nature portfolio

https://doi.org/10.1038/s41586-021-03436-z

Supplementary information

Accelerated global glacier mass loss in the early twenty-first century

In the format provided by the authors and unedited

Supplementary Information for

Accelerated global glacier mass loss in the early twenty-first century

Romain Hugonnet, Robert McNabb, Etienne Berthier, Brian Menounos, Christopher Nuth,
Luc Girod, Daniel Farinotti, Matthias Huss, Ines Dussaillant, Fanny Brun and Andreas Kääb

Correspondence to: romain.hugonnet@gmail.com

This document includes (navigation links embedded):

Supplementary Methods

Supplementary Discussion

Supplementary Figures

Supplementary Tables

Supplementary Methods

1. ASTER processing

We downloaded 440,548 ASTER L1A[43] granules totalling around 30TB of data from NASA's Land Processes Distributed Active Archive Center via EarthData Search (https://search.earthdata.nasa.gov). Daytime granules with a cloud coverage below 99% were selected, from the first acquisition date of 4th March 2000 until 30th September 2019. Spatially, they cover all terrain in 1 by 1 degree tiles containing glaciers. Our tiling was extended both in latitude and longitude to retrieve at least one additional granule in each direction, based on a 60 km by 60 km granule footprint and the along-track angle of the Terra satellite (boarding ASTER) at a given latitude.

This coverage extension increased the amount of stable terrain available in each DEM in order to improve the along-track corrections and co-registration (later described) and mitigate correction errors at the edges. We bypassed the arbitrary 60 km by 60 km splitting of the ASTER archive and combined granules containing a valid 3N and 3B Visible and Near Infra-Red (VNIR) band in groups of at most three, depending on whether they were acquired sequentially (less than 12s apart), thereby forming 60 km by 180 km image strips. This "stitching" was done in image geometry, before stereo matching. We chose a maximum of three granules in order to avoid effects from the curvature of the Earth appearing in our corrections. We also observed that three granules were sufficient to improve corrections for the frequencies typically observed in ASTER along-track undulations biases[50]. For occurences with less than three consecutives granules, we stitched two granules if possible or kept only one. We left one overlapping granule in between sequential strips to mitigate the edge effects of later corrections. The resulting ASTER L1A strips were processed in UTM zones based on their new centroid. To generate ASTER DEMs, we used MicMac ASTER (MMASTER)[50], a procedure that is part of the MicMac photogrammetric processing library[49]. In total, we generated 198,339 ASTER DEM strips posted at 30 m resolution (Table S1) which required about 5 million compute hours.

Our last step was to correct ASTER DEM strips for systematic biases using TanDEM-X as a reference. In this last step, prior to any correction, we performed an initial co-registration. We first removed cross-track biases by selecting the best-performing polynomial fit within orders 1 to 6 based on their RMSE (Fig. S1a). We then corrected along-track low-frequency and high-frequency undulations simultaneously using a sum of sinusoids with specific frequencies and amplitudes (Fig. S1b), conditioned by priors and optimized through basin-hopping[51]. Finally, ice-free terrain from the DEMs was co-registered a second time off-ice to TanDEM-X. We successfully corrected and co-registered 154,565 ASTER DEM strips (Extended Data Fig. 2, Table S1).

2. Elevation time series

The ASTER elevations are annotated as $h_{AST}(t,x,y)$. The ArcticDEM and REMA elevations, originating mainly from the WorldView satellite sensors, are annotated as $h_{WV}(t,x,y)$. The TanDEM-X reference elevations are annotated as $h_{TDX}(x,y)$.

2.1. Time stacking

Following co-registration, we created three-dimensional arrays (time t, space x and y; hereafter referred to as “stacks”) of elevation using all available ASTER DEMs, ArcticDEM DEMs and REMA DEMs within a 1 by 1 degree tiling. For ASTER DEMs generated with MicMac, a raster indicating the quality of stereo-correlation was also stacked in time for each corresponding DEM. We did not filter low-correlation pixels and relied solely on the statistical filtering and measurement error assessment later described. Only pixels (x,y) within a buffer of 10 km of the glacier inventory were kept for further processing. They together represent a surface of about 4,000,000 km² including about 700,000 km² of glacierized terrain[85].

The stacked elevations from ASTER $h_{AST}(t,x,y)$ and ArcticDEM/REMA $h_{WV}(t,x,y)$ are annotated as $h_{AST/WV}(t,x,y)$. The stacked quality of stereo-correlation from ASTER are annotated as $q(t,x,y)$ and varies from 0 to 100%.

2.2. Elevation filtering step 1: reference elevations

A large number of outliers is present in elevation data due to photogrammetric blunders, presence of clouds or low image contrast. We performed an initial filtering of stacked elevations to remove extreme outliers using our reference DEM TanDEM-X (Extended Data Fig. 3c).

First, we implemented a spatial filter. We excluded elevation observations $h_{AST/WV}$ for which the absolute elevation difference to the maximum or minimum reference elevation h_{TDX} found within a disk D of radius r was larger than a vertical elevation threshold Δh_D:

$$min(h_{TDX}(x_D,y_D)) - \Delta h_D < h_{AST/WV}(t,x,y) < max(h_{TDX}(x_D,y_D)) + \Delta h_D \qquad \text{(S1)}$$

where (x_D,y_D) are pixels in the disk D of radius r centered on (x,y).

This procedure was performed for each pixel with $r = 200\,m$, $\Delta h_D = 700\,m$ and repeated with $r = 500\,m$, $\Delta h_D = 500\,m$ and $r = 1000\,m$, $\Delta h_D = 300\,m$. These large threshold values and the ones detailed hereafter were defined and tested over the glacier HPS12, Southern Patagonian Icefield, experiencing to our knowledge the most rapid sustained elevation change in the world[53].

Then, we applied a temporal filter to all pixels (x,y) with a valid reference value $h_{TDX}(x,y)$ (Extended Data Fig. 3c). To constrain this filtering, we assumed the time stamp of TanDEM-X, t_{TDX}, to be 1st January 2013 which corresponds to the middle of the period used to generate the TanDEM-X mosaic DEM[46]. To account for possible time differences, we included all values within a vertical threshold Δh_0 of 100 m starting 1st January 2013. From this date, we propagated the filtering in time allowing for a maximum and minimum linear elevation change rate $\frac{\Delta h}{\Delta t}_T$ of 50 meters per year (glacier HPS12). The vertical threshold ensures the conservation of observations of rapidly evolving surfaces, for example that of surging glaciers, near the reference date t_{TDX}.

$$h_{AST/WV}(t,x,y) < \left| h_{TDX}(x,y) + \Delta h_0 + (t - t_{TDX}) \cdot \frac{\Delta h}{\Delta t}_T \right| \tag{S2}$$

2.3. Elevation measurement error

Inspection of our elevation observations $h_{AST/WV}$ revealed a significant variability in elevation precision (heteroscedasticity). To account for it in subsequent analysis, we evaluated the dependence of the elevation variance $\sigma_{h_{AST/WV}}{}^2$ on two factors: terrain slope $\alpha(x,y)$ and quality of stereo correlation $q(t,x,y)$. To assess this dependency, we randomly drew without replacement up to 10,000 off-ice elevation differences to the reference TanDEM-X, $h_{AST/WV} - h_{TDX}$, for varying categories of terrain slope and quality of stereo-correlation. We did this for all tiles globally. For each category, we estimated the variance as the square of the Normalized Median Absolute Deviation (NMAD) to mitigate the effect of elevation outliers[58].

Elevation measurement error $\sigma_{h_{AST/WV}}$ increased with terrain slope and, for ASTER DEMs, decreased with quality of stereo-correlation (Extended Data Fig. 3a). The dependency of error on surface slope is well documented[55,56] while the one to the quality of stereo-correlation is inherent to photogrammetry[49] but, despite this, both are rarely accounted for. As an example, for a typical ASTER DEM with a co-registration RMSE of 5 m, we found the NMAD to be about 3 m over low slopes (0 to 10 degrees) and exceeding 20 m for steep slopes (> 40 degrees). Similarly, we found the NMAD of elevation differences for $q = 100\%$ to be around 3 m, while it was over 30 m for $q = 40\%$.

The elevation measurement error $\sigma_{h_{AST/WV}}$ was estimated by a simple model, calibrated on the empirical variance (Extended Data Fig. 3a), to yield an error for all elevation observations in space (x,y) and time t (Equation 1, repeated below).

$$\sigma_{h_{AST/WV}}{}^2(t,x,y) = \sigma_c{}^2(t,x,y) + \sigma_\alpha{}^2(\alpha,q) + \sigma_q{}^2(q) \tag{1}$$

where $\alpha(x,y)$ is the terrain slope, $q(t,x,y)$ is the quality of stereo-correlation and $\sigma_c(t,x,y)$ is the co-registration error, specific to the DEM from which the elevation observation $h_{AST/WV}(t,x,y)$ originates. The co-registration error σ_c would ideally have to be estimated on pixels with low slopes and good qualities of stereo-correlation to avoid double-counting the effect of other errors. However, as this is not possible for some DEMs because of the limited amount of flat terrain available, we conservatively used the RMSE of elevation differences over all available stable terrain to derive σ_c.

The errors due to slope σ_α and quality of stereo-correlation σ_q were described as:

$$\sigma_\alpha = (a_\alpha + a_q(1 - q)^{b_\alpha}) \cdot \tan(\alpha) \tag{S3}$$

$$\sigma_q = a_q \cdot (1 - q)^{b_q} \tag{S4}$$

where we found $a_\alpha \approx 20$ m, $b_\alpha \approx 1$, $a_q \approx 20$ m and $b_q \approx 1.25$ by manually combining independent least squares optimizations for the slope and the quality of stereo-correlation on the empirical variance (Extended Data Fig. 3a). Against the above-mentioned approximations, the simplicity of this process was sufficient to calibrate the elevation variance at the right order of precision. The slope and quality of stereo-correlation errors were not independent, as it proved necessary to integrate some effects of the quality of stereo-correlation in the slope error to yield satisfying results (Equation S3). ArcticDEM and REMA DEMs being already filtered for low stereo-correlations without any product on quality of stereo-correlation provided[44,45,52], we assigned them a constant worst-case quality of stereo-correlation of 60%.

Over vegetated terrain, the variance in elevation increased by a factor of 1.5-2. We interpreted this high variance from vegetation that can dynamically change over time due to forest harvesting, wildfire, regrowth and snow-covered vegetation[60]. This does not significantly affect elevation measurement error over glacierized terrain and thus was omitted from our workflow. We did, however, verify its limited impact in subsequent analysis of elevation time series (Section 3).

2.4. Elevation filtering step 2: linear elevation change and reference elevations

In most regions of the world, the largest glacier elevation change rates are far lower than the extreme rate $\frac{\Delta h}{\Delta t}_T$ of 50 meters per year previously adopted. In order to improve the performance of our filtering (Extended Data Fig. 3c), we estimated a maximum and minimum acceptable linear elevation change rate at the pixel-scale.

First, we estimated a robust linear elevation change rate $\frac{dh}{dt}_{WLS}(x,y)$. For each pixel, we performed two successive weighted least-squares (WLS) fits, filtering outliers outside the 99 percent confidence interval of the first fit, and keeping only the second fit[54]. The elevation measurement error $\sigma_{h_{AST/WV}}$ (see above) was used for weighting.

Then, for each pixel, we derived the 20th and 80th percentiles of linear elevation change rate within a disk. We computed the maximum absolute value of the two percentiles and conservatively used twice this value to constrain the maximum linear elevation change rate allowed at the pixel-scale.

$$\frac{\Delta h}{\Delta t}_T(x,y) = 2 \cdot max\left(\left|\frac{dh}{dt}_{WLS}(x_D,y_D)_{20th}\right|, \left|\frac{dh}{dt}_{WLS}(x_D,y_D)_{80th}\right|\right) \tag{S5}$$

where (x_D, y_D) are pixels in the disk D of radius $r = 1000$ m centered on (x,y) and the subscripts 20th and 80th denote the 20th and 80th distribution percentiles.

