the PC1 (a), PC2 (b) and PC3 (c), as a result of the second PCA attempt.

### *3.3. Temperature sensitivity per cluster*

One obvious outcome of the clustering procedure is that glaciers do not conform regions or zones with strict limits and rather it seems that this data-driven approach indicates long distance connections while unveiling particular regimes associated with elevation, moisture availability and radiation. It follows that agglomerating glaciers with similar regime relative to the global population should allow for building surface mass balance composites more representative of the glacio-climatic regime diversity at this large scale. Our results also allow us to suggest that surface mass balance for glaciers within a similar regime should respond in a roughly similar fashion to climate and thus average composites of all available time series would be more representative of

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

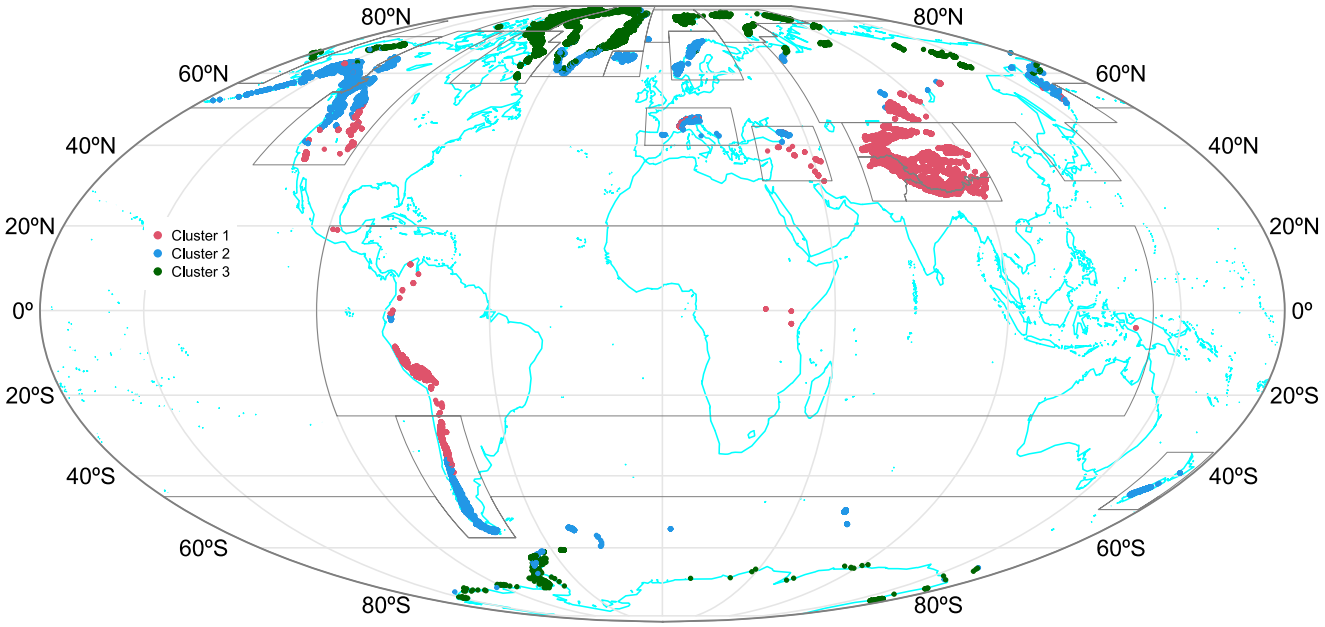



**Fig. 8.** Clustering results over the world's map. Each glacier is colored according to the cluster they are classified. Boundaries of RGI regions included in light gray.

sensitivity to climate changes. Figure 10 presents an analysis of surface mass balance for each cluster, where the proportion of glaciers per year since 1950 reveals the fact that all regimes are represented since that year, although cluster 3 presents systematically less members (Fig. 10a). surface mass balance composites for each proposed regime show a decreasing trend which in all cases becomes more evident since late 1960s and even more notorious after 1980 (panels b to d). The linear trend 1950–2017 for each group is only statistically significant for clusters 1 and 2, with $-8.84$ and $-9.26$ mm w.e. $a^{-1}$ ($p < 0.001$ in both cases) respectively, while for cluster 3 it is $-2$ mm w.e. $a^{-1}$ ($p > 0.1$). However, after 1980 all groups' negative trends speed-up, with $-14.33$ mm w.e. $a^{-1}$ ($p < 0.001$) for cluster 1, $-21.29$,mm w.e. $a^{-1}$ ($p < 0.001$) for 2 and $-12.18$ mm w.e. $a^{-1}$ ($p < 0.01$) for 3, i.e. from almost twofold to sixfold increase. Cluster 3 shows the largest uncertainty bounds, expressed in the 95% confidence interval. Computation of the EPS with 20-year windows for each group reaffirms the finding on uncertainty in cluster 3 (panel e). While within each 20-year window EPS remains above 0.6 for much of the period since 1960 for clusters 1 and 2, in 3 the shared variance actually shows a declining trend. Reasons behind this result are perhaps the fact that for every 20-year window <6 glaciers contribute to the cluster composite and that only by 1970 there is enough number of continuous records allowing calculation of the respective surface mass balance composites (Figs 10f–h).

Figure 11 shows global surface mass balance time series computed using different combinations of available data, which allows comparison of surface mass balance trends. A global trend averaging the three clusters (red curve in Fig. 11) delivers $-6.27$ mm w.e. $a^{-1}$ ($p < 0.001$), i.e. less pronounced than the trend computed from the mean of reference glaciers (black curve in Fig. 11), $-7.54$ mm w.e. $a^{-1}$ ($p < 0.001$). The cumulative surface mass balance of the clusters' average composite is also less negative than that determined from the reference glaciers, $-25.74$ versus $-28.04$ m w.e.; by comparison, the same figure calculated from averaging all glaciers per year (blue curve in Fig. 11) results in a loss equivalent to $-26.07$ m w.e. On average, the difference between the global surface mass balance composite computed from the clusters and the reference glaciers' time series is 0.034 m w.e., with a linear correlation of 0.71 ($p > 0.001$). Unlike the global mean computed from either reference glaciers or all glaciers, which present negative surface mass balance since mid 1970s, the composite from clusters evidences this since 1986, after two positive spikes. Since 2000, our averaging points to a stronger negative trend.

Unlike calculation of sensitivity using modeling approaches, where simulations using tweaks of climatic variables are used to calculate the rate of change of surface mass balance, in the regression approach we test the behavior of the β parameter. In addition, self-correlations and cross-correlations can be indicative of the hysteresis of each individual variable and the delay of the response variable adjusting to a given stimulus from the independent variable. In our case, we interpret this time window, should it exist, as the temporal scale where sensitivity maximizes. A first hint can be appreciated in Figure 12, where both the global mean of each climatology and the average surface mass balance curve per cluster show significant self-correlations. At the 95% confidence level, all three climatologies show statistical relationships with delays up to 15 years, with the stronger values (>0.65) within a 4-year window. For surface mass balance, significant self-correlations show up from lag-1 to about decadal lags for clusters 1 and 2, while for cluster 3 only at lags 1 and 4, although in all cases these self-correlations plummet after lag-1 when compared to the climatologies.

Cross-correlations between the climatologies and the surface mass balance per cluster reveal number 3 with the smallest figures (Fig. 13). The plots suggest a slightly significant cross-correlation at the 95% confidence level for a window from $-2$ to $+7$ years. Wider windows and higher correlations can be seen for the mean from the WGMS reference, clusters 1 and 2. While at the zero lag the values are near 0.6, windows extending to $\sim -5$ to $+10$ years remain above 0.4, and even $-10$ to $+15$ years are still significant for clusters 1 and 2 considering all climatologies. These cross-correlations allow us to infer that at this global scale the surface mass balance average time series have a memory of ~15–20 years relative to global air temperature averages, supporting the choice of 20-year windows for most of the time series analysis included in this paper. In addition, this result may mean that surface mass balance sensitivity corresponds to a non-stationary property of glacier response.

