Reanalysis of the US Geological Survey Benchmark Glaciers: long-term insight into climate forcing of glacier mass balance

Shad O’Neel1, Christopher McNeil1, Louis C. Sass1, Caitlyn Florentine2, Emily H. Baker1, Erich Peitzsch2, Daniel McGrath3, Andrew G. Fountain4 and Daniel Fagre2

1U.S. Geological Survey Alaska Science Center, Anchorage, AK, USA; 2U.S. Geological Survey Northern Rocky Mountain Science Center, West Glacier MT, USA; 3Department of Geosciences, Colorado State University, Fort Collins, CO, USA and 4Department of Geology, Portland State University, Portland, OR, USA

Abstract

Mountain glaciers integrate climate processes to provide an unmatched signal of regional climate forcing. However, extracting the climate signal via intercomparison of regional glacier mass-balance records can be problematic when methods for extrapolating and calibrating direct glaciological measurements are mixed or inconsistent. To address this problem, we reanalyzed and compared long-term mass-balance records from the US Geological Survey Benchmark Glaciers. These five glaciers span maritime and continental climate regimes of the western United States and Alaska. Each glacier exhibits cumulative mass loss since the mid-20th century, with average rates ranging from −0.38 to −0.30 m w.e. a⁻¹. We produced a set of solutions using different extrapolation and calibration methods to inform uncertainty estimates, which range from 0.22 to 0.44 m w.e. a⁻¹. Mass losses are primarily driven by increasing summer warming. Continentality exerts a stronger control on mass loss than latitude. Similar to elevation, topographic shading, snow redistribution and glacier surface features often exert important mass-balance controls. The reanalysis underscores the value of geodetic calibration to resolve mass-balance magnitude, as well as the irreplaceable value of direct measurements in contributing to the process-based understanding of glacier mass balance.

1. Introduction

Implications of glacier mass loss span global (Gardner and others, 2013; Marzeion and others, 2017; Zemp and others, 2019) to local scales (e.g. Moore, 1992; O’Neel and others, 2015; Sass and others, 2017; Schoen and others, 2017). Yet resolving regional patterns of mountain glacier mass loss remains challenging (e.g. Gardner and others, 2013; Box and others, 2018; Menounos and others, 2018), in part due to the paucity of long-term records, and in part due to the complex response of glaciers to climate forcing (Oerlemans, 2000; Roe, 2011). Modern remote-sensing capabilities enable rich analytical insight, but direct glaciological observations remain irreplaceable for capturing short-term (seasonal, interannual) changes and contributing to the process-based understanding of mass balance. However, comparisons drawn from these field-based records can be complicated by differences in data collection and analysis methods. For example, glacier-wide mass-balance estimates can be produced using multiple approaches to extrapolate point measurements to the entire glacier area, including mass-balance profiles (Fountain and Vecchia, 1999; Krimmel, 2000), an area-weighting (site-index) method (e.g. March and Trabant, 1996; van Beusekom and others, 2010), linear regressions between a single stake to robust observations from previous years (Krimmel, 1989), contour mapping (Meier and others, 1971), as well as inclusion of ad-hoc correction factors (e.g. van Beusekom and others, 2010). The magnitude of variance and resultant uncertainty from mixing these approaches has never been quantified and could be significant, highlighting the need for common intercomparison and interpretation frameworks for mass-balance datasets (Huss and others, 2009). Additionally, systematic reanalyses of mass-balance time series (e.g. Zemp and others, 2013) helps facilitate refinement of ongoing practices of long-term field programs by providing insight into biases introduced as technology evolves, and by linking individual steps of the analysis to overall estimates of uncertainty in mass-balance estimates.

The main objective of this study is to present a unified mass-balance reanalysis for the five US Geological Survey (USGS) Benchmark Glaciers (Gulkana, Wolverine, Lemon Creek, South Cascade and Sperry glaciers). To accomplish this, we first describe in detail a consistent, adaptable approach for estimating glacier-wide mass balance at annual and seasonal time steps from an inconsistent suite of observations. Our reanalysis includes an examination of several point-balance extrapolation and geodetic calibration approaches that capitalize on the increased availability and quality of aerial and space-borne imagery to minimize systematic errors in
field data. Finally, we use the reanalysis results to gain broad spatial and temporal insight into glacier response to climate forcing.