We refined the temporal filter previously described (Equation S2, Extended Data Fig. 3c) with the linear elevation change rate $\frac{\Delta h}{\Delta t}_T(x,y)$ (instead of a constant of 50 meters per year). We also modified the vertical threshold Δh_0 to better account for the varying timestamp of TanDEM-X:

$$\Delta h_0(x,y) = \Delta h_0 + \Delta t_{TDX-TDX} \cdot \frac{\Delta h}{\Delta t}_T(x,y) \qquad \text{(S6)}$$

where $\Delta t_{TDX-TDX} = 2$ yr is the maximum number of years that can separate a TanDEM-X acquisition and the middle date t_{TDX} chosen as 1st January 2013.

We refined the spatial filter previously described (Section 2.2, Equation S1, Extended Data Fig. 3c) by further constraining the vertical elevation threshold Δh_D:

$$\Delta h_D(x,y) = \Delta h_0 + \Delta t_{AST-TDX} \cdot \frac{\Delta h}{\Delta t}_T(x,y) \qquad \text{(S7)}$$

with $\Delta h_0 = 100$ m, and $\Delta t_{AST-TDX} = 15$ yr is the maximum number of years that can separate a TanDEM-X acquisition in 2015 to the furthest observation in 2000. We applied the spatial filter with $r = 1000$ m. This step was especially helpful to filter remaining outliers for pixels with no valid reference elevation h_{TDX} where the temporal filter cannot be applied.

2.5. Temporal covariance of glacier elevation

Extracting a continuous time series from glacier elevation data is complex[61,62,76]. At the pixel scale, glacier surface elevation undergoes changes at different time scales (seasonal, annual, decadal) with trends and amplitudes that significantly vary in space (region, glacier, zone of the glacier). Parametric methods such as least squares are thus not particularly well-suited to fitting a temporal series to glacier elevation data. If an underlying parametrization is erroneous, both the temporal interpolation and the propagation of uncertainties will be negatively affected.

We instead chose to interpolate our elevation measurements using non-parametric, empirically-based interpolation methods based on the covariance of the data, a technique referred to as Gaussian Processes[59]. These methods find the local minima of variance propagated from observations. Assuming that the covariance model chosen is statistically representative of the underlying process, it provides the best unbiased estimator. These methods also yield empirical confidence intervals, having the benefit of being representative of the uneven temporal sampling inherent to the data.

We harnessed the repeat temporal coverage of DEMs to study the temporal covariance of glacier elevation change in order to better constrain both our temporal filtering and interpolation. Once the temporal covariance is estimated, we derive our best interpolator using Gaussian Process (GP) regression[63]. For this type of application, GP regression is equivalent to kriging[71,78,79]. Here, we do not optimize covariances for each pixel based on priors with a maximum likelihood function, as is usually the case in machine learning applications. Instead, our objective is to model variograms with characteristics representative of many pixels at once, and to apply these variograms directly in the regression. The rationale behind this approach is to mitigate the sparse sampling of elevations in time at the pixel scale by utilizing the repeat spatial coverage of our observations.

To derive temporal variograms, we randomly drew up to 10,000 pixels (x,y) of locations $h_{AST/WV}(t,x,y)$ containing at least 10 valid observations in time. We did so for all RGI regions. To reduce the effect of outliers, the variograms were estimated by the median of squared residuals for each time lag instead of the mean.

We aggregated our variograms depending on the linear trend of elevation of pixels $\frac{dh}{dt}_{WLS}(t,x,y)$ (Section 2.4) to better identify variance components independent of this underlying linear elevation change. The empirical temporal variances varied little between regions, and we found no significant variability with external factors (such as slope), as we did for the elevation measurement error.

Through analysis of the empirical variograms, we found that the data temporal covariance (Extended Data Fig. 3b) consisted of a sum of:

- a pairwise linear (PL) kernel, manifesting parabolically in variance, that represents the long-term, decadal linear elevation trend of the pixel (over 20 years in our case)
- a periodic, exponential sine-squared (ESS) kernel, corresponding to how seasonality is captured by the elevation data, for example by the fact that summer-to-summer observations of glacier elevations can be much closer than summer-to-winter observations
- a local, radial basis-function (RBF) kernel, showing how close elevation observations are to each other with varying time differences
- white noise, representative the average of the measurement errors $\sigma_{h_{AST/WV}}$.

To capture nonlinear elevation change trends (with time scales larger than the local RBF kernel), we added a local linear kernel, described by a rational quadratic (RQ) kernel multiplied by the pairwise linear kernel. We chose to model local linear changes using a RQ kernel times PL kernel instead of solely a RQ kernel. The latter was discarded after initial testing due to undesired effects at the temporal boundaries (2000 and 2020). The RQ times PL implies that the "mean" linear trend can vary locally and, when no observation is available at the boundaries, the extrapolated trend falls back towards the local "mean" linear trend. Physically, this is behaviour consistent with existing observations and with known decadal and sub-decadal climatic oscillations that influence glacier change. In practice, there is little extrapolation made in our study due to the dense repeat data coverage (Fig. 1).

We used our measurement error $\sigma_{h_{AST/WV}}$, having a value specific to each elevation observation in space and time, instead of the average measurement error (white noise) sampled by the empirical variograms. We semi-automatedly modelled the temporal covariance with time scales and amplitudes estimated from the empirical covariance (Extended Data Fig. 3b) and used the same ESS, RBF and RQ parameters for all pixels given the absence of significant nonstationarities. The pairwise linear kernel was estimated independently for each pixel (x,y) (Equation 2, repeated below):

$$\sigma_{h_{AST/WV}}(x,y,\Delta t)^2 = PL(x,y,\Delta t) + ESS(\phi_p, \sigma_p{}^2, \Delta t) + RBF(\Delta t_l, \sigma_l{}^2, \Delta t) + RQ(\Delta t_{nl}, \sigma_{nl}{}^2, \alpha_{nl}, \Delta t) \cdot PL(x,y,\Delta t) + \sigma_{h_{AST/WV}}(t,x,y)^2 \tag{2}$$

where Δt is the time lag between observations.

We found $\phi_p = 1$ yr, $\sigma_p \approx 5$ m, implying a seasonal periodicity component of 5 m on average. We found that the local signal was best decomposed into a sum of three RBF kernels with $\Delta t_{l1} \approx 0.75$ yr, $\sigma_{l1} \approx 5$ m, $\Delta t_{l2} \approx 1.5$ yr, $\sigma_{l2} \approx 4$ m and $\Delta t_{l3} \approx 3$ yr, $\sigma_{l3} \approx 2$ m, which suggests that, once the underlying linear trend and periodicity is removed, inter-annual glacier elevations are on average within 5 m of each other within a year, within 9 m within 1.5 year and within 11 m within 3 years. Finally, based on pixel-scale testing (for filtering purposes) and the temporal range of the underlying linear trend observed in our empirical variograms, we constrained the local linear values to $\sigma_{nl} \approx 10$ m, $\alpha_{nl} \approx 10$ and $\frac{1}{2\alpha_{nl}} \cdot \Delta t_{nl} \approx 5$ yr. Those values mean that, on average, local nonlinearity lasts around 5 years and within 10 m of the underlying linear trend. Our primary objective was to ensure a low sensitivity to outliers, which were not effectively filtered out when using shorter-time scale parameters. In order to avoid removing glacier surges, we included a conditional loop in our procedure, which was calibrated on Nathorstbreen glacier, Svalbard (the largest surge observed during our period of study, Fig. S3).

The variances described above do not directly condition the mean of the GP elevation time series, which is interpolated from available observations, but only leave the opportunity to find periodicity and local variations in those observations within an order of magnitude. For example, the periodic kernel (with $\sigma_p \approx 5$ m) applied to elevation observations in the Low Latitudes yields elevation temporal series with no marked seasonality, despite leaving the opportunity to find amplitudes of the order of 5 m. This is because, in this region, the elevation observations categorized by seasons do not show a significant seasonal trend (in front of linear, local trends) and thus the periodic kernel does not find significant seasonality to propagate from observations.

With the same rationale, sensitivity tests showed limited influence of the seasonal, local and non-linear variances σ_p, σ_l and σ_{nl}, time parameters Δt_l and Δt_{nl}, and scale parameter α_{nl} for values within the same order of magnitude (Supplementary Discussion section "Sensitivity to the Gaussian Process hyperparameters".). The limited influence of GP parameters within the same order of magnitude is due to the dense repeat coverage, and the relatively large measurement error of ASTER elevations (of about 5 m) which generally prevents complete deconvolution of local and periodic signals. This effect is later accounted for by our uncertainty propagation of interpolation biases (Section 4.3) and, when aggregated at different spatial scales, is essentially what defines the temporal resolution of our dataset (Supplementary Discussion section "Time series comparison and temporal resolution"). We found that the credible interval of the GP regression was the most impacted by parameter changes and was thus validated in a later analysis (Section 3.4). For estimating changes over stable terrain in the surroundings of glaciers, we used only a linear and seasonal kernel.

2.6. Elevation filtering step 3: iterative GP filtering of elevation

We applied GP regression iteratively, fitting an interpolated time series $h_{GP}(t,x,y)$ to our elevations $h_{AST/WV}(t,x,y)$ at a monthly temporal resolution from January 1st, 2000 to January 1st, 2020 for each pixel at a posting of 100 m. The interpolation period is 20 years, with boundaries positioned exactly three months before the earliest (early March, 2000) and after the latest (late September, 2019) observations. We used the GP error $\sigma_{h_{GP}}(t,x,y)$ propagated in time to filter any remaining outliers.

Ideally, we would filter the largest outliers first, and repeat this process iteratively outlier by outlier to avoid biasing new fits with previous outliers. However, as this process would be too computationally intensive, we chose instead to remove several outliers at once. For each pixel, we performed five iterative GP regressions for filtering, where we removed outliers successively from a 20-sigma ($\sigma_{h_{GP}}$) interval, a 12-sigma interval, a 9-sigma interval, a 6-sigma interval and, finally, a 4-sigma interval (Extended Data Fig. 3d).

2.7. Temporal GP interpolation of glacier elevation

After excluding the outliers, we performed a final GP regression at a monthly temporal resolution, to yield the final interpolated elevation time series $h_{GP}(t,x,y)$ with error $\sigma_{h_{GP}}(t,x,y)$ (Extended Data Fig. 3e). The repeat temporal coverage allowed us to correctly identify the expected seasonal minimum of glacier elevation at the pixel-scale (e.g. around the end of September in the Northern Hemisphere, and around the end of March in the Southern Hemisphere).

Qualitative evaluation of the time series showed that the fit and the GP credible interval performed well for typical glacier elevation change signals (Fig. S2). However, in the case of abrupt changes such as glacier surges, the method fails to represent the true temporal evolution of elevation even if the overall elevation change signal is captured (Fig. S3). Improving these aspects would require a classification of glacierized terrain for extreme events prior to constraining the temporal covariance and performing temporal interpolation, which was not feasible at a global scale.

3. Validation of elevation time series

The ICESat elevations are annotated as $h_{ICS}(t,x,y)$. The IceBridge elevations are annotated as $h_{ICB}(t,x,y)$. The elevations from high-resolution DEMs are annotated as $h_{HR}(t,x,y)$.

3.1. Comparing to ICESat

Before comparing ICESat elevations $h_{ICS}(t,x,y)$ to our interpolated elevation time series $h_{GP}(t,x,y)$, we verified that ICESat was aligned off-ice with our reference TanDEM-X $h_{TDX}(x,y)$ by computing the co-registration shifts between the two datasets over all stable terrain in the region[48]. ICESat showed very negligible horizontal and vertical shifts with the reference TanDEM-X in all regions. This is expected given that during its production the

TanDEM-X was co-registered to ICESat data[46]. Thus, we chose not to apply any horizontal or vertical shift.