Using the 20-year window for calculating moving linear correlations since 1950 further support the finding of surface mass

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

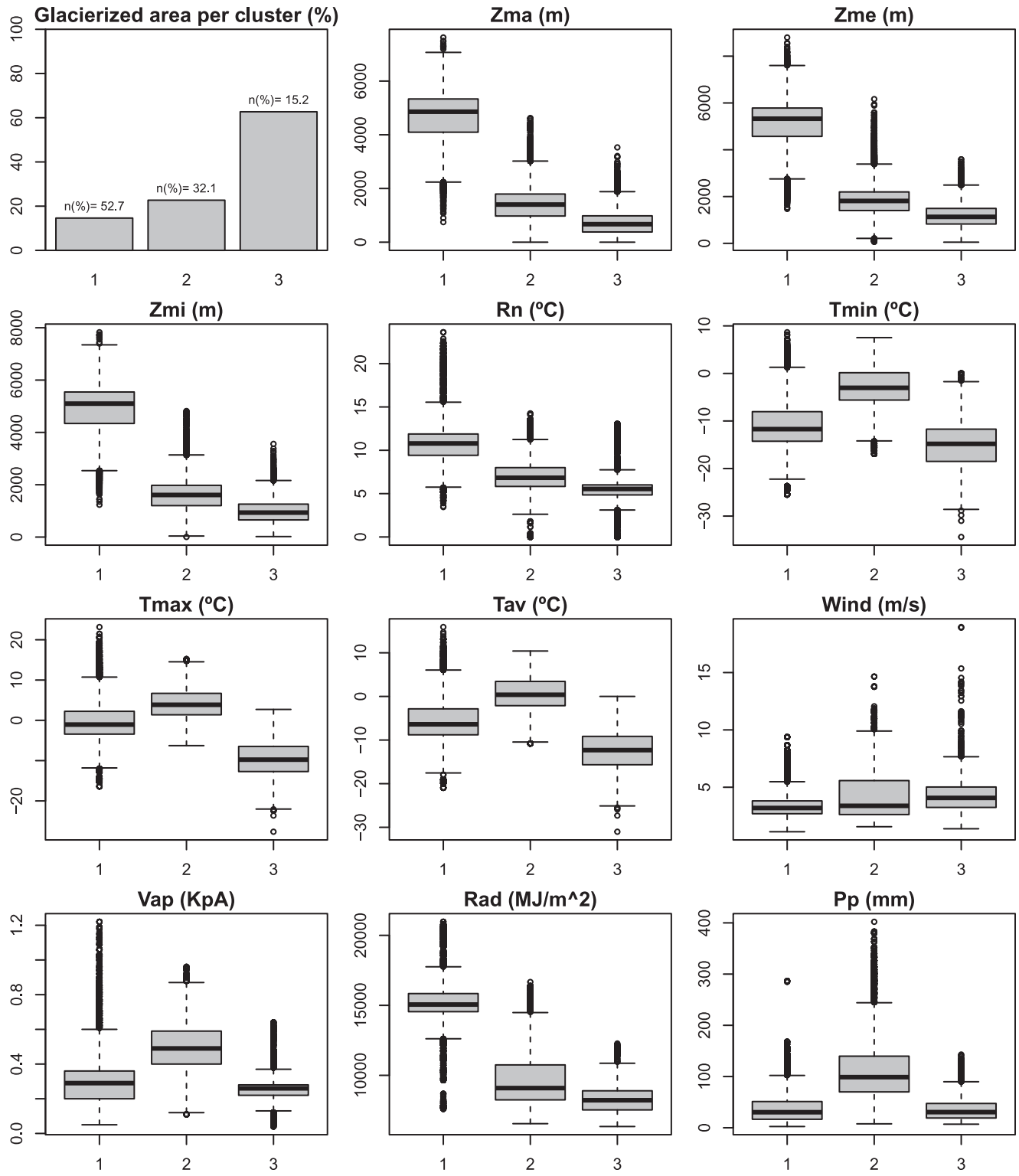



**Fig. 9.** Statistical characterization of clusters (indicated in the *x*-axis) according to the input data used for the classification. The top-left panel is a barplot where the bars represent the percentage of global glacierized area covered by each cluster, while the 'n' on the top of each bar indicates the percentage of glaciers. The rest of the panels are boxplots showing clusters' characteristics for each variable included in the PCA second attempt (see Fig. 6). Abbreviations are according to Table 1.

balance sensitivity as non-stationary. Figure 14 displays these linear correlations computed for the WGMS average and the three clusters, relative to each global climatology. Considering only the climatologies, results are remarkably similar for each cluster. The closer to the present, the correlations seem to become more negative, revealing a stronger inverse relationship with temperature. However, the clusters do not show the positive values between 1965 and 1975 that can be appreciated for the WGMS average. In fact, even averaging the three clusters this positive spike does not show up, indicating that the averaging process using the clusters identifies a slightly different surface mass balance sensitivity. While in all clusters the correlation remains consistently negative, with very few exceptions, becoming statistically significant ($p<0.005$) during the 1980s, for the WGMS-CRUTEM correlation that only occurs in 20-year periods starting in 1978, 1996 and 1997. However, considering all the time series only in

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

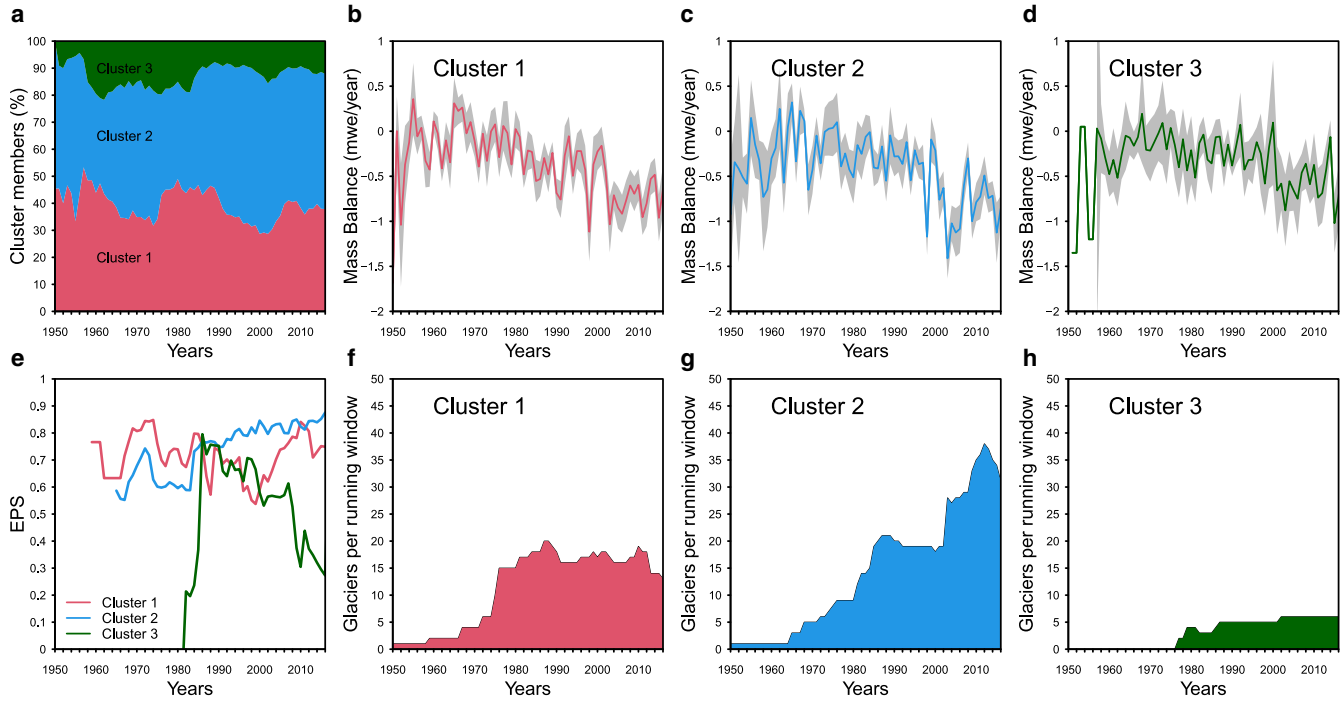



**Fig. 10.** Mass-balance characterization according to the corresponding cluster. Panel (a) indicates the proportion of measured glaciers per cluster relative to the respective annual global total, (b)–(d) show the mass-balance time series per cluster since 1950, with the gray envelope indicating the 95% confidence interval; (e) the annual evolution of EPS per cluster; while (f)–(h) the number of glaciers considered in the calculation per year.