2. Background and setting

The concept of a continental scale investigation of North American glacier-climate linkages dates to the 1957 International Geophysical Year (IGY; American Geographical Society, 1960). USGS initiated a mass-balance program at Washington’s South Cascade Glacier as part of the IGY (Meier, 1958), which was the first step towards a long-term programmatic effort to document and understand glacier-climate linkages. During 1966, as part of the International Hydrological Decade (IHD), the project expanded to Gulkana and Wolverine glaciers in Alaska (Meier and others, 1971) and more recently to Montana’s Sperry Glacier in 2005 (Clark and others, 2017). We also include mass-balance data from southeast Alaska’s Lemon Creek Glacier, traditionally maintained by the Juneau Icefield Research Program (JIRP). This glacier fills the geographic gap between the glaciers of the contiguous USA and mainland Alaska, prompting USGS to initiate seasonal mass-balance observations that augment the ongoing JIRP’s summer-season efforts (Pelto and others, 2013). Logistics and scientific value played fundamental roles in site selection, with consideration of access, hazards and in some cases a history of previous research (Meier and others, 1971; Fountain and others, 1997). Representativeness of regional climate also influenced the selection process. The study glaciers span the primary North American climate regimes that support glaciers: midlatitude maritime, midlatitude continental, high-latitude maritime and high-latitude continental (Fig. 1a).

Gulkana Glacier (GG; Fig. 1b) is located in the high-latitude continental climate regime of Alaska’s Delta Mountains. USGS established an observation program here that included measuring snow and ice mass balance at three sites (nominally), a weather station (air temperature, precipitation, wind) installed on the glacier margin below the long-term equilibrium line (ELA) and a streamgauge on the outlet stream below the glacier. Two auxiliary weather stations were added to the network after 2010. The basin area above the USGS streamgauge is 31.5 km², and the glacier area decreased from 18.7 to 16.0 km² from 1967 to 2016 (Table 1).

Wolverine Glacier (WG; Fig. 1c) is found in the high-latitude maritime climate regime of Alaska’s Kenai Mountains. In 1966, simultaneous with GG, USGS established an observation program that included measuring snow and ice mass balance at three sites (nominally) (Mayo and others, 2004; O’Neel and others, 2014), a weather station (air temperature, precipitation, wind) installed on the glacier margin below the long-term equilibrium line (ELA) and a streamgauge on the outlet stream below the glacier. An additional weather station was added in 2012. The basin area above the USGS streamgauge is 24.6 km², and the glacier area decreased from 17.1 to 15.6 km² from 1969 to 2018 (Table 1).

Lemon Creek Glacier (LCG; Fig 1d) is located in the maritime high-latitude climate regime. Snow and ice mass-balance observations began in 1953, making LCG the longest record presented here (Miller and Pelto, 1999; Pelto and others, 2013). Extensive observations characterize the efforts through the first two decades of research, which transitioned into the modern five-site network following the IHD, ~1970. The early program included weather observations near the glacier, and a streamgauge on the outlet stream. However, in our analyses, frequent and extended discontinuities required using nonlocal meteorological data from the Juneau Airport. The airport is located 13 km from the glacier at sea level (Station ID: USW00025309; ncdc.noaa.gov). The basin area above the USGS streamgauge is 31.9 km² and the glacier area decreased from 12.8 to 9.7 km² between 1948 and 2018 (Table 1).

Research began at Washington State’s South Cascade Glacier (SCG; Fig. 1e) in 1958 (Meier, 1958). SCG represents the midlatitude maritime climate, characteristic of the North Cascade Mountains. The glacier program included weather measurements (air temperature, precipitation, intermittently maintained) and a streamgauge on the outlet stream of the lake in front of the glacier. A new, seasonal gauge was installed just below the glacier in 2002. Measurements strategies (e.g. stake networks, measurement dates) varied through time, but, on average, five seasonally measured sites form a continuous record of mass-balance observations. The basin area above the USGS streamgauge is 6.1 km² and the glacier area decreased from 2.9 to 1.8 km² between 1958 and 2015 (Table 1). SCG is the only Benchmark Glacier that calved icebergs for a portion of its record. The glacier terminated in a lake from ~1930–1975, but calving is thought to not have influenced mass balance significantly (Horlings, 2016).

Sperry Glacier (SG; Fig. 1f) is located along the Continental Divide within Glacier National Park, Montana, and represents the midlatitude continental or transitional climate. Seasonal mass-balance observations began in 2005 at six sites, complemented by historic photographic and topographic map records (Johnson, 1980) that provide longer-term context to the 14-year field program. A weather station installed on the glacier’s western margin during 2006 records temperature, humidity, wind speed, wind direction and solar radiation. The glacier resides in a ~4 km² hanging cirque, that contains several small, ungaugeable outlet streams that pour over cliffs at the cirque boundary. As of 2015, the glacier had an area of 0.80 km² (Fagre and others, 2017), only slightly reduced from the 1950 area of 1.3 km², making SG the smallest of the benchmark glaciers (Table 1).

3. Data

Glacier-wide mass-balance estimates require several classes of data, including direct glaciological measurements (accumulation and ablation), meteorological data (temperature and precipitation) and descriptions of the glacier hypsometry (area, area-elevation distribution), which are derived from the air- and space-borne imagery. The comprehensive dataset used in the reanalysis is available for download (Baker and others, 2018).