To compare ICESat elevations $h_{ICS}(t,x,y)$ at the same points in time t_{ICS} as our elevation time series $h_{GP}(t,x,y)$, we performed a linear interpolation of our monthly elevation time series to extract elevation at the center date of each ICESat campaign. Most ICESat campaigns last 30 to 40 days, thus the longest possible time lag to their center date is 15 to 20 days. This 15 to 20-day period of time corresponds to about half of the temporal resolution of our monthly time series, which is why we considered using the middle ICESat campaign date a sufficient approximation for this analysis.

Finally, for each individual ICESat laser shot (point data), we performed a spatial bilinear interpolation of our elevations $h_{GP}(t,x,y)$ (gridded data) to estimate the elevation $h_{GP}(t_{ICS},x_{ICS},y_{ICS})$ at the center of each ICESat footprint (x_{ICS},y_{ICS}).

3.2. Comparing to IceBridge

For IceBridge elevations, we bilinearly downsampled our two sets of IceBridge elevations $h_{ICB}(t,x,y)$ (DMS-based and lidar-based) to 50 m (half the horizontal resolution of our time series). This resampling aimed simply at reducing the large amount of repeat spatial samples while conserving the vertical precision of the data. We then extracted elevations at the center of each pixel containing valid data, and subsequently compared to our interpolated time series by estimating $h_{GP}(t_{ICB},x_{ICB},y_{ICB})$ using the same procedure as described for ICESat data.

3.3. Systematic errors

We intersected over ten million ICESat and a hundred million IceBridge measurements in space and time with our interpolated elevation time series on both glacierized and stable terrain. We annotate validation elevations composed of both ICESat and IceBridge h_{IC}.

$$\Delta h_{GP/IC} = h_{GP} - h_{IC} \tag{S8}$$

To assess whether our elevation estimates were unbiased, we studied the median of elevation differences $\Delta h_{GP/IC}$. For all statistical operations, IceBridge points - about 40 times more dense spatially than ICESat points - were weighted at 1/40th of ICESat points to represent a comparable spatial sampling.

On average, glacierized terrain was found to be lower than true elevation, while stable terrain was slightly higher (Fig. S4a,b). These biases between stable terrain and glacierized terrain varied between regions (Table S3) and seasons (Fig. S4c). Our reference, TanDEM-X was also found to be higher over stable terrain, while we previously described the absence of vertical shift with ICESat (Section 3.1). The reason for this difference is that terrain here is limited to a 10 km buffer around glaciers instead of the whole region.

We attributed these biases to snow cover, originating from two distinct sources. First, we explained the seasonal variations by the fact that snow-covered terrain is not masked out from stable terrain during co-registration of the ASTER, ArcticDEM and REMA DEMs (Fig.

S5). The varying height of snow-cover in those DEMs directly leads to the observed seasonal biases. Second, we attributed the average differences between stable, glacierized terrain and our reference TanDEM-X to the fact that TanDEM-X has no defined timestamp but is a mosaic of different seasons, which also unevenly contains snow-covered terrain (Fig. S5a). At present, the lack of a homogeneous and global DEM with a well-defined seasonal timestamp prevents correcting these issues to improve the co-registration in regards to snow-covered terrain.

In order to derive volume change, our interest however lies in elevation change and not absolute elevation. Thus, the mean elevation bias between stable and glacierized terrain does not have an impact on elevation differences. The seasonal cycle of this bias has a systematic impact on seasonal elevation differences, however. We estimated this seasonal bias of co-registration by fitting a sinusoidal function (Fig. S4c ; Table S3). These sinusoidal fits revealed that the maximum and minimum bias occurred symmetrically around the dates of September 30th and March 30th for the Northern and Southern hemispheres, agreeing with prior knowledge. Additionally, the amplitude of the snow-cover biases were found to be relatively small for Arctic regions (RGI regions 3, 4, 5, 7, 9), large for mountainous areas (RGI regions 2, 8, 11, 12, 13, 14, 15, 17, 18) and absent in the Tropics (RGI region 16) (Table S3).

We used these sinusoidal fits to remove the seasonal biases in our elevation differences for further validation purposes. We derived a linear trend (WLS) in time from the remaining differences to ICESat and IceBridge (Extended Data Fig. 4d). We thereby verified the absence of elevation change bias in our GP time series globally and along several variables of interest. Less sampled regions yield larger trends, but do not statistically differ from zero (Table S3).

We also identified elevation biases with curvature (Fig. S4d), on both stable and glacierized terrain. This bias was also observable when differencing to the reference DEM TanDEM-X on stable terrain, suggesting that it originates from an inherent difference between the datasets rather than our temporal interpolation method, as shown in previous studies[69]. The lower resolution of our DEMs does not allow us to reliably capture elevation in places of high curvatures such as peaks, ridges or narrow valleys. Being independent of time, this bias does not need to be accounted for when differencing the elevation time series into elevation change and does not affect our glacier mass balance estimates.

3.4. Random errors

For each ICESat and IceBridge observation, we also derived a z-score, or elevation difference divided by our time series error at a given point of time and space:

$$z_{GP/IC} = \frac{(h_{GP} - h_{IC})}{\sigma_{h_{GP}}} \qquad \text{(S9)}$$

This standardized metric allows us to compare the performance of the elevation time series h_{GP} in relation to its GP credible interval $\sigma_{h_{GP}}$ which varies significantly in space and time based on factors such as elevation measurement errors, number of observations, or time lag to the closest observations (Figs. S2, S3). Ideally, for a performant interpolation method, the z-score would be normally distributed, centered on 0 with a standard deviation of 1. A mean of 0 would signify that our elevation estimates h_{GP} are unbiased and a standard deviation of 1 would signify that our modelled error $\sigma_{h_{GP}}$ is representative of the elevation error at the right confidence level.

We used the NMAD as a robust estimator of the standard deviation of the z-score distribution (Fig. S6). The standard deviation of z-scores was verified to be independent of several factors: the terrain slope, the terrain curvature, the time of the year, the time lag to the closest elevation observation, and the size of the GP credible interval. The lack of dependency on these two last factors validates the reliability of our empirically-based GP parameters (Section 2.5). The z-score standard deviation was found to be very close to 0.5 for glaciers (Fig. S6, Extended Data Fig. 4d) and close to 1 for stable terrain (Table S3) in all regions, implying that our GP method is able to estimate elevation in space and time at the right confidence level. Over glaciers, our elevation uncertainties are thus conservative by a factor of over two.

4. Spatial correlation of elevation change time series

We identified three sources of correlation to account for when propagating uncertainties from our elevation change time series into volume change time series. The first source is the short range spatial correlation originating from instrument resolution (Section 4.1). The second is the long range spatial correlation originating from instrument noise, generally not accounted for in earlier studies (Section 4.2). The third and final source originates from our temporal interpolation (Section 4.3). We accounted for these three sources of spatial correlation when aggregating both from pixels to individual glaciers and from glaciers to regional volume change estimates (see Aggregation to regions in Methods).

4.1. Spatial correlation due to instrument resolution

Conservatively, we estimated spatial uncertainties that originate from instrument resolution solely on ASTER, which has the coarsest resolution. We derived a short-range spatial variogram[80] over ASTER DEMs to assess the short-range spatial correlation that can be attributed to the effects of spatial resolution $\sigma_{dh,res}$:

$$\sigma_{dh,res}(d)^2 = S(d, s_0, r_0) \qquad \text{(S10)}$$

where d is the spatial lag, or distance between pixels, $S(d,s,r)$ the spherical model of partial sill s and range r. We found a range of $r_0 = 150$ m. The partial sill s_0 derived in our spatial variograms is only a representative average of the many factors affecting elevation variance (Section 2.3). Instead of using this average, we estimated s_0 from our individual pixel errors

$\sigma_{dh_{GP}}$ over glacierized terrain after removing contributions from longer range correlations (described in Section 4.3).

4.2. Spatial correlation due to instrument noise

We accounted for long-range correlations due to instrument noise, such as the ones present in ASTER DEMs that we described in Section 1[50,82]. Depending on the distribution of stable terrain in a specific DEM, it can be difficult to fully correct the low-frequency along-track undulations, and even more so for the high-frequency along-track undulations (also called "jitter")[83]. We know from ASTER corrections that the low-frequency along-track undulation has a range of about 5 to 20 km, while the jitter has a range of 2 km. Residuals from this correlated noise can be described by a sum of nested spatial variograms at these correlation lengths with specific variances, which in turn can be used to derive the uncertainty in the volume changes obtained by spatial aggregation[72].

In order to account for improvements brought by corrections (Section 1), filtering and interpolation (Section 2) we deduced our spatial correlations directly from the elevation differences to ICESat and IceBridge data $\Delta h_{GP/IC}$ (Section 3). The procedure for estimating the spatial correlation is the same as used for longer range correlation which is described extensively in the next section. We modelled our empirical variogram by a sum of spherical variograms $\sigma_{dh,noise}$ with correlation lengths of 2 km, 5 km and 20 km:

$$\sigma_{dh,noise}(d)^2 = S(d,s_1,r_1) + S(d,s_2,r_2) + S(d,s_3,r_3) \qquad \text{(S11)}$$

with $r_1 = 2$ km, $r_2 = 5$ km and $r_3 = 20$ km.

The sills we estimated indicate that, for the average ASTER DEM, we corrected on average around 90% of the typically observed low-frequency along-track undulation of range 20 km, and about 70% of the high-frequency undulation (jitter) of range 2 km.

4.3. Spatial correlation due to temporal interpolation

Previously, we treated our elevation time series uncertainty $\sigma_{h_{GP}}$ as a random error term (Section 3.4), but this qualification does not always hold. Because of the nature of the spatial coverage of DEMs, temporal data gaps are often spatially correlated over large areas. Although we know that our GP credible interval $\sigma_{h_{GP}}$ contains the true elevation at the right confidence level, the elevation difference to the true elevation might be spatially correlated. For example, let's assume that there are three years without data for the entire region of Svalbard and that we independently interpolate all pixels of the region in time. Assume also that in the middle of that period, one winter had stronger accumulation than usual. At this point in time, our GP credible interval might still contain the true values of glacier elevations at the pixel scale (Figs. S2, S3), but at the regional scale this stronger accumulation manifests as a systematic error, or as a random error correlated over large distances.

We accounted for the spatial correlation inherent to the temporal interpolation to correctly propagate uncertainties into the resulting volume changes. For this, we used the ICESat validation data only, which has a more diverse regional coverage than IceBridge. We categorized our validation elevation differences $\Delta h_{GP/ICS}$ by date, corresponding to ICESat campaigns. For each campaign, at the pixel scale, we also categorized $\Delta h_{GP/ICS}$ by time lag $\Delta t_{AST/WV}$. The time lag is the time between the ICESat acquisition and the closest valid ASTER/WV observation $h_{AST/WV}(t_{AST/WV}, x, y)$ used to derive the time series during the temporal interpolation:

$$\Delta t_{AST/WV}(t, x, y) = \left| t - t_{AST/WV} \right| \tag{S12}$$

For each region, date and time lag bin, we randomly drew up to 10,000 observations of validation elevation differences $\Delta h_{GP/ICS}$ and quantified their spatial correlation[81]. We did so by computing their variance correlations at distances of 0.15 km, 2 km, 5 km, 20 km, 50 km, 200 km and 500 km (Extended Data Fig. 5a,b). The first four ranges of 0.15 km, 2 km, 5 km and 20 km correspond to the correlation ranges for instrument resolution and instrument noise (preceding sections). We found similar variograms between well-sampled regions and chose to aggregate and apply a single variogram globally for robustness in less-sampled regions. This approach is expected to be conservative for instrument noise, due to the larger residuals of correlated signals captured over large ice caps (corrected with less stable terrain) that dominate the ICESat sample. Aggregating all available variograms, we derived the median variance of elevation differences $\Delta h_{GP/IC}$ at different correlation lengths depending on the time lag to the closest observation. We showed that there was no significant correlation in the variance beyond 500 km, a finding consistent with known estimates of correlations of glacier mass balances[77], and we thus considered elevation differences to be independent beyond this range. We modelled these correlations as a sum of spherical variograms:

$$\sigma_{dh}(d, \Delta t)^2 = \sum_{i=0}^{6} S(d, s_i(\Delta t), r_i) \tag{S13}$$

where $r_4 =$ 50 km, $r_5 = 200$ km. We found $s_6(\Delta t) = 0$ for $r_6 = 500$ km. The partial sills s_0 to s_5 depend on the time lag to the closest valid observation Δt. For s_1 to s_5, we used the empirical values of the partial sills to estimate it as a function of the time lag. We found a good fit between the complete sills (sum of partials sill) and a sum of quadratic and squared sinusoidal functions by least squares optimization (Extended Data Fig. 5b):

$$\sum_{i=1}^{k} s_i(\Delta t) = (a_k \cdot \Delta t + b_k)^2 + c \cdot sin(d \cdot \Delta t)^2 \tag{S14}$$

where we found two constants for the sinusoidal component $c = 1.4$ m^2 and $d = 1$ yr. At $\Delta t = 0$, we deduced the values of $s_1(0)$, $s_2(0)$ and $s_3(0)$ described in the previous section on instrument noise.