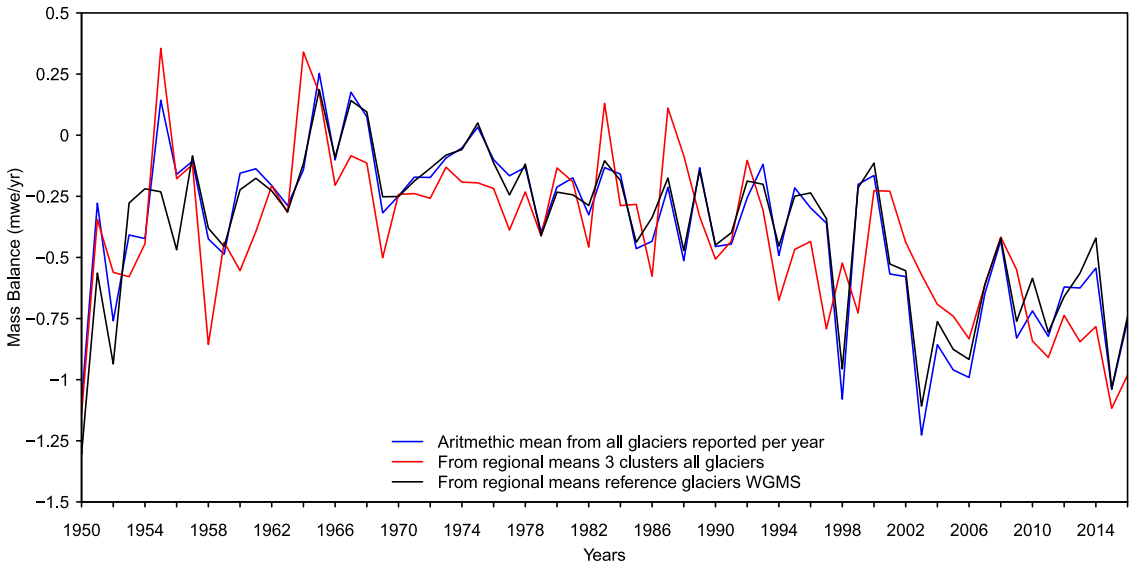



**Fig. 11.** Mass-balance curves calculated with different combinations of annual data. The blue line is the arithmetic average using all glaciers reported per year since 1950, the red line is the average calculated from annual means of each cluster, while the black line is the average of the reference glaciers available from the WGMS database.

a couple of cases the explained variance was higher than 50%, all of them during the 1980s too. Another noticeable finding for that decade is that it is the only one in which the 95% uncertainty envelope is always negative. For cluster 3's average, however, negative values only appear after 1956, which is probably linked to the fact that fewer surface mass balance time series were available during those years. For all the clusters, windows starting in the late 1990s tend to show fewer negative correlations. These results are somewhat different when compared to the correlations calculated for the whole period (Table 3), as the WGMS average from reference glaciers shows values close to the correlations for the 1980s in the 20-year window analysis. The global average built from the clusters displays higher and more consistent correlations with respect to each climatology (− 0.62). Analyzed separately per cluster, number 3 shows the lowest correlation, below − 0.3, much smaller than the values computed with the 20-year window. For cluster 1, correlations are close to those for the early 1980s in Figure 14, while for cluster 2, all are nearly the same as for the period 1980–1990.

Slopes of regression models using 20-year windows also reveal a time-varying sensitivity to global climatology trends for each cluster (Fig. 15). Statistically significant slopes (at $p <$ 0.005) are indicative of maximum sensitivity to temperatures of global climatologies since the 1980s, with values larger than

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

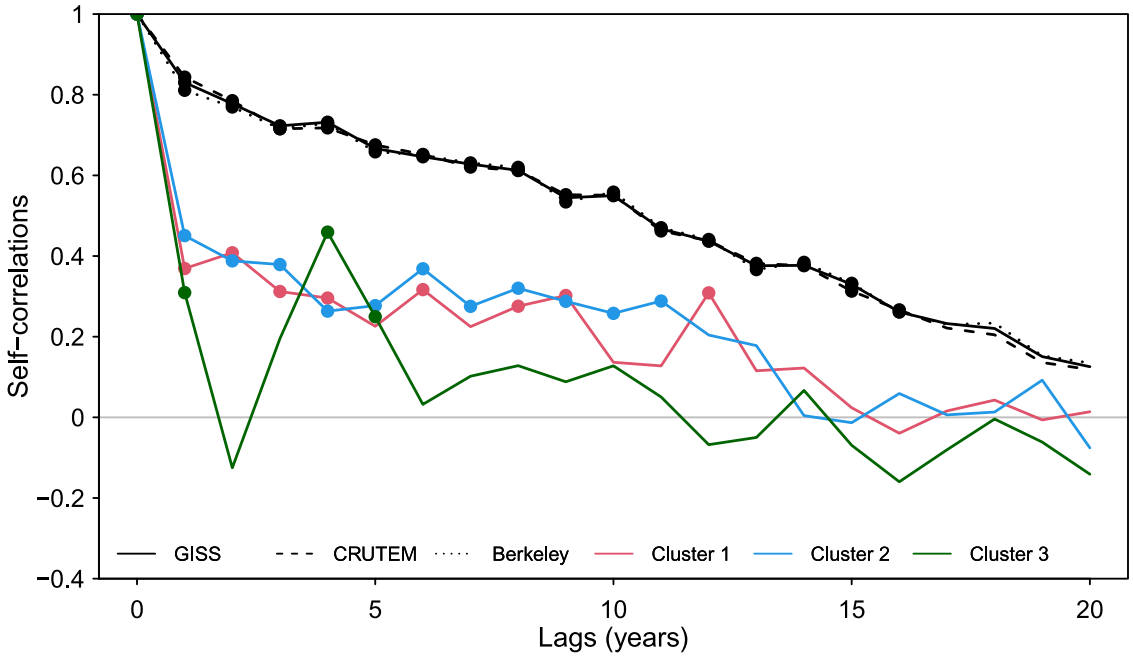



**Fig. 12.** Self-correlations since 1950 for all climatologies and mass-balance time series of each cluster. Dots indicate correlations significant at the 95% confidence level.

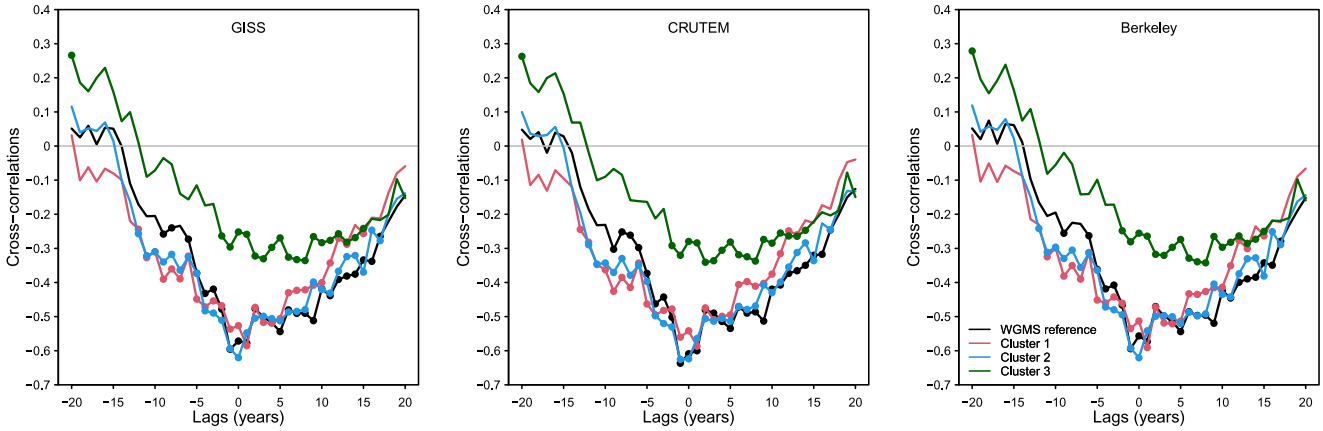



**Fig. 13.** Cross-correlations since 1950 between clusters' mass-balance time series and climatologies, with the reference glaciers' time series also included. Dots indicate correlations significant at the 95% confidence level.