3.1 Point mass-balance data

A primary emphasis of this reanalysis involved consistent aggregation of ablation stake and snowpit data to estimate point mass balances. All possible data were incorporated into surface height change estimates before they were scaled by material (snow, firn, ice) density to obtain a mass balance (m w.e. a −1). We prescribe an ice density of 900 kg m⁻³. For snow pits or cores, density is measured through the entire snowpack, from which a column-average density is calculated. Stakes are used to estimate ablation in a stratigraphic, floating-date time system by measuring the snow or ice level relative to the presumably fixed stake (Cogley and others, 2011). Measurement sites were established at fixed locations (index sites) early in each project. Stakes are installed upstream of, or at index sites, then replaced seasonally (SCG, SG), biennially (GG, WG, LCG) or as needed to prevent meltout or flow into a different mass-balance regime. When the glacier surfaces are snow-covered, depth measurements to ice or the previous summer surface (firm) are performed in snow pits or by extracting a core if snowpack exceeds ~2 m. Identification of the previous summer surface can be challenging in cores, which intermittently led to this surface being marked with sawdust, colored chalk or hardware cloth to facilitate interpretation. Probe observations often augment snow pits to better capture spatial variability of seasonal snow depths.

Field visits are scheduled to approximately coincide with the mass extrema, though our reanalysis approach (Section 4.1)
accounts for mass changes that may occur between measurement
date and seasonal mass extrema. To minimize the magnitude of
these modeled adjustments, we measure winter ablation that
occurs after fall field visits (measured the following spring) and
summer accumulation (measured during fall visits) to ensure
that these components are distributed to the appropriate balance
years. Reconstructing these mass-balance components was not
always possible in the early record when mass-balance practices
were still being refined.

Measurement networks and protocols vary among the glaciers
and in time commensurate with changes in staff, funding and the
ever-evolving knowledge base. In all cases, measurement sites
were located in areas thought to be representative of broad regions
of the glacier (Fig. 1). Except for SG, original site selections pre-
date all current project staff.

Three fixed sites characterized the WG and GG programs over
most of the record, distributed with the intention to represent the
mass-balance regime of the ablation zone, the equilibrium line
altitude (ELA) region and the accumulation zone. At SCG and
LCG, site consistency was not emphasized over much of the
record; rather, measurements occurred in general regions (typic-
ally, 5–10 sites). JIRP assembled the majority of the LCG record,
which is characterized by a single late-summer, accumulation
dominated set of annual observations (Pelto and others, 2013;
McNeil, 2016). Data from LCG and SCG required additional
efforts to prepare input for standard processing (Supplementary
Material). At SG, the influence of wind and avalanching was
recognized early on, leading to a stake network that spans both
the width and length of the glacier.

Upgrades to observation networks occurred throughout the
project history. Since the late 1990s, SCG sites were distributed
at quasi-regular elevation intervals along the glacier centerline.
Network size increased two- to threefold at GG and WG after
2010 (Fig. 1). Seasonal measurements began at LCG in 2013.
The number of snow pits used to determine accumulation rates
and gradients remained variable among the glaciers, ranging
from one annually on SG to five on LCG. New index sites have
been periodically established to accommodate retreat beyond the
lowest elevation sites at many of the glaciers. In accord with the
concepts of the USGS Benchmark Glaciers, extensive ablation
stake networks were used early on to determine sparse-network
site representativeness (e.g. Tangborn and others, 1975), but
these efforts were later halted, and only recently reinitiated.

3.2 Meteorological data

For WG and GG, meteorological stations installed within the
watershed (Fig. 1) provide daily estimates of average air tempera-
ture and total precipitation catch, but discontinuous local records
for SCG, LCG and SG required the use of continuously operating

![Fig. 1.](image-url)
distant weather stations (Wheler and others, 2014). We chose to use the local stations at GG and WG, as relatively few data gaps exist (Table S1), and regional lapse rates are poorly defined. At these two stations, we fill short (<3 day) gaps using linear interpolation and long gaps via monthly regressions with data from distant stations, typically at lower elevations (van Beusekom and others, 2010). Further details of the USGS-maintained instrumentation, changes to instrumentation and quality control are provided in the Supplementary Material.

### 3.3 Geospatial data

A suite of 33 DEMs provide geometric and geodetic constraints on the mass-balance reanalysis (Tables S2–S6). Through ~2008, stereo aerial photography of SCG, LCG, WG and GG was episodically acquired as time and budgets allowed. These historic data were reprocessed using Structure from Motion (SfM) photogrammetric methods to produce higher-accuracy DEMs (e.g. Westoby and others, 2012). Since 2008, airborne imaging has largely been replaced by optical commercial satellite imagery from the DigitalGlobe constellation (Neigh and others, 2013). This sub-meter resolution imagery along with automated processing engines has streamlined the DEM production process (e.g. AMES stereo pipeline; Shean and others, 2016). Exceptions to optical acquisitions include the 2000 (SRTM; https://earthexplorer.usgs.gov/) and 2013 (IfSAR; http://ifsar.gina.alaska.edu) LCG DEMs, which were produced from synthetic aperture radar products, and the 2016 WG DEM, which was produced from airborne lidar.