We accounted for the random error due to instrument resolution, represented by the partial sill s_0 at a correlation length of $r_0 = 150$ m, using the original pixel error $\sigma_{dh}(t,x,y)$. To avoid double-counting errors in s_0, we quadratically subtracted the partial sills of range r_1 to r_5 to $\sigma_{dh}(t,x,y)$:

$$\sigma_{dh,0}(t,x,y)^2 = \sigma_{dh}(t,x,y)^2 - \sum_{i=1}^{6} s_i(\Delta t) \tag{S15}$$

For each glacier and time step, we used the mean of all pixel errors $\sigma_{dh,0}{}^2$ over glacierized terrain to estimate s_0. We then spatially integrated this sum of variograms[72] over each glacier, for each time step, accounting for each individual glacier pixel's distance to the closest observation Δt at this time step to yield the uncertainty in the mean elevation change $\sigma_{\overline{dh}}$ (see Methods, Equations 4 and 5).

Supplementary Discussion

Improved elevation change estimation

Our surface elevation change maps (Extended Data Fig. 6, and Data availability statement) visually testify to the improved quality brought out by our novel methodology in comparison to preceding ASTER-based studies[28,53,54,74]. It is most notable through the reduction of elevation change outliers and through stable terrain surfaces that consistently show elevation changes close to 0 m. The improved quality and increased spatial coverage is the combined result of (i) the inclusion of low stereo-matching correlation data that is statistically weighted, (ii) a multi-step filtering of outliers conditioned by a reference elevation model, an elevation measurement error and temporal covariances throughout the different steps of our statistical modelling approach, (iii) the mitigation of seasonality, and other non-linear responses during the final interpolation into surface elevation time series. Although having fewer data gaps than preceding studies, our time series of elevation change still contain uncertainties that vary in space and time for different regions, periods, and for different areas of a given glacier (e.g. low-contrast accumulation areas). Stable terrain is not directly comparable to glacierized terrain, as it only includes a pairwise linear kernel and a periodic kernel during GP regression (i.e. a linear trend with mitigated seasonality). Non-linear kernels were not applied to stable terrain as those were estimated using glacier elevation observations only, non-applicable to other types of terrain (bare-ground, vegetated, ...). Nonetheless, the linear estimation allows similar 20-year changes to be captured at the boundary of glacierized and stable terrain. For instance, unmapped debris-covered tongues treated as stable terrain show long-term elevation changes consistent with the rest of the glacier.

Subaqueous mass loss

Our analysis is limited to measuring glacier elevation change above water by assessing only topographic changes and not bathymetric ones. The subaqueous mass loss due to retreat of ice fronts over water for marine- or lake-terminating glaciers cannot be captured, leading to an underestimation of the total glacier mass loss[29]. This omitted underwater contribution is largely dominated by marine-terminating glaciers, however, and thus does not impact the contribution to sea-level rise. This limitation is shared between geodetic, gravimetric (for marine-terminating glaciers) and altimetric (ICESat) estimates.

Time series comparison and temporal resolution

We compare our results to the time series of previous studies[19,21] (Fig. S7) and, for this exercise, use the same density conversion factor and errors for annual mass change rates despite the known limitations of density conversion assumptions at such short time scales[23]. We find good agreement to gravimetric time series[19] in regions where competing mass change signals are weak (e.g., Russian Arctic), and to geodetic and glaciological-based time series in regions with dense in-situ measurements[21]. At the regional scale, we generally observe a temporal autocorrelation, or "smoothing" of extreme glaciological years in our time series, either due to the limits of local and seasonal elevation change deconvolution from our Gaussian Processes method (where repeat data coverage is good, e.g. Iceland), or originating from the near-linear interpolation over long observational data gaps (where data coverage is more limited). In practice, the rapid change of extreme years is redistributed into adjacent years. This effect is accounted for by our improved uncertainty approaches calibrated on spatial correlations with observational time lag, validated with high-resolution data (Extended Data Fig. 5d-f). Based on our volume change uncertainties, the temporal resolution at which volume changes are statistically significant at the regional scale (95% confidence interval <0.2 m yr^{-1}) is of 3-7 years depending on the spatial domain and temporal coverage. This is confirmed by the inter-comparison to regional estimates of temporally resolved, gravimetric studies.

Decadal changes in summer temperature and winter precipitation

For the Northern Hemisphere, we define winter to coincide with the accumulation period which, for high-altitude glaciers is typically October-April whereas we define the summer season to include the months May-September. Winter and summer seasons in the Southern Hemisphere are respectively defined as the months April-October and November-March. Decadal changes between annual and seasonal components show minor differences of precipitation patterns (Fig. S8). However, the increase of summer temperature over glaciers (corresponding to a global trend of 0.037 K yr^{-1}) is slightly larger than that of annual temperature (0.031 K yr^{-1}) and thus yields a slightly lower global mass balance sensitivity to temperature of -0.24 m w.e. yr^{-1} K^{-1}. We also find little difference in temperature and precipitation change conditions between tidewater and non-tidewater glaciers aggregated at the global-scale.

Uncertainty propagation and limits of density-based mass change uncertainties

Volume change uncertainties sources are dominated by short- to long-range spatial correlations (2-200 km) and our pixel-wise GP uncertainties only have influence for very small glaciers (<0.1 km², Extended Data Fig. 5h,g). While our volume change time series have rigorously constrained uncertainties (Extended Data Fig. 5d-f), density-based mass change uncertainties are still poorly known. We thus use conservative approaches that likely lead to an overestimation of mass change uncertainties. We conservatively applied our uncertainties by considering them completely correlated in space at the scale of RGI regions, which leads to larger regional mass change uncertainties than previous DEM-based studies[29,53,54] that considered subregions of RGI regions as independent. Additionally, the current formulation of the density uncertainty[23] linearly scales with specific elevation changes while it is known that it is with most negative elevation change rates that the density conversion factor is best constrained towards 850 kg m^{-3}. This effect likely provides uncertainties that are too large, especially for regions with strong mass losses (e.g., Fig. 3).

This limitation is currently shared among all existing geodetic studies due to the poorly constrained dependencies between density conversions and specific mass change rates, and poorly known spatial and temporal correlations of density conversions. Progress in the understanding of these conversion factors might be made possible by coupling our observational baseline with glaciological models.

Sensitivity to the Gaussian Process hyperparameters

In addition to the validation, we performed sensitivity analysis of the parameters to ensure that our results were not significantly affected by the choice of using a global median temporal covariance. By making all parameters vary within an order of magnitude from the values defined through our empirical covariance (Section 2.5), we computed all elevation time series for two glacierized regions, Iceland, with the densest temporal sampling; and Scandinavia, with the sparsest temporal sampling (Antarctic and Subantarctic excluded). We selected these two regions as potentially the most sensitive ones because of (i) their nonlinear evolution during 2000-2019 (Extended Data Table 1), and (ii) their small size which implies strongly spatially correlated signals. Our results show little sensitivity (Fig. S9), with a mean absolute deviation of less than 3%, falling well within volume change uncertainties for these regions (<30% of the 95% uncertainty range of volume change, <10% for that of mass change). This low sensitivity is due to the dense temporal coverage of our observations (on average 39 in 20 years), which results in a minor overall influence from the interpolation approach based on GP regression. The GP approach is harnessed in great part to better filter outliers and mitigate the effects of seasonality. While there is no doubt that further characterization of underlying nonstationarities in the temporal covariance could allow for improved analysis of local glacier elevation change, especially for elevation data of higher vertical precision, the current parametrization is therefore sufficiently robust for the dense repeat coverage of ASTER and its vertical precision..

Inventory biases

Based on visual inspection of glacier inventories superimposed with our elevation change maps over the entire period of 2000 to 2019, we noticed that glacier outlines based on recent imagery (acquired later than 2000) were in some cases omitting part of the glaciers lowest elevations. Factors that could explain this pattern include exclusion of debris-covered ice in the glacier inventory or substantial glacier retreat over the period 2000 to 2019[74]. Therefore, we note that the errors due to inventory and outlines, considered as random errors in our analysis, most likely integrate some systematic effects that might lead to a slight underestimation, within uncertainties, of global glacier mass loss.

Supplementary Figures

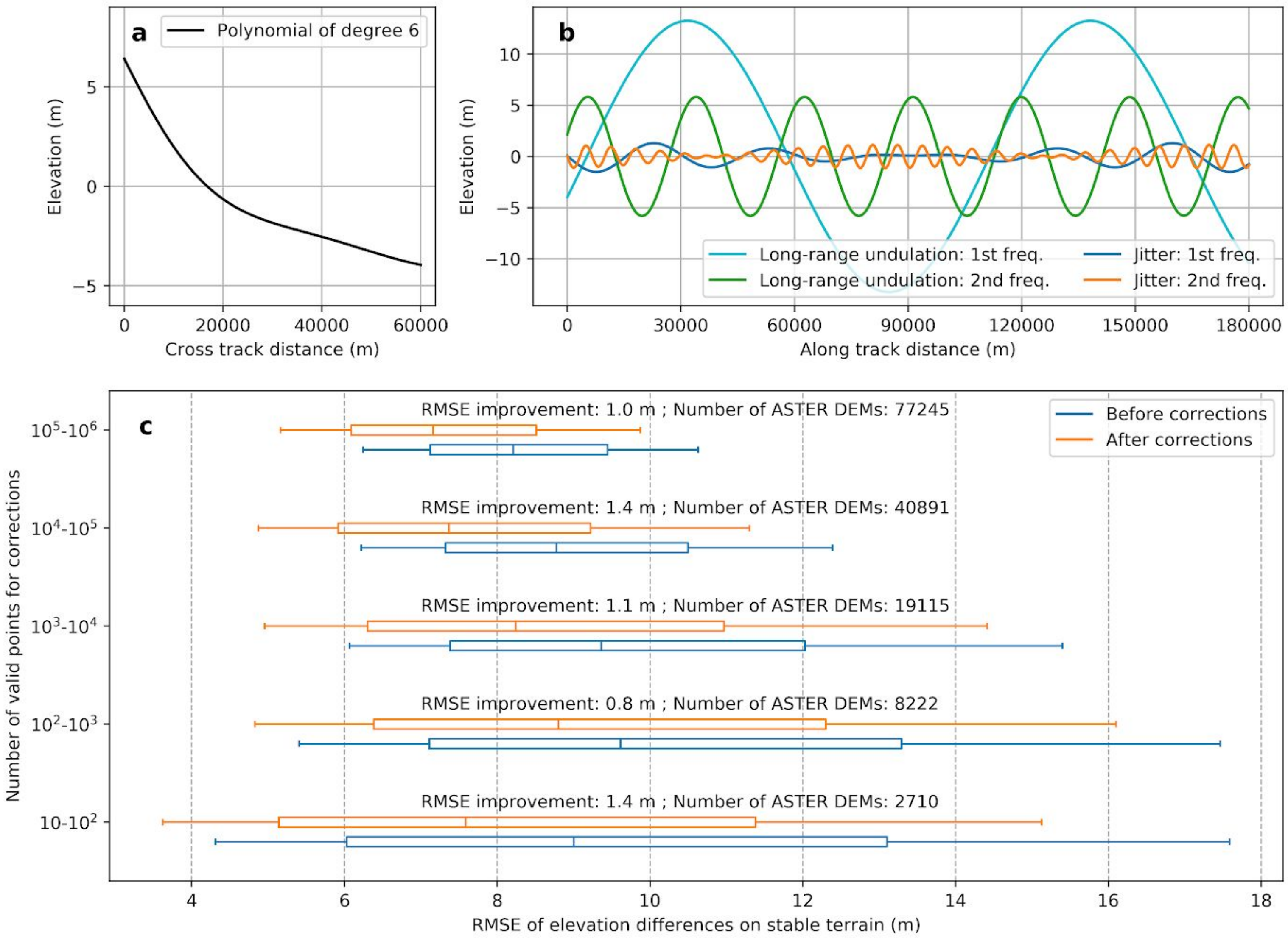

Fig. S1. ASTER bias corrections.