$-0.5\ \mathrm{m\ K^{-1}}$. The Durbin–Watson test applied to the residuals indicates serial correlation for the linear model relating cluster 1 and Berkeley at the beginning of the study period, and for all the models in cluster 3, prior to 1960. As with linear correlations, in all groups slopes tend to be more negative toward the present, although there is an incipient upward trend after 1996. For the three clusters, the 95% confidence interval becomes negative only after 1980. The confidence interval in all clusters also reveals that the average, although different, has considerable overlap. Linear models applied to the whole study period since 1950 for each clusterreveal that for cluster 2 with the highest figures, larger than $-0.49\ \mathrm{m\ K^{-1}}$, followed by cluster 1, with non-statistically significant slopes ($p > 0.1$), and then by cluster 3, group with slopes smaller than $-0.21\ \mathrm{m\ K^{-1}}$, i.e. less than half the sensitivities of the other two (Table 3). The most negative slopes for cluster 1 are larger than $-0.8\ \mathrm{m\ K^{-1}}$ in the late 1980s and early 1990s. In turn, slopes for cluster 2 are generally the most negatives of all, with statistically significant values as low as $-1.1\ \mathrm{m\ K^{-1}}$ since the mid-1980s. Cluster 3 presents similar values as cluster 1, although this is the only case where linear models in particular 20-year explain more than 50% of the variance, periods starting in 1986 and 1987 for CRUTEM, and 1986 for GISS and Berkeley.

## 4. Discussion

### 4.1. Interpretation of the clustering results

The more surface mass balance data available and the increasing coverage of the glacier inventory motivated us to put forward a somewhat different take on studying sensitivity to temperature at the global scale. We assess that the exploratory analysis of available data, the combination of machine learning for classification and computation of mass balance according to the results of the clustering allow for an alternative and balanced appraisal of the relationship between temperature and glaciers of comparable validity relative to the most common approach using modeling (e.g. Oerlemans and Fortuin, 1992). While our exploratory analysis corroborates the notion of a much wider coverage of mass-

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

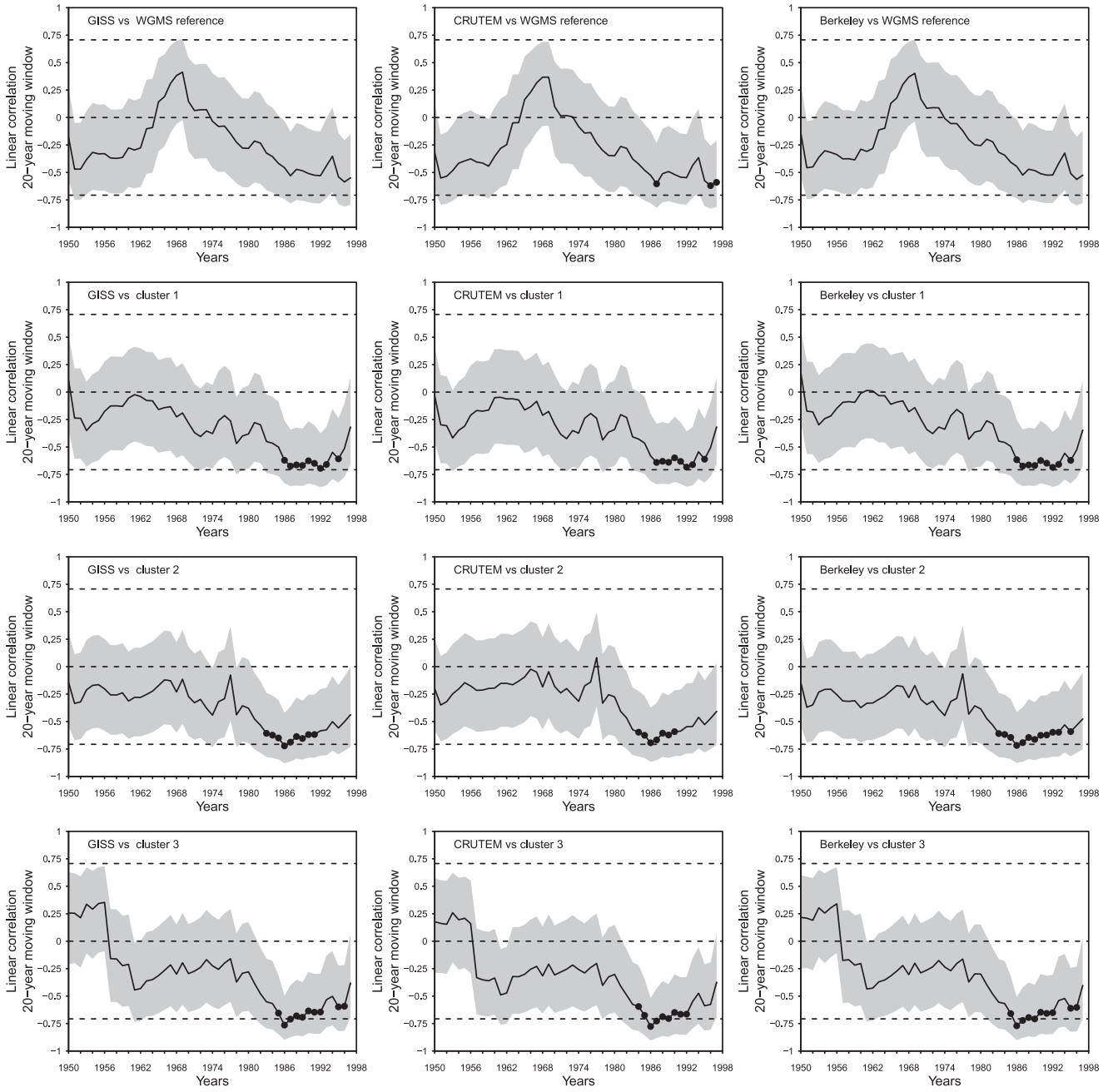



**Fig. 14.** Linear correlations, using 20-year moving windows, between the surface mass balance time series and the climatologies' global time series. Gray envelopes represent the 95% confidence level, segmented lines indicate the 0 and 50% of explained variance, and black dots indicate the starting year of the window when the linear correlation was significant at $p < 0.005$.

**Table 3.** Linear correlations and sensitivity of surface mass balance time series relative to global climatologies

| | WGMS reference glaciers | Average all clusters | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|---|---|
| Linear correlations | | | | | |
| GISS | − 0.57 | − 0.62 | − 0.53 | − 0.62 | − 0.25 |
| CRUTEM | − 0.61 | − 0.62 | − 0.54 | − 0.62 | − 0.28 |
| Berkeley | − 0.56 | − 0.62 | − 0.51 | − 0.62 | − 0.26 |
| Sensitivity (m $K^{-1}$) | | | | | |
| GISS | − 0.38 | − 0.41 | − 0.43 | − 0.49 | − 0.19 |
| CRUTEM | − 0.41 | − 0.41 | − 0.44 | − 0.50 | − 0.21 |
| Berkeley | − 0.38 | − 0.42 | − 0.53 | − 0.51 | − 0.20 |

balance time series across the world and an increasingly coherent trend, consistent with global shrinking detected from satellite imagery (Hugonnet and others, 2019), they also point to a persistent spatial bias toward a relatively narrow latitudinal band, where most glaciers of the Southern Hemisphere appear to be outliers in terms of their locations. Although we assess the spatial bias is much reduced than decades ago (Cogley and Adams, 1998), a reasonable inquiry about this result is whether that bias really matters for the analysis of climatic sensitivity. We contend that our regionalization approach suggests that this is partially irrelevant at the global scale and that following our strategy it becomes possible to reassess previous studies that considered this as a potential limiting factor to draw conclusions solely from surface mass balance data (Braithwaite and Hughes, 2020). The clustering, by