### Table 1. Geodetic analyses data including acquisition year, mass balance, glacier area (km²) and fractional area change relative to the reference DEM

<table>
<thead>
<tr>
<th>Glacier</th>
<th>Year</th>
<th>Geodetic mass balance (m w.e.)</th>
<th>Uncertainty (m w.e.)</th>
<th>Glacier area (km²)</th>
<th>Area change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gulkana</td>
<td>1967</td>
<td>−26.6</td>
<td>2.80</td>
<td>18.7</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>1974</td>
<td>−23.6</td>
<td>2.49</td>
<td>18.4</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>1979</td>
<td>−19.2</td>
<td>1.86</td>
<td>18.1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>−16.9</td>
<td>2.52</td>
<td>18.0</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>2005</td>
<td>−6.9</td>
<td>1.28</td>
<td>16.9</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>−5.9</td>
<td>0.60</td>
<td>16.9</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2016</td>
<td>0</td>
<td>0</td>
<td>16.0</td>
<td>0</td>
</tr>
<tr>
<td>Wolverine</td>
<td>1969</td>
<td>−17.2</td>
<td>2.86</td>
<td>17.1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1972</td>
<td>−18.0</td>
<td>2.24</td>
<td>17.1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1979</td>
<td>−17.1</td>
<td>3.43</td>
<td>17.0</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>−14.9</td>
<td>1.72</td>
<td>16.8</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td>−8.7</td>
<td>2.14</td>
<td>16.3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2016</td>
<td>−2.4</td>
<td>0.24</td>
<td>15.9</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2018</td>
<td>0</td>
<td>0</td>
<td>15.6</td>
<td>0</td>
</tr>
<tr>
<td>Lemon Creek</td>
<td>1948</td>
<td>−49.6</td>
<td>11.83</td>
<td>12.8</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>1957</td>
<td>−39.9</td>
<td>3.68</td>
<td>12.4</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>1979</td>
<td>−36.7</td>
<td>3.03</td>
<td>12.1</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>1989</td>
<td>−32.5</td>
<td>2.67</td>
<td>11.7</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>1999</td>
<td>−23.4</td>
<td>2.33</td>
<td>10.7</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>−11.2</td>
<td>1.74</td>
<td>10.4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2016</td>
<td>−5.3</td>
<td>1.11</td>
<td>10.1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2018</td>
<td>0</td>
<td>0</td>
<td>9.7</td>
<td>0</td>
</tr>
<tr>
<td>South Cascade</td>
<td>1958</td>
<td>−33.1</td>
<td>3.50</td>
<td>2.9</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>1970</td>
<td>−26.1</td>
<td>1.90</td>
<td>2.7</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>1979</td>
<td>−25.8</td>
<td>1.90</td>
<td>2.7</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>1986</td>
<td>−23.2</td>
<td>2.06</td>
<td>2.6</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>1992</td>
<td>−16.4</td>
<td>1.23</td>
<td>2.3</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>2001</td>
<td>−11.7</td>
<td>1.01</td>
<td>2.2</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>2004</td>
<td>−8.1</td>
<td>1.06</td>
<td>2.1</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>−5.0</td>
<td>0.41</td>
<td>2.0</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>2015</td>
<td>0</td>
<td>0</td>
<td>1.8</td>
<td>0</td>
</tr>
<tr>
<td>Sperry</td>
<td>2005</td>
<td>−0.9</td>
<td>1.59</td>
<td>0.86</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>2014</td>
<td>0</td>
<td>0</td>
<td>0.81</td>
<td>0</td>
</tr>
</tbody>
</table>

The reference DEM is indicated with bold, italicized text.
orthophotographs to obtain total glacier area. The elevation distribution was then assembled by partitioning glacier surface elevations (from available DEMs) into elevation bins commensurate with the total elevation range of each glacier (SG: 30 m, SCG: 50 m; LCG, GG, WG: 100 m). Because the DEMs exist on approximately decadal time steps, we linearly interpolate the changes in the area within each bin between subsequent DEMs to produce annual hypsometries.

4. Analysis
To improve confidence in and reproducibility of mass-balance estimates, as well as to allow direct intercomparison of otherwise disparate glacier records, we developed a flexible yet consistent reanalysis approach (Fig. 2). The approach provides the capacity to evaluate glacier-wide seasonal and annual surface mass balance in a variety of