MMASTER corrections[50] for cross-track (**a**) and along-track biases (**b**) and overall (**c**). For (**a**) and (**b**), elevation corrections shown correspond to the median of parameters (i.e. the most typical case of correction, which vary greatly between images) found when optimizing corrections individually for 154,565 ASTER DEM strips. For the along-track corrections, we show the median of parameters decomposed by additive frequencies of long range undulations and jitter. **c**, Boxplots of off-ice RMSE between ASTER DEMs and the reference TanDEM-X before (blue) and after (orange) bias correction with the number of valid points for corrections. Boxes show the distribution from first to third quartiles, with the median represented as a line and the whiskers extending to the 10th and 90th percentiles. RMSE improvement is a limited but simple indicator of the underlying removal of large-scale correlated noise on top of the inherent random noise, typically observed at an amplitude of 5 to 10 m in ASTER DEMs. Most DEMs are confidently corrected with a large number of points which reflects in their lesser spread of RMSE, later used to assess measurement error and weighting.

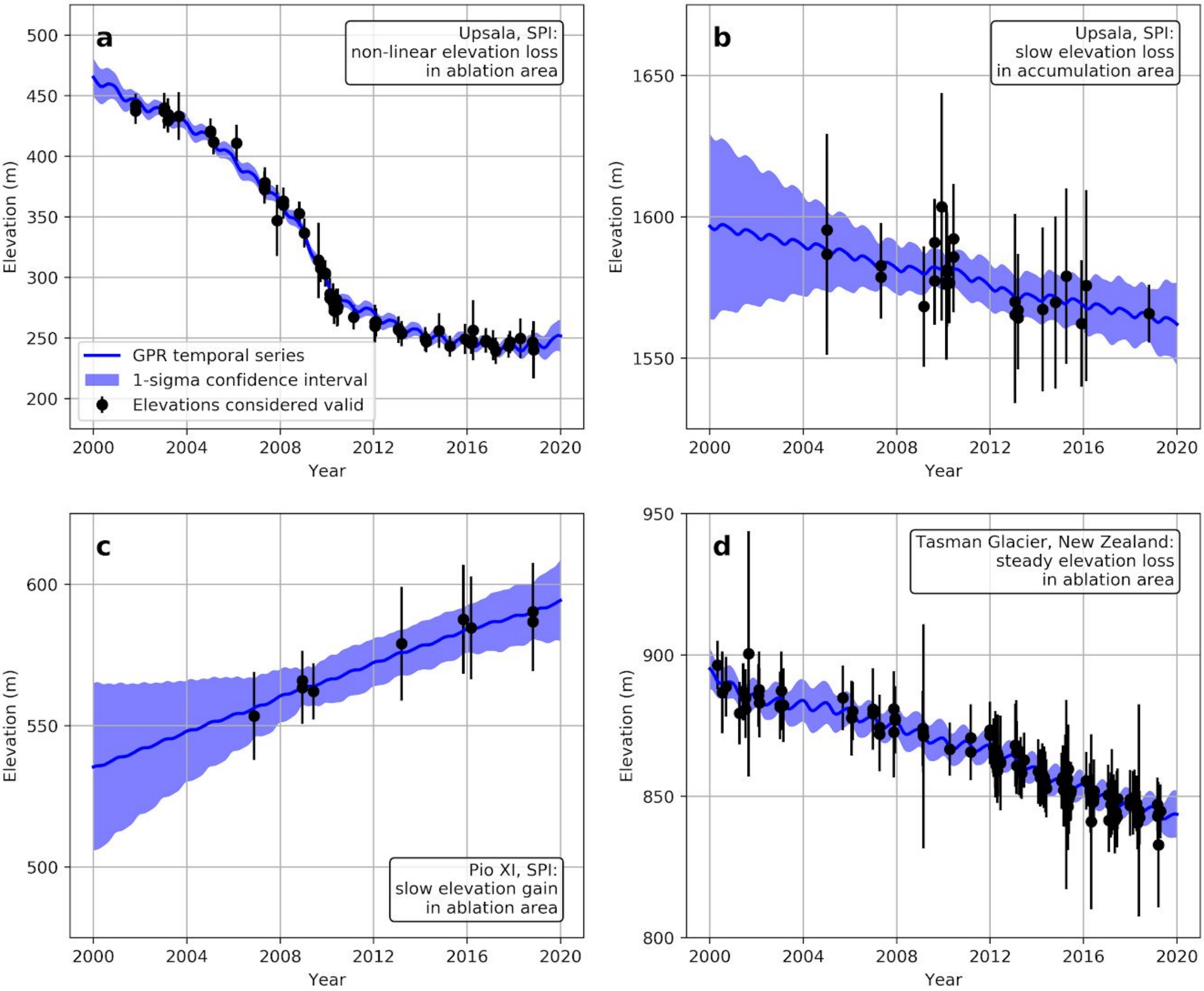

Fig. S2. Gaussian Process regression elevation time series.

Time series of final Gaussian Process Regression fits and standard deviation after the removal of outliers. These temporal fits are shown for: a pixel in the ablation area of Upsala where a strongly nonlinear elevation loss occurred[99] which is used as an example in Extended Data Fig. 3c-e (**a**), a zone of low quality of stereo-correlation in the accumulation area of Upsala, Southern Patagonian Icefield, Argentina that is undergoing slow elevation loss (**b**), the data-scarce ablation area of Pio XI, Southern Patagonian Icefield, Chile, facing a steady elevation gain (**c**) and the highly sampled tongue of Tasman Glacier, New Zealand showing steady elevation loss (**d**).

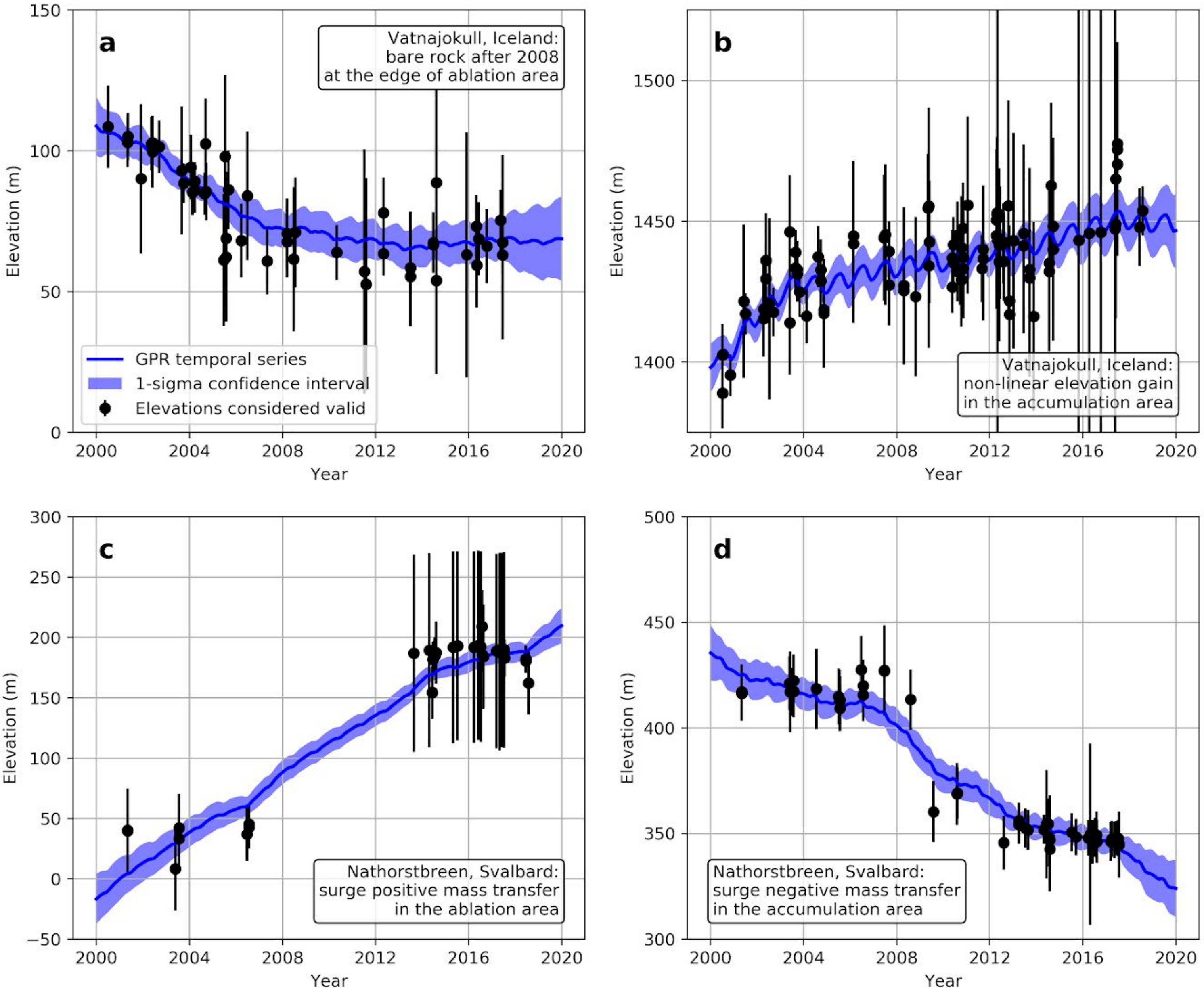

Fig. S3. Gaussian Process regression elevation time series for extreme observations.

Time series of final Gaussian Process Regression fits and standard deviation over extreme elevation observations, after the removal of outliers. We show pixels in: the edge of the ablation area of outlet glacier Breidamerkurjökull of Vatnajökull, Iceland, showing rapid thinning and then no elevation change after 2008 due to complete deglaciation (**a**), the nonlinear thickening on the outlet glacier Dyngjujökull, in the accumulation area of Vatnajökull, Iceland (**b**), the elevation gain at the bottom of Nathorstbreen glacier, Svalbard, after a massive surge in 2009[40] (**c**) and the mass loss in the accumulation area of Nathorstbreen glacier, Svalbard, from the same surge event (**d**).