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

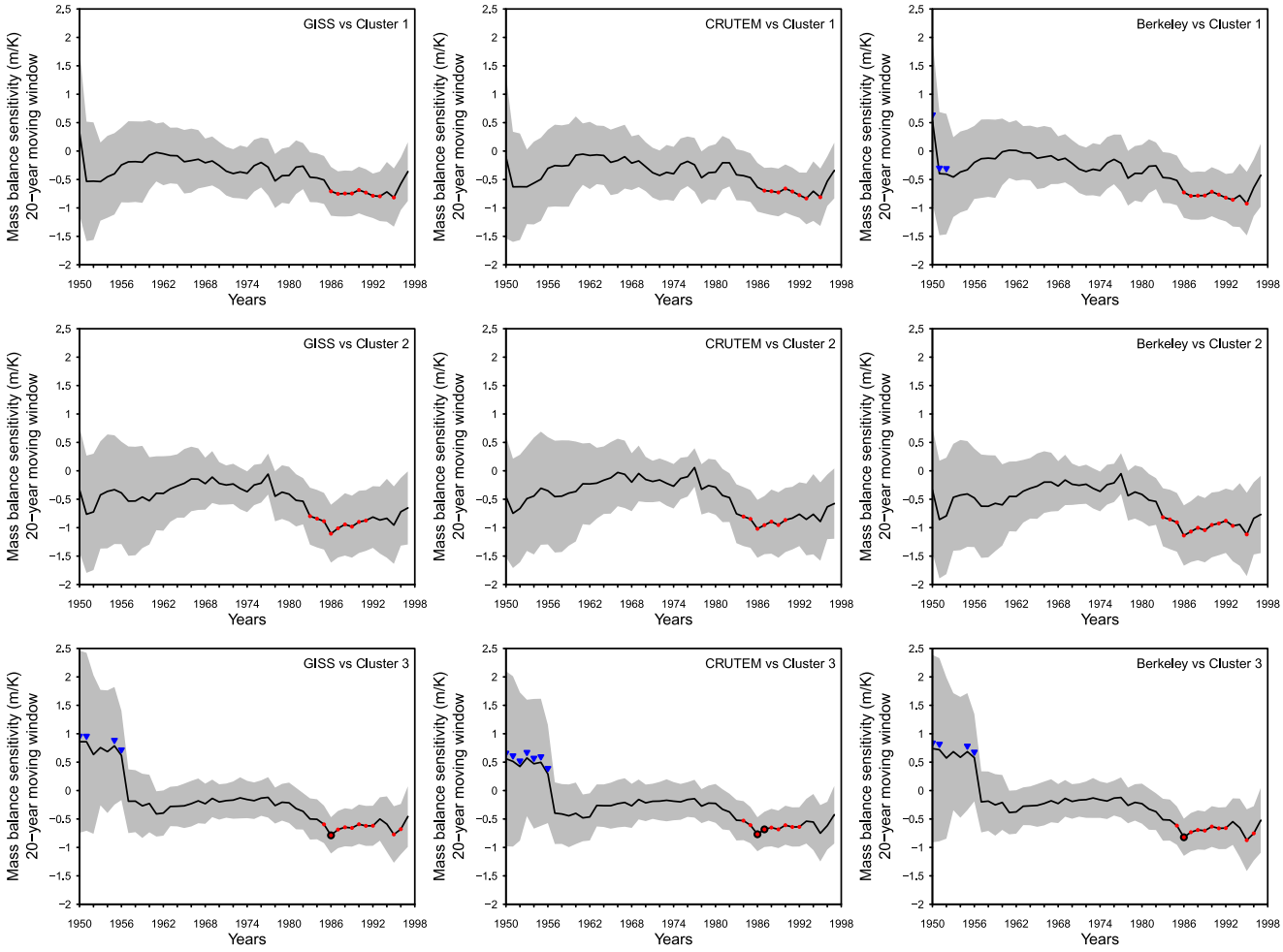



**Fig. 15.** Slopes of the linear models, using 20-year moving windows, between the surface mass balance time series and climatologies' global time series. Gray envelopes represent the 95% confidence interval, red dots correspond to slopes significant at $p < 0.005$, black circles indicate linear models that explain more that 50% of the variance, and blue triangles point to the initial year of the 20-year period in which the Durbin–Watson test finds serial correlation in the residuals.

delivering only three categories at this scale, reaffirms the understanding that glaciers located at long distances can share similar climatic response (Dyurgerov and Meier, 2000), suggesting that just a few glaciers of each group would be representative of climate–glacier relationships of large areas. Misfit glaciers within the identified clusters are very few and thus allow us to argue that at this scale inner variability matters less. Under a similar premise, Pelto (1990) used cluster analysis to classify glaciers located north from 35°N according to mass-balance gradients, identifying five groups. Furbish and Andrews' (1984) paper on glacier hypsometry demonstrated that glaciers within a similar climatic setting may respond differently to climate changes. We posit that our study provides a framework to identify large-scale glacier regimes where the climatic signal emerges despite internal geometric differences, as suggested by the low weight of length, area and hypsometry in the classification procedure. Certainly, a number of additional morphometric parameters can be computed and added to the database, but the fact that the ones used in this study do not correlate with the climatic variables suggest morphometry impacts within and not between regions. For instance, Manquehual-Cheuque and Somos-Valenzuela (2021) modeled climate change refugia for glaciers in Northern Patagonia, dividing the region of interest according to CECS (2009) subdivisions to improve model performance since local variables such as aspect and slope turned out as significant predictors differentiating glacier groups at this scale. In our work, the heavy weight of *Zmed*, *Zmi* and *Zma* indicates that the groups respond primarily to climatic conditions. This is demonstrated by clusters 1 and 2 that show respectively high shared variance since 1960 (above 50%, as expressed by the EPS), with narrow confidence limits in the average mass-balance curve. However, cluster 3 goes in a divergent path (declining EPS), suggesting that for this group the spatial bias still matters. The small and discontinuous sample of <6 mass-balance time series per 20-year period, mostly of Northern Hemisphere glaciers, may result in a low probability of averaging out local geometric characteristics and thus within this regime a regional signal might yet be difficult to extract. Conversely, this could also mean that some of these glaciers represent different regimes that cannot be separated under the present method. However, the high consistency between Antarctica and Greenland mass balances revealed in recent studies (Hanna and others, 2020) suggests the lack of data as the main cause.

Glacier–climate studies using regionalizations have normally used a larger number of groups than the three clusters identified in our work (e.g. Mernild and others, 2013). The classification in our study differs from the RGI regions (Pfeffer and others, 2014). As indicated in the introduction, during our literature review, we did not find explanation on the criteria used for defining these regions. Certainly, the goals of initiatives such as the WGI were to establish a geographic (i.e. glacierized mountain ranges) and country-based representation in order to obtain the most complete coverage of glaciers (WGMS, 1989, 2020). We assess our results add a new view on the data available by delivering a sort

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

of climate-based first-order classification of glaciers, as morphometric descriptors do not seem to be important at this scale. Meier (1984) determined the contribution of non-polar glaciers to sea level based upon 31 regions that do not resemble the locations of regime transitions determined in our work but instead agree with the view aimed at maximizing geographic coverage. An example is the 55°N boundary in North America; in our case, we found a limit between cluster 1 and 2 within the band 49°N to 45°N. That study also differentiated the Andes in three large areas separated at the equator and at 30°S, while in our work there is a transition between 34°S and 36°S only. These latitudes showing up as a significant glacier regime change in South America coincide with previous work on glacier classification. Using a more complete database than Sagredo and Lowell (2012), our study finds the separation between cluster 1 and 2 roughly matching the first branch of the objective function employed in that work that focused on South America, a limit previously proposed for Chilean glaciers too (CECS, 2009; Fernández, 2009). Therefore, our work seems to provide a global framework to support further regional disaggregations or fine-tuning existing proposals (e.g. Pelto, 1990; Caro and others, 2020). It is clear that the clusters seem to coincide with some of the common categorizations of glacier regimes (e.g. De Woul and Hock, 2005) such as continental (cluster 1) and maritime (cluster 2), while a polar category emerges as well (cluster 3). However, particularly with the split between clusters 1 and 2, we interpret that the notions of maritime and continental are mostly meaningful for the Northern Hemisphere as there is a clear correlation between longitude and descriptors of glacier elevation, indicating a non-negligible effect of the distance from the ocean. The high positive correlation of longitude relative to *Zma*, *Zme* and *Zmi* in the Northern Hemisphere reaffirms previous work on spatial distribution by Schytt (1967), who analyzed glaciers located north from 50°N finding that the more inland, the highest the elevation of glacier coverage. Notwithstanding in the high mountains of Asia, we only found glaciers of cluster 1, with the exception of less than ten glaciers around the Pamir and Altai mountains, and thus the distance to the Indian – to the south– or to the Atlantic – to the east – oceans seem to have no impact on the glacier regime at the global scale. However, these few glaciers classified in cluster 1 but located within a larger area of cluster 2 serve as an example of the usefulness of our approach as a tool to study long-distance similarities between glacier behavior and climate forcings. When looked at the global scale, these glaciers, according to our clustering method (see Fig. S15), pertain to a regime that is more sensitive to temperature, where *Tav*, *Tmax*, *Tmin* tend to be closer to the melting point relative to the other groups and with relatively high precipitation, although with spread (Fig. 9). Certain large-scale features in that region, especially the importance of westerlies' patterns, are also common in other regions that fall within cluster 2, such as the North America west coast, Patagonia and New Zealand (Sturman and Wanner, 2001; Kittel and others, 2002; Garreaud and others, 2013). As expected from the distance from the ocean in the case of these Asian glaciers of cluster 2, this influence should not be as widespread as in the other regions, but the clustering correctly shows that certain features that determine glacier regimes are comparable. According to our results, in the Southern Hemisphere, the longitudinal pattern does not exist when analyzed from a global scale perspective. On the contrary, the regime separation in the Southern Hemisphere, particularly in the Andes, is latitudinal in essence and seems more a reflection of the decreasing elevation of the Andes toward the south, and effect that influences climate and snow regimes too (Sarricolea and others, 2017; Saavedra and others, 2017). Therefore, we assess the spatial consistency proofs that the three clusters defined here constitute first-order spatial distribution of climate–glacier regimes and its transitions at the global scale.

### 4.2. The use of clusters for analyzing temperature sensitivity

Using the clustering results to compute mass-balance time series delivered slightly different global trends relative to using reference glaciers only. We interpret the result of the clustering from a climatic perspective in the sense that each surface mass balance cluster time series is representative of the respective category even if the composite is built from a handful of glaciers, thus the behavior after 1980 may correspond to a window important to focus on. Moreover, as indicated in the analysis of Figure 4, uncertainties tend to be lower after 1980 as the number of glaciers monitored and the represented latitudes increased. But, perhaps more interestingly, the analysis of clusters led us to experiment with studying surface mass balance sensitivity using linear regressions for the different regimes identified and characterized. Our approach to studying surface mass balance sensitivity can be deemed unorthodox as treatises hardly mention regression as a suitable technique for this kind of work. For example, Cuffey and Paterson (2010) suggest that calibrated mass-balance models that compute the annual cycle of accumulation and melt are a better approach because they can account for different processes contributing to the surface energy and mass balance more explicitly. Braithwaite and others (2002) compiled a list of studies on surface mass balance sensitivity that included eight employing regression models, with these applications using windows shorter that the 20 years we found relevant to the present work (e.g. Young, 1981). Regression resulted in figures that are comparable with previously published values using both static and dynamic approaches. The global mean computed from averaging clusters' time series for the period 1950 to the present indicates a surface mass balance sensitivity of $-0.37$ (GISS) and $-0.38\,\mathrm{m\,K^{-1}}$ (CRUTEM and Berkeley), closely matching previous results (Oerlemans and Fortuin, 1992; Dyurgerov and Meier, 2000; Raper and Braithwaite, 2006). Remarkably, each cluster presents a different mean sensitivity, with glaciers in cluster 2 appearing to be the most sensitive and those in cluster 3 the least. Cluster 2 contains glaciers in Patagonia and New Zealand, arguably the most sensitive in the world as detected by modeling (Mackintosh and others, 2017). On the other hand, glaciers in cluster 3 tend to be in cold and dry environments where sublimation becomes a significant phase change in the mass balance (both for ablation and accumulation), potentially leading to decoupling from temperature. Glaciers in cluster 1 are in a middle range of sensitivity, closer to global averages. Hugonnet and others (2019) determined a global temperature sensitivity $-0.27\,\mathrm{m\,K^{-1}}$ using an area-averaged method to compare mass-balance change (thickness change computed from a geodetic mass-balance method) and air temperature at the 700 geopotential level. This value, generally lower than ours and previous estimates, is nearly identical to the sensitivity of cluster 3 from our study and may indicate the large impact of these glaciers in the computation of sensitivity at the global scale. However, it is worth indicating that using our 20-year window we detect a global trend in sensitivity reduction since the late 1990s (Fig. 15), roughly corresponding to Hugonnet and others (2019) period of analysis.

Sensitivities for each cluster show interesting temporal patterns that we have not found previously reported in the literature. Using a 20-year window, we detected that regression slopes fluctuated quite significantly, indicating a maximum period of sensitivity for all clusters from the mid-1980s to mid-1990s. This period with statistically significant slopes coincides with previous findings on stronger negative trends in mass-balance and glacier contribution to sea level rise (Dyurgerov and Meier, 2000; Mernild

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

and others, 2013; Zemp and others, 2019), but to our knowledge, this has not been described for temperature sensitivity. Coherence in sensitivity and the surface mass balance trend since the mid-1980s provides an additional perspective on the timing of the homogenization of glacier fluctuations as a result of a sufficiently large global warming. Also, and as indicated above, the recent decrease in sensitivity (Fig. 15) agrees with latest findings (Hugonnet and others, 2019). Limitations of our results are associated with the fact that at this global scale we cannot yet account for the impact of particular features of the monitored glaciers that may affect mass-balance sensitivity. For instance, supraglacial debris may decouple them from regional climatic conditions (Nicholson and others, 2021; Rounce and others, 2021) and thus impact mass balance. Therefore, future refinement of our analysis may include merging the global glacier inventory with recently published databases on debris cover (Herreid and Pellicciotti, 2020) to then account for that variable in the clustering in order to reassess our classification.

The fluctuation in sensitivities as detected by the regression slopes for the studied period also implies a non-stationary glacier–climate relationship, with temperature in certain periods acting more strongly over the surface mass balance. The overlap among confidence intervals for the three clusters suggests that surface mass balance sensitivity spread within a specific range while behaving differently on average. Thus, our view is that these time series of sensitivity represent the aggregate conditions of a particular glacier regime, different of the other regimes determined in this study. Variations in temperature sensitivity have been previously described for glaciers in the Alps (Leonelli and others, 2011) and globally in the intercomparison of several glacier models, with the growth in sensitivity slowing down as temperature increases (Hock and others, 2019). Leonelli and others (2011) investigated the correlations between tree-ring chronologies, glacier mass-balance time series and several climatic time series extracted from gridded products, finding statistically significant correlations for the period 1952–2003 between surface mass balance and boreal summer temperatures; more interestingly, these authors found increasing negative correlations to about the year 2000 to then begin to become less negative. As glaciers shrink their hypsometry changes by losing volume in the lower sections, thus receding to higher elevations where variations in lapse rates and the zero isotherm may impact less the mass balance. Modeling work has shown that considering hypsometry and minimum elevation in glacier changes controls sensitivity (Marzeion and others, 2014). Although WGMS reference glaciers mass-balance time series are generally computed using the conventional balance method, i.e. accounting for the glacier area and hypsometry of the year measured, for the rest of the glaciers, it is more difficult to ascertain time series built with that method or the reference-surface method, that assumes a fixed area (Elsberg and others, 2001; Cogley and others, 2011). However, it has been demonstrated that this does not exert a significant impact on the response to climatic variability (Huss and others, 2012). Ablation areas are logically the glacier sections where most melt occurs (e.g. Azam and others, 2012; Marinsek and Ermolin, 2015) and thus even without hypsometric adjustment, their size relative to the whole glacier surface area will contribute to the strength of the correlation to temperature change because it determines the phase change from solid to liquid. Ultimately, a glacier that reduces its ablation area to a minimum may become closer to equilibrium to its climate (Pelto, 2010) or decouple from fluctuations in the regional climate (Colucci, 2016). The fact that glaciers have continued retreating worldwide allows us to infer that the consistent loss of statistical significance of the regression slope for all clusters might be indicating a logical trend to less sensitivity as they recede to higher elevations or irremediably shrink.

## 5. Conclusions

Our alternative approach to computation and analysis of surface mass balance sensitivity aimed to take full advantage of the richer data available. Although using machine learning methods to generate classifications is not new, we see this as the first study providing an unbiased way to revisit surface mass balance sensitivity informed by a global clustering method. While the clustering results can be improved by including other variables and testing other data-driven approaches, the RGI fields that resulted more important in this study suggest that just a few indicators are enough to characterize glacier regimes and response to climate at the global scale. The three clusters delivered by our unsupervised procedure reveal that at the global scale it is climate which largely determines the location and regime of glaciers (cf. Cuffey and Paterson, 2010: 5.4.5) and that morphometric descriptors may become important differentiating factors only within these groups. Our work can be used as a starting point to test other variables for finer disaggregation or to analyze whether a particular glacier represents the regime of a region. Our attempt employing our regionalization for merging all available surface mass balance time series tried to maximize the extraction of information by devising a method to integrate all available time series. The method thus assumes that glaciers of a similar regime behave uniformly so their differences in surface mass balance sensitivity in the long term are minimal. Despite that the magnitudes of global surface mass balance trends and temperature sensitivity agree with most previous research, further consideration of the impact of debris cover and other dynamics that influence glacier response to climate is needed, which requires expanding monitoring programs and inventory databases. However, the high degree of consistency between sensitivities and glacier regimes characterized with the clustering method suggests that our averaging method has been able to discriminate differences in the absolute response to temperature changes, expressed in m $K^{-1}$, while detecting similarities in the trend of sensitivity fluctuations. These findings point to a scenario where glaciers are showing a coherent signal of change no matter their regime. However, that the surface mass balance sensitivity resulted non-stationary, renders fundamental to account for negative feedbacks that may partially decouple glaciers from climatic changes at times. Therefore, we assess our work has revealed new insights into glacier–climate relationships that can guide the development of observational and analytical strategies alike.