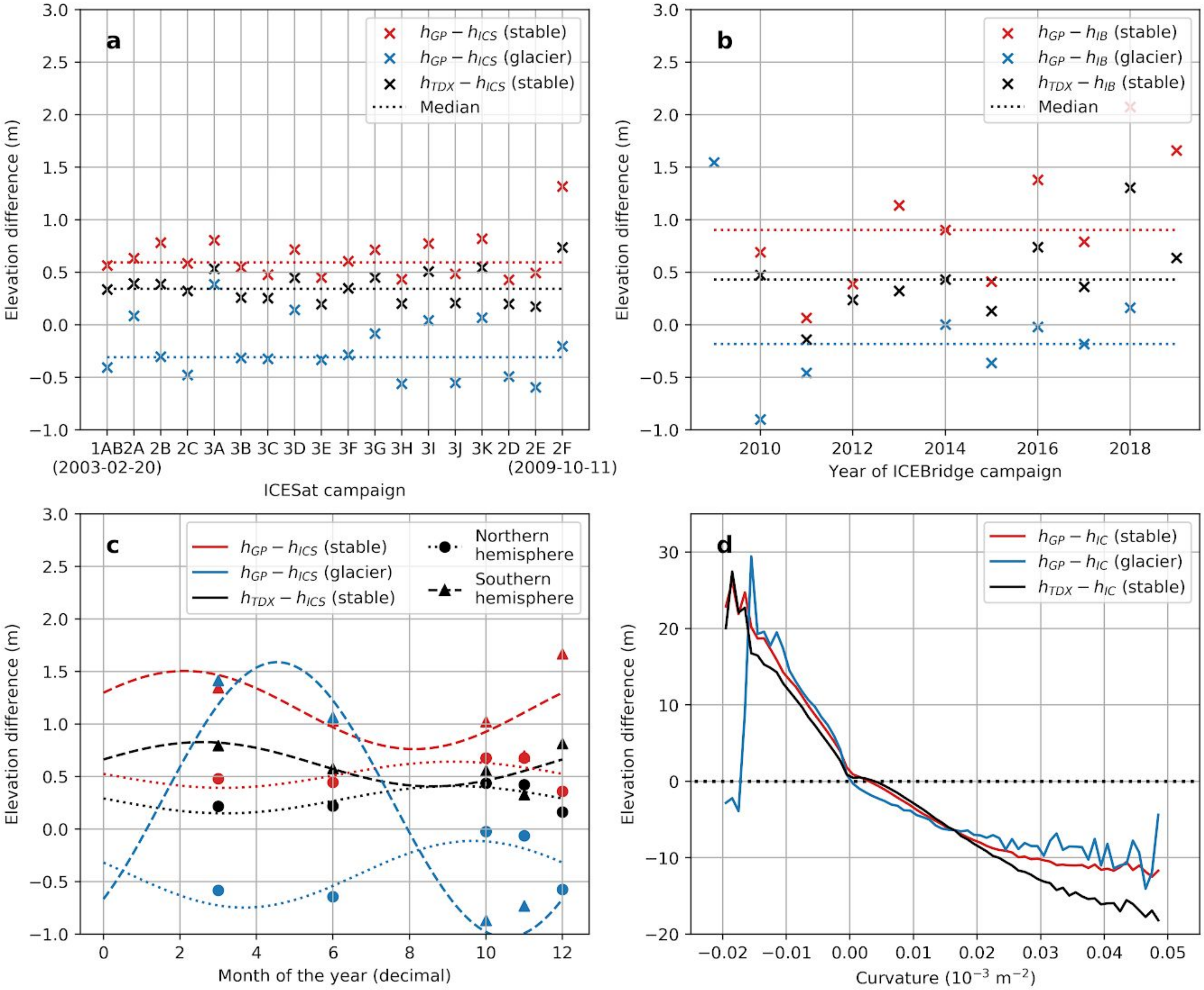

Fig. S4. Systematic error analysis.

Median elevation differences between ICESat/IceBridge and our GP regression elevation time series on stable and glacierized terrain, and for TanDEM-X on stable terrain. Elevation differences are shown globally for all ICESat campaigns (**a**), all IceBridge campaigns (**b**), seasonal timestamps independently for the Northern and Southern hemisphere (ICESat samples only, Antarctic and Subantarctic excluded) (**c**) and terrain maximum curvature (**d**). Panels (**a**), (**b**) and (**c**) are shown on the same scale.

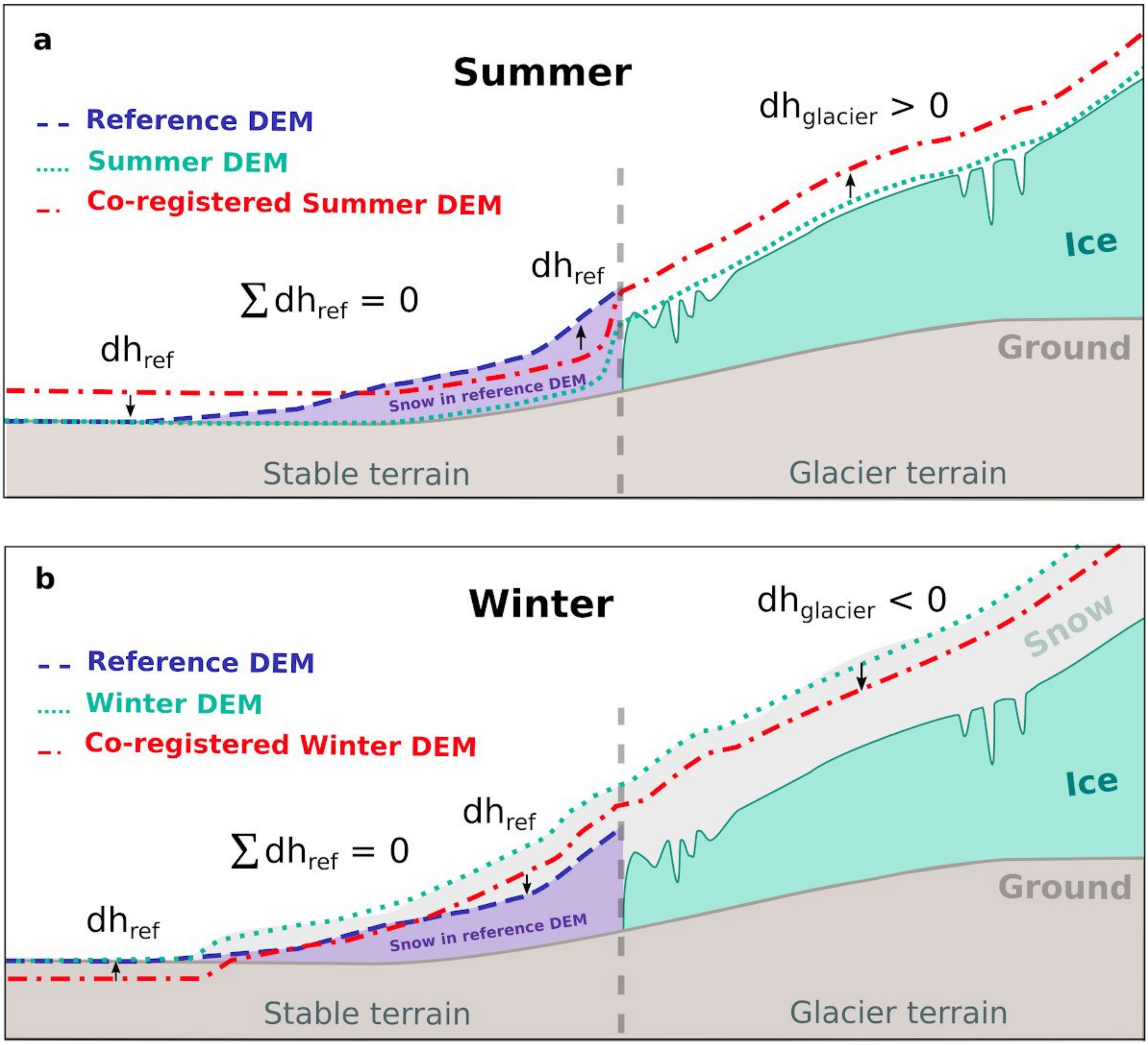

Fig. S5. Schematic representation of the effects of snow-covered terrain on co-registration.

Summer and winter co-registration biases. **a**, The coregistration of summer ASTER DEMs to a reference DEM (TanDEM-X) that includes snow cover on stable terrain creates a positive elevation bias over glaciers. **b**, The coregistration of winter ASTER DEMs to a reference DEM (TanDEM-X) that, on average, includes less snow, creates a negative elevation bias over glaciers.

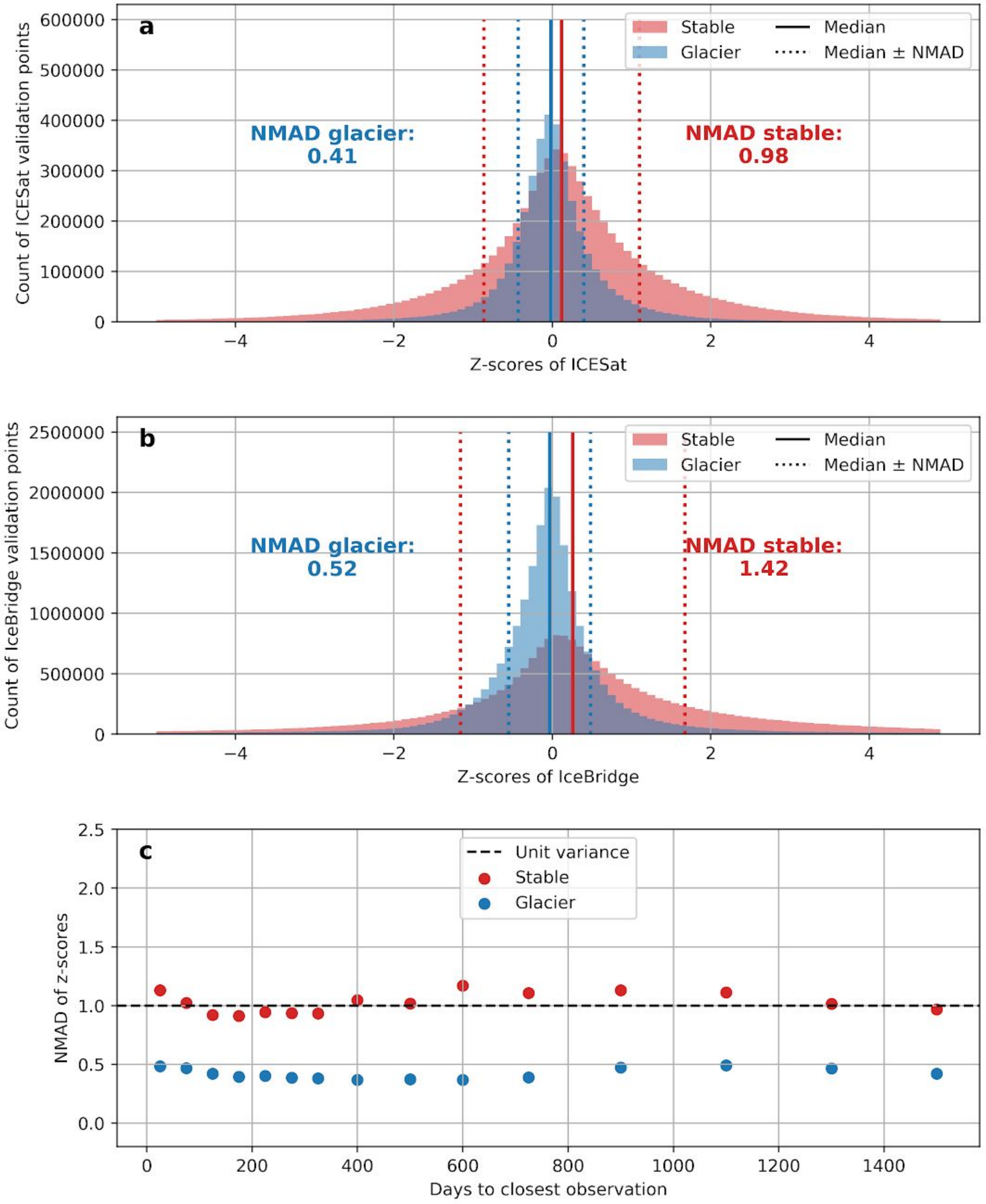

Fig. S6. Random error analysis.

Distribution of z-scores of elevation differences for stable and ice-free surfaces. The distribution of ICESat (**a**) and IceBridge (**b**) z-scores for stable and glacierized terrain with the corresponding median and Normalized Median Absolute Deviation. **c**, NMAD of the z-score with the time lag to the closest observation. IceBridge points are weighted 1/40th to represent similar spatial sampling as ICESat points.

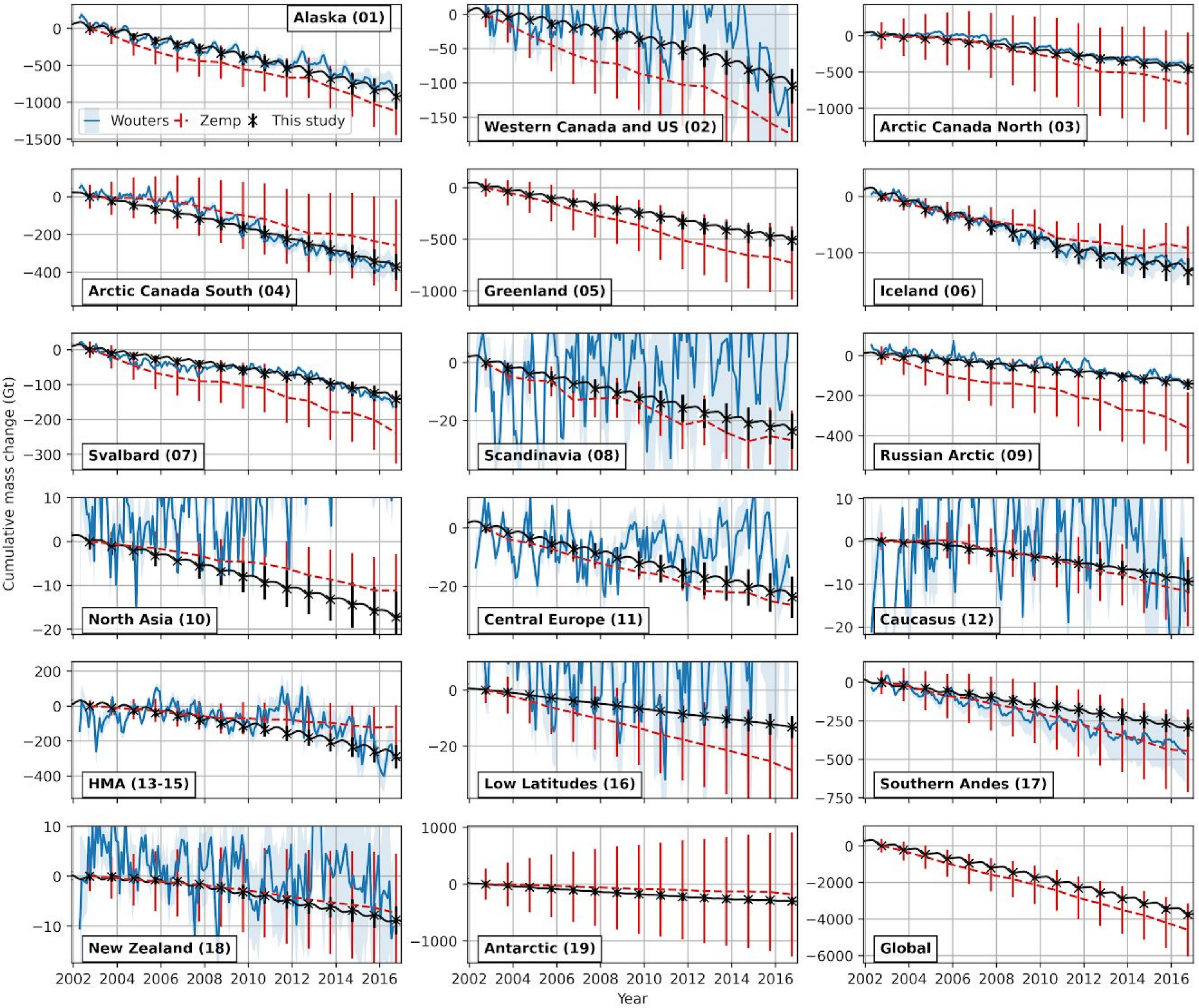

Fig. S7. Comparison of mass change time series with earlier studies.

Regional and global time series of cumulative glacier mass change between 2002 to 2016 compared to earlier global or near-global studies[19,21]. Time series are zeroed at the starting date of September 2002 to coincide with the northern hemisphere glaciological years reported in an earlier global study[21]. Uncertainties are shown at 95%. For our study, we show annual uncertainties despite known limitations for density assumptions and show our monthly time series uncorrected for seasonal biases.

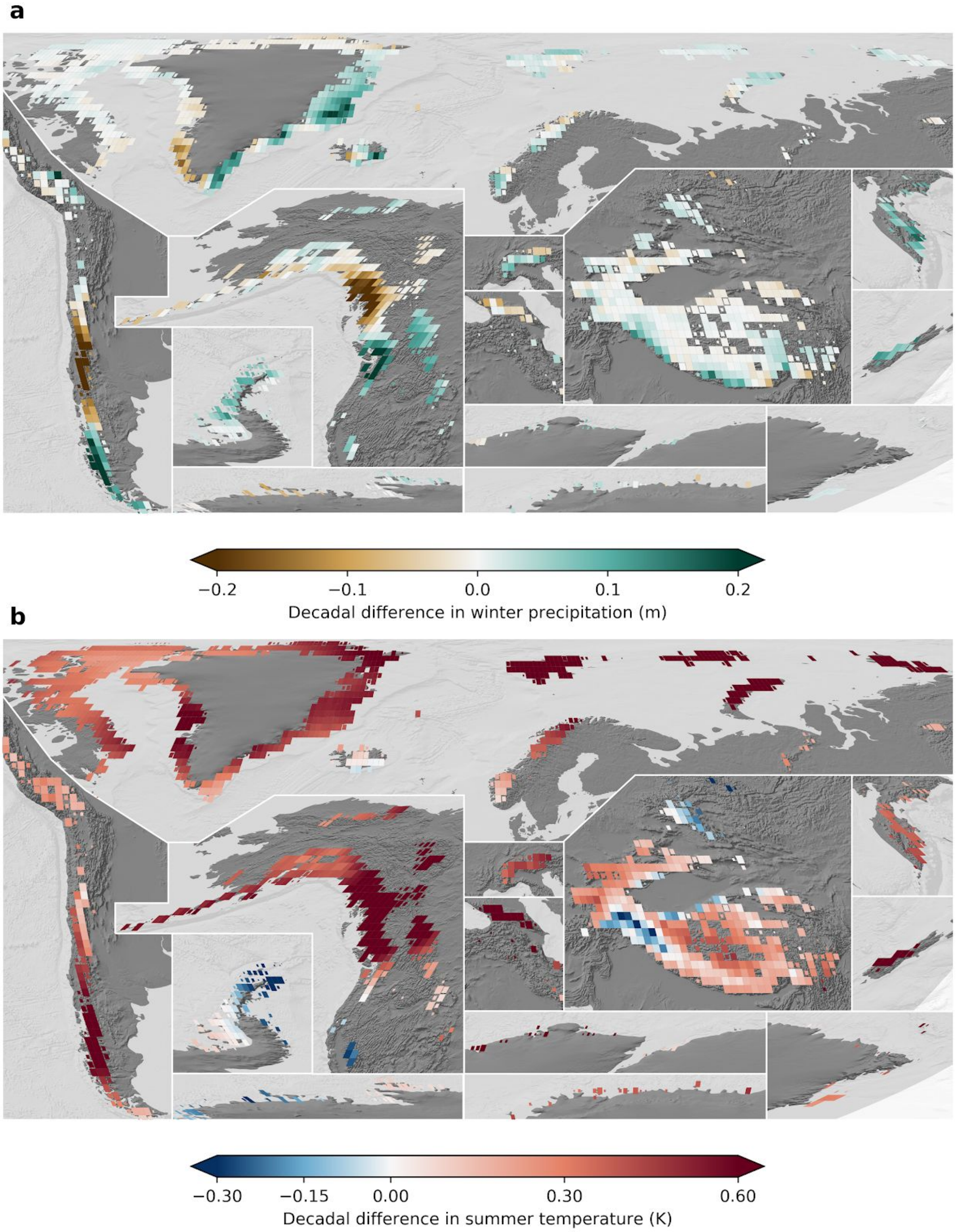

Fig. S8. Decadal changes in summer temperature and winter precipitation.

Difference between 2010-2019 and 2000-2009 for summer temperature (May-September in the northern hemisphere, October-April for southern hemisphere) (**a**) and winter precipitation (November-March in the northern hemisphere, April-October in the southern hemisphere) (**b**). Decadal patterns are similar to those of annual temperature and precipitation (Fig. 4).

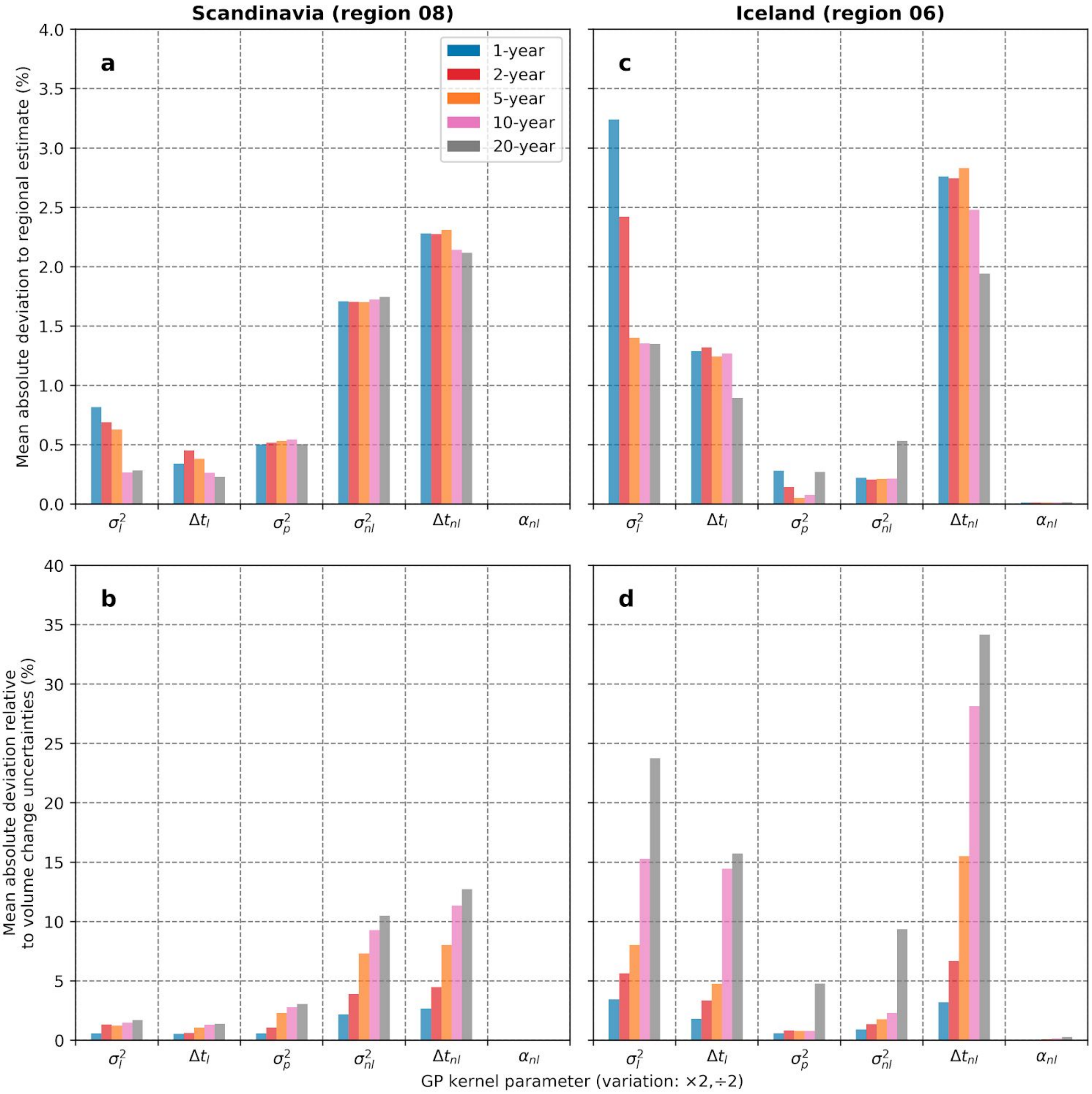

Fig. S9. Sensitivity to Gaussian Process kernel parameters.