**Supplementary material.** The supplementary material for this article can be found at https://doi.org/10.1017/jog.2022.16.

**Acknowledgements.** We are funded by the Chilean Science Council (ANID): FONDECYT Regular grant N°1201429 and Anillo ACT210080. We also acknowledge all the open databases developed by the WGMS. This work is also a tribute to the hundreds of groups gathering, curating and sharing mass-balance and inventory data for more than a century. We are grateful to Mario Pino for early discussions on this work, Bryan Mark for last-minute revisions of the language style, and to the editors and reviewers.

## References

**Akinduko AA, Mirkes EM and Gorban AN** (2016) SOM: stochastic initialization versus principal components. *Information Sciences* **364-365**, 213–221. doi: 10.1016/j.ins.2015.10.013.

**Azam MF and 10 others** (2012) From balance to imbalance: a shift in the dynamic behaviour of Chhota Shigri glacier, western Himalaya, India. *Journal of Glaciology* **58**(208), 315–324. doi: 10.3189/2012JoG11J123.

**Braithwaite RJ, Zhang Y and Raper SCB** (2002) Temperature sensitivity of the mass balance of mountain glaciers and ice caps as a climatological characteristic. *Zeitschrift für Gletscherkunde und Glazialgeologie* **38**(1), 35–61.

**Braithwaite RJ and Hughes PD** (2020) Regional geography of glacier mass balance variability over seven decades 1946–2015. *Frontiers in Earth Science* **8**, 1–14. doi: 10.3389/feart.2020.00302.

**Caloiero T** (2014) Analysis of daily rainfall concentration in New Zealand. *Natural Hazards* **72**(2), 389–404. doi: 10.1007/s11069-013-1015-1

**Caro A, Gimeno F, Rabatel A, Condom T and Ruiz J** (2020) Glacier clusters identification across Chilean Andes using topo-climatic variables. *Investigaciones Geográficas* **60**, 119–133. doi: 10.5354/0719-5370.2020.59009

**CECS** (2009) *Estrategia nacional de glaciares: fundamentos*. S.T.I. N°205, Dirección General de Aguas – Ministerio de Obras Públicas, Chile.

**Charrad M, Ghazzali N, Boiteau V and Niknafs A** (2014) NbClust: an R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software* **61**(6), 1–36. doi: 10.18637/jss.v061.i06.

**Ciracì E, Velicogna I and Swenson S** (2020) Continuity of the mass loss of the world's glaciers and ice caps from the GRACE and GRACE follow-on missions. *Geophysical Research Letters* **47**(9), 1–11. doi: 10.1029/2019GL086926.

**Cogley J and Adams W** (1998) Mass balance of glaciers other than the ice sheets. *Journal of Glaciology* **144**(147), 315–325.

**Cogley J and 10 others** (2011) Glossary of glacier mass balance and related terms. IHP-VII Technical Documents in Hydrology No. 86, IACS Contribution No. 2, UNESCO-IHP, Paris.

**Colucci RR** (2016) Geomorphic influence on small glacier response to post-Little Ice Age climate warming: Julian Alps, Europe. *Earth Surface Processes and Landforms* **41**(9), 1227–1240. doi: 10.1002/esp.3908.

**Cuffey KM and Paterson WSB** (2010) *The Physics of Glaciers*, 4th Edn. Burlington:Butterwoth-Heinemann.

**De Woul M and Hock R** (2005) Static mass-balance sensitivity of Arctic glaciers and ice caps using a degree-day approach. *Annals of Glaciology* **42**(1), 217–224. doi: 10.3189/172756405781813096.

**Demšar U, Harris P, Brunsdon C, Fotheringham AS and McLoone S** (2013) Principal component analysis on spatial data: an overview. *Annals of the Association of American Geographers* **103**(1), 106–128. doi: 10.1080/00045608.2012.689236.

**Dyurgerov MB** (2002) Glacier mass balance and regime: data of measurements and analysis. Technical report, Institute of Arctic and Alpine Research, University of Colorado.

**Dyurgerov MB and Meier MF** (2000) Twentieth century climate change: evidence from small glaciers. *Proceedings of the National Academy of Sciences of the United States of America* **97**(4), 1406–1411. doi: 10.1073/pnas.97.4.1406.

**Elsberg DH, Harrison WD, Echelmeyer KA and Krimmel RM** (2001) Quantifying the effects of climate and surface change on glacier mass balance. *Journal of Glaciology* **47**(159), 649–658. doi: 10.3189/172756501781831783.

**Etzelmüller B** (2000) On the quantification of surface changes using grid-based digital elevation models (DEMs). *Transactions in GIS* **4**(2), 129–143. doi: 10.1111/1467-9671.00043.

**Fernández A** (2009) *Patrones y controles en la distribución de los glaciares de los Andes chilenos: la memoria del paisaje*. Unpublished MSc. Thesis, Universidad Austral de Chile, Valdivia.

**Fernández A, Araos J and Marín J** (2010) Inventory and geometrical changes in small glaciers covering three Northern Patagonian summits using remote sensing and GIS techniques. *Journal of Mountain Science* **7**(1), 26–35. doi: 10.1007/s11629-010-1066-7.

**Fick SE and Hijmans RJ** (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**(12), 4302–4315. doi: 10.1002/joc.5086.

**Furbish D and Andrews J** (1984) The use of hypsometry to indicate long-term stability and response of valley glaciers to changes in mass transfer. *Journal of Glaciology* **30**(105), 199–211.

**Garreaud R, Lopez P, Minvielle M and Rojas M** (2013) Large-scale control on the Patagonian climate. *Journal of Climate* **26**, 215–230.

**Gregory JM and Oerlemans J** (1998) Simulated future sea-level rise due to glacier melt based on regionally and seasonally resolved temperature changes. *Nature* **391**(1), 474–477.

**Hanna E and 10 others** (2020) Mass balance of the ice sheets and glaciers – progress since AR5 and challenges. *Earth-Science Reviews* **201**(10). doi: 10.1016/j.earscirev.2019.102976.

**Herreid S and Pellicciotti F** (2020) The state of rock debris covering Earth's glaciers. *Nature Geoscience* **9** (13) 621–627. doi: 10.1038/s41561-020-0615-0.

**Hock R and 7 others** (2019) GlacierMIP-A model intercomparison of global-scale glacier mass-balance models and projections. *Journal of Glaciology* **65**(251), 453–467. doi: 10.1017/jog.2019.22.

**Hugonnet R and 10 others** (2019) Accelerated global glacier mass loss in the early twenty-first century. *Nature* **592**(7856), 726–731. doi: 10.1038/s41586-021-03436-z.

**Huss M, Hock R, Bauder A and Funk M** (2012) Conventional versus reference-surface mass balance. *Journal of Glaciology* **58**(208), 278–286. doi: 10.3189/2012JoG11J216.

**Kittel T, Thornton PE, Royle AJ and Chase TN** (2002) Climates of the Rocky Mountains: historical and future patterns. In *Rocky Mountain Futures: An Ecological Perspective*. Covelo, CA: Island Press (doi: 10.13140/RG.2.1.1683.0487).

**Kohonen T** (1990) The self-organizing map. *Proceedings of the IEEE* **78**(9), 1464–1480(doi: 10.1109/5.58325).

**Lenssen NJ and 6 others** (2019) Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres* **124**(12), 6307–6326. doi: 10.1029/2018JD029522.

**Leonelli G, Pelfini M, D'Arrigo R, Haeberli W and Cherubini P** (2011) Non-stationary responses of tree-ring chronologies and glacier mass balance to climate in the European Alps. *Arctic, Antarctic, and Alpine Research* **43**(1), 56–65. doi: 10.1657/1938-4246-43.1.56.

**Mackintosh AN, Anderson BM and Pierrehumbert RT** (2017) Reconstructing climate from glaciers. *Annual Review of Earth and Planetary Sciences* **45**(1), 649–680. doi: 10.1146/annurev-earth-063016-020643.