Sensitivity of the regional estimates of glacier volume changes to the Gaussian Process kernels parameters for Scandinavia **(a,b)** and Iceland **(c,d)**. The kernel parameters are varied by multiplying and dividing the value used in this study by 2, and refer to Equation 2 (see Methods, or Supplementary Methods section 2.5). Iceland and Scandinavia were selected as they are potentially the most sensitive to Gaussian Process kernel parameters. This is due to both their small size (spatially correlated signal) and the fact that they show strong nonlinear changes during the past two decades (Extended Data Table 1). Additionally, they include a wide spectrum of temporal coverage, as Iceland is the region with the largest repeat coverage (~66 observations in 20 years per pixel) while Scandinavia is the region with the lowest repeat coverage (~27 observations in 20 years per pixel), excluding the Antarctic and Subantarctic. Panels **(a)** and **(c)** show the mean absolute deviation relative to the regional estimate and panels **(b)** and **(d)** the mean absolute deviation relative to the estimated volume change uncertainty. The mean absolute deviation is computed from all possible successive time periods of a certain length in 2000-2019 (e.g., 5-year periods indicate 2000-2004, 2005-2009, 2010-2014 and 2015-2019) and varied parameters (x2,÷2). Overall, varying all Gaussian Process kernel parameters within this order of magnitude impacts the estimates less than 3%, which is well within estimated volume change uncertainties (at most 30% of uncertainty range) and estimated mass change uncertainties (at most 10% of uncertainty range). The maximum absolute deviation is within the same range and does not exceed 1.5 times the mean absolute deviation.

Supplementary Tables

Region	ASTER initial DEMs	ASTER final DEMs	ArcticDEM & REMA final DEMs	ICESat points	IceBridge DEMs
01, 02 Alaska & Western Canada and USA	28,705	24,681	3,763	1,137,548	273
03 Arctic Canada North	9,610	7,645	4,209	2,220,084	78,434
04 Arctic Canada South	3,702	3,013	2,832	584,499	39,113
05 Greenland	24,290	16,837	14,353	2,375,214	334,721
06 Iceland	3,439	2,569	701	72,811	0
07 Svalbard and Jan Mayen	2,870	1,750	3,507	350,053	4,882
08 Scandinavia	3,432	2,770	1,143	16,157	0
09 Russian Arctic	9,137	4,485	5,635	596,456	0
10 North Asia	15,693	12,867	4,248	309,226	0
11 Central Europe	4,499	4,304	0	2,277	0
12 Caucasus and Middle East	5,745	5,563	0	31,580	0
13-15 High Mountain Asia	31,676	30,774	0	2,169,746	0
16 Low Latitudes	15,689	14,800	0	61,845	0
17 Southern Andes	9,076	8,124	0	237,559	495
18 New Zealand	2,221	1,922	0	13,998	0
19 Antarctic and Subantarctic	28,555	12,461	3,456	559,639	154,770

Table S1. Regional data coverage for ASTER, ArcticDEM, REMA and IceBridge DEMs.

Number of DEMs and corresponding area covered for RGI regions. ASTER initial DEMs are the strips generated from stereo imagery. ASTER, ArcticDEM and REMA final DEMs are the DEM strips successfully corrected (for ASTER), merged (for ArcticDEM and REMA) and co-registered to TanDEM-X (for both). ASTER initial and final DEM footprint is generally 180 km x 60 km, ArcticDEM and REMA DEM footprint is on average 50 km x 15 km and IceBridge DEMs 500 x 500 m. ICESat points are only considered in a buffer of 10 km around glaciers, the count of points on glacier and stable terrain is reported in Table S3.

Region	Site	Early DEM date	Early DEM source	Late DEM date	Late DEM source	Resolution (m)	Glacier area (km²)	Number of glaciers covered
01 Alaska	Yukon	2007-09-03	SPOT-5HRS	2018-10-01	SPOT6/7	20	2387.4	139
02 Western Canada and USA	Place	2006-08-15	LiDAR	2018-09-04	LiDAR	1	4.9	7
02 Western Canada and USA	Sentinel	2006-08-15	LiDAR	2018-09-04	LiDAR	1	22.3	10
02 Western Canada and USA	Bridge	2006-08-15	LiDAR	2017-09-27	LiDAR	1	142.6	23
02 Western Canada and USA	Weart	2006-08-15	LiDAR	2018-09-04	LiDAR	1	12.8	9
02 Western Canada and USA	Conrad	2005-07-31	Aerial Photo	2017-09-17	Aerial Photo	5	19.4	11
02 Western Canada and USA	Nordic	2004-08-21	Aerial Photo	2017-09-27	Aerial Photo	5	5	3
11 Central Europe	Mont Blanc	2003-08-21	SPOT5-HRG	2018-08-09	Pleiades	10	141.1	105
11 Central Europe	Gries	2012-08-27	Aerial Photo	2018-08-19	Aerial Photo	1	6	10
11 Central Europe	Silvretta	2012-08-20	Aerial Photo	2018-08-16	Aerial Photo	1	6.8	7
11 Central Europe	Plaine Morte	2012-09-14	Aerial Photo	2018-08-28	Aerial Photo	1	11.6	10
11 Central Europe	Aletsch	2009-09-08	Aerial Photo	2017-08-29	Aerial Photo	25	120	61
11 Central Europe	Gorner	2007-09-13	Aerial Photo	2015-08-26	Aerial Photo	25	54.1	8
11 Central Europe	Rhone	2000-08-24	Aerial Photo	2007-09-12	Aerial Photo	25	16	6
11 Central Europe	Morteratsch	2008-09-09	Aerial Photo	2015-08-29	Aerial Photo	25	17	14
11 Central Europe	Unteraar	2003-07-14	Aerial Photo	2009-08-19	Aerial Photo	25	23.9	6
13 Central Asia	Abramov	2003-08-27	SPOT5-HRS	2015-09-01	Pleiades	40	114.7	121
14 South Asia West	Mera	2012-11-25	Pleiades	2018-10-28	Pleiades	4	33.6	44
14 South Asia West	Chhota Shigri	2005-09-21	SPOT5-HRG	2014-09-26	Pleiades	10	97.9	65
15 South Asia East	Gangotri	2004-11-26	SPOT5-HRG	2014-08-25	Pleiades	10	199.9	17

Table S2. High-resolution DEMs.

High-resolution DEM pairs used to validate glacier volume changes and their related uncertainty propagation. Only glaciers covered over more than 70% of their surface are considered. A minimum time interval of 5 years between the DEMs was chosen, corresponding to the length of the periods reported in our study.

RGI Region number	*Glacierized terrain (Stable terrain)*						
	ICESat points	IceBridge points	Elevation bias			Elevation change bias (m yr^{-1})	Standardized elevation uncertainty
			Amplitude (m)	Phase (decimal month)	Summer vertical bias (m)		
01	330,297 (658,370)	8,523,287 (463,310)	2.0 (0.1)	5.5 (3.5)	0.8 (1.1)	-0.046±0.033 (0.076±0.055)	0.51 (0.99)
02	30,011 (118,880)	0 (0)	1.8 (0.3)	5.9 (2.8)	2.2 (2.4)	0.035±0.216 (0.043±0.165)	0.54 (1.31)
03	1,139,069 (1,081,015)	4,833,055 (1,789,711)	1.8 (0.9)	6.6 (0.2)	2.1 (-0.6)	0.029±0.048 (0.007±0.104)	0.47 (1.01)
04	178,770 (405,729)	2,060,069 (1,381,223)	0.1 (0.1)	3.2 (1.8)	0.0 (0.5)	-0.001±0.014 (0.013±0.018)	0.32 (0.60)
05	878,141 (1,497,073)	3,150,014 (13,772,122)	0.4 (0.1)	6.3 (1.0)	0.0 (0.3)	0.008±0.014 (0.007±0.018)	0.47 (1.02)
06	27,480 (45,331)	0 (0)	1.3 (0.3)	4.3 (6.2)	1.0 (0.7)	0.023±0.071 (-0.041±0.034)	0.32 (0.77)
07	214,912 (135,141)	179,851 (18,227)	0.6 (0.4)	5.7 (5.5)	0.2 (0.4)	-0.026±0.039 (-0.046±0.037)	0.38 (0.63)
08	5,421 (10,736)	0 (0)	2.3 (0.4)	4.5 (4.3)	1.7 (1.6)	-0.092±0.160 (0.062±0.145)	0.55 (0.97)
09	383,224 (213,232)	0 (0)	0.2 (0.3)	6.0 (5.4)	-0.4 (0.1)	-0.015±0.022 (-0.013±0.014)	0.28 (0.45)
10	3,183 (306,043)	0 (0)	1.4 (0.3)	3.2 (6.0)	1.2 (0.8)	0.206±0.126 (-0.007±0.067)	0.60 (0.93)
11	1,891 (46,722)	0 (0)	1.2 (0.0)	6.2 (4.7)	2.1 (1.5)	0.052±0.263 (0.066±0.189)	0.67 (1.48)
12	929 (30,651)	0 (0)	1.7 (0.4)	6.0 (3.3)	1.3 (0.3)	-0.183±0.440 (-0.031±0.378)	0.69 (1.67)
13	105,218 (1,238,894)	0 (0)	0.5 (0.1)	3.5 (0.9)	1.2 (0.4)	0.110±0.090 (0.018±0.062)	0.68 (1.16)
14	54,382 (423,848)	0 (0)	0.7 (0.0)	3.2 (0.4)	1.4 (0.4)	0.102±0.083 (0.017±0.070)	0.64 (1.15)
15	19,961 (327,443)	0 (0)	0.2 (0.2)	4.3 (1.5)	1.1 (0.5)	0.122±0.106 (0.050±0.074)	0.63 (1.10)
16	897 (60,948)	0 (0)	0.0 (0.4)	2.7 (1.7)	0.3 (1.0)	0.045±0.500 (0.086±0.159)	0.58 (1.14)
17	21,157 (216,402)	109,339 (178,910)	2.5 (1.0)	11.8 (0.8)	2.0 (2.0)	-0.007±0.082 (0.010±0.215)	0.44 (1.29)
18	401 (13,597)	0 (0)	2.5 (0.8)	9.5 (1.2)	-0.4 (2.7)	-0.142±0.671 (0.167±0.685)	0.71 (1.47)
19	445,523 (114,116)	2,208,460 (3,997,523)	0.0 (0.9)	2.2 (3.0)	2.2 (4.4)	-0.004±0.048 (0.150±0.214)	0.33 (1.47)

Table S3. Validation of elevation time series with ICESat and IceBridge.

Elevation biases are decomposed in a sinusoidal function of amplitude, phase at maximum and vertical bias at the end of summer: mid-September for the northern hemisphere (decimal month: 8.5), mid-March for the southern hemisphere (decimal month: 2.5). Elevation change bias is the residual linear trend in time (weighted least squares) of seasonally de-biased mean elevation differences, shown with 95% CI. Standardized elevation uncertainty is the standard deviation of z-scores (conservative if less than 1).