**Manquehual-Cheuque F and Somos-Valenzuela M** (2021) Climate change refugia for glaciers in Patagonia. *Anthropocene* **33**, 100277. doi: 10.1016/j.ancene.2020.100277.

**Marinsek S and Ermolin E** (2015) 10 Year mass balance by glaciological and geodetic methods of Glaciar Bahía del Diablo, Vega Island, Antarctic Peninsula. *Annals of Glaciology* **56**(70), 141–146. doi: 10.3189/2015AoG70A958.

**Mark BG and Fernández A** (2017) The significance of mountain glaciers as sentinels of climate and environmental change. *Geography Compass* **11** (6), e12318. doi: 10.1111/GEC3.12318.

**Markonis Y and Strnad F** (2020) Representation of European hydroclimatic patterns with self-organizing maps. *The Holocene* **30**(8), 1155–1162. doi: 10.1177/0959683620913924.

**Marzeion B, Jarosch AH and Hofer M** (2012) Past and future sea-level change from the surface mass balance of glaciers. *The Cryosphere* **6**(6), 1295–1322. doi: 10.5194/tc-6-1295-2012.

**Marzeion B, Jarosch AH and Gregory JM** (2014) Feedbacks and mechanisms affecting the global sensitivity of glaciers to climate change. *The Cryosphere* **8**(1), 59–71. doi: 10.5194/tc-8-59-2014.

**Meehl GA and 13 others** (2007) Global climate projections. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge and New York: Cambridge University Press.

**Meier MF** (1984) Contribution of small glaciers to global sea level. *Science* **226** (4681), 1418–1421. doi: 10.1126/science.226.4681.1418

**Mernild SH, Lipscomb WH, Bahr DB, Radić V and Zemp M** (2013) Global glacier changes: a revised assessment of committed mass losses and sampling uncertainties. *The Cryosphere* **7**(5), 1565–1577. doi: 10.5194/tc-7-1565-2013.

**Nicholson L, Wribel A, Mayer C and Lambrecht A** (2021) The challenge of non-stationary feedbacks in modeling the response of debris-covered glaciers to climate forcing. *Frontiers in Earth Science* **9**, 1–18. doi: 10.3389/feart.2021.66269.

**Nussbaumer SU and 5 others** (2017) Glacier monitoring and capacity building: important ingredients for sustainable mountain development. *Mountain Research and Development* **37**(1), 141–152. doi: 10.1659/MRD-JOURNAL-D-15-00038.1.

**Oerlemans J and Fortuin JP** (1992) Sensitivity of glaciers and small ice caps to greenhouse warming. *Science* **258**(5079), 115–117. doi: 10.1126/science.258.5079.115.

**Osborn TJ and Jones PD** (2014) The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google earth. *Earth System Science Data* **6**(1), 61–68. doi: 10.5194/essd-6-61-2014.

**Pelto MS** (2010) Forecasting temperate alpine glacier survival from accumulation zone observations. *The Cryosphere* **4**(1), 67–75. doi: 10.5194/tc-4-67-2010.

**Pelto MS** (1990) Modeling mass-balance changes during a glaciation cycle. *Annals of Glaciology* **14**, 238–241. doi: 10.3189/S0260305500008661.

**Permana DS and 20 others** (2019) Disappearance of the last tropical glaciers in the Western Pacific Warm Pool (Papua, Indonesia) appears imminent. *Proceedings of the National Academy of Sciences* **116**(52), 26382–26388. doi: 10.1073/pnas.1822037116.

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press

**Pfeffer WT and 18 others** (2014) The Randolph Glacier Inventory: a globally complete inventory of glaciers. *Journal of Glaciology* **60**(221), 537–552. doi: 10.3189/2014JoG13J176.

**Pike RJ and Wilson SE** (1971) Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. *Bulletin of the Geological Society of America* **82**(4), 1079–1084. doi: 10.1130/0016-7606(1971)82[1079:ERHIAG]2.0.CO;2.

**Radić V and Hock R** (2010) Regional and global volumes of glaciers derived from statistical upscaling of glacier inventory data. *Journal of Geophysical Research: Earth Surface* **115**(1), 1–10. doi: 10.1029/2009JF001373.

**Raper SCB and Braithwaite RJ** (2006) Low sea level rise projections from mountain glaciers and icecaps under global warming. *Nature* **439**(7074), 311–313. doi: 10.1038/nature04448.

**Reusch DB, Alley RB and Hewitson BC** (2005) Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data. *Polar Geography* **29**(3), 188–212. doi: 10.1080/789610199.

**RGI Consortium** (2017). Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0: Technical Report, Global Land Ice Measurements from Space, Colorado, USA. Digital Media. doi: 10.7265/N5-RGI-60

**Roe GH, Baker MB and Herla F** (2017) Centennial glacier retreat as categorical evidence of regional climate change. *Nature Geoscience* **10**, 95–99. doi: 110.1038/ngeo2863.

**Rohde R, Muller R, Jacobsen R, Perlmutter S and Mosher S** (2013) Berkeley earth temperature averaging process. *Geoinformatics & Geostatistics: An Overview* **1**(2), 1–13. doi: 10.4172/2327-4581.1000103.

**Rounce DR and 10 others** (2021) Distributed global debris thickness estimates reveal debris significantly impacts glacier mass balance. *Geophysical Research Letters* **8**(48). doi: 10.1029/2020GL091311.

**Sagredo EA and Lowell TV** (2012) Climatology of Andean glaciers: a framework to understand glacier response to climate change. *Global and Planetary Change* **86-87**, 101–109. doi: 10.1016/j.gloplacha.2012.02.010.

**Sarricolea P, Herrera-Ossandon M and Meseguer-Ruiz O** (2017) Climatic regionalisation of continental Chile. *Journal of Maps* **13**(2), 66–73. doi: 10.1080/17445647.2016.1259592.

**Saavedra FA, Kampf SK, Fassnacht SR and Sibold JS** (2017) A snow climatology of the Andes Mountains from MODIS snow cover data. *International Journal of Climatology* **37**(3), 1526–1539. doi: 10.1002/joc.4795.

**Schaefer M, Fonseca-Gallardo D, Fariás-Barahona D and Casassa G** (2020) Surface energy fluxes on Chilean glaciers: measurements and models. *The Cryosphere* **14**(8), 2545–2565. doi: 10.5194/tc-14-2545-2020.

**Schumacher V, Fernández A, Justino F and Comin A** (2020) WRF high resolution dynamical downscaling of precipitation for the Central Andes of Chile and Argentina. *Frontiers in Earth Science* **8**, 328. doi: 10.3389/feart.2020.00328.

**Schytt V** (1967) A study of ablation gradient. *Geografiska Annaler. Series A, Physical Geography* **49**(2/4), 327–332. doi: 10.2307/520899.

**Sturman A and Wanner H** (2001) A comparative review of the weather and climate of the Southern Alps of New Zealand and the European Alps. *Mountain Research and Development* **21**(4), 359-369.

**Vesanto J** (2000) Neural network tool for data mining: SOM toolbox. *Proceedings of symposium on tool environments and development methods for intelligent systems*, 184–196.

**Viale M and Nuñez MN** (2011) Climatology of winter orographic precipitation over the subtropical central Andes and associated synoptic and regional characteristics. *Journal of Hydrometeorology* **12**(4), 481–507. doi: 10.1175/2010JHM1284.1.

**Ward JH** (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**(301), 236–244. doi: 10.1080/01621459.1963.10500845.

**WGMS** (1989). World Glacier Inventory – Status 1988 IAHS (ICSI) / UNEP / UNESCO, World Glacier Monitoring Service.

**WGMS** (2020). Global Glacier Change Bulletin No. 3 IAHS (ICSI) / UNEP / UNESCO, World Glacier Monitoring Service.

**Wigley TML, Briffa KR and Jones PD** (1984) On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. *Journal of Climate and Applied Meteorology* **23**(2), 201–213. doi: 10.1175/1520-0450(1984)023<0201:OTAVOC>2.0.CO;2.

**Young GJ** (1981) The mass balance of Peyto glacier, Alberta, Canada, 1965 to 1978. *Arctic and Alpine Research* **13**(3), 307–318. doi: 10.2307/1551037.

**Zemp M and 38 others** (2015) Historically unprecedented global glacier decline in the early 21st century. *Journal of Glaciology* **61**(228), 745–762. doi: 10.3189/2015JoG15J017.

**Zemp M and 14 others** (2019) Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**(7752), 382–386. doi: 10.1038/s41586-019-1071-0.

https://doi.org/10.1017/jog.2022.16 Published online by Cambridge University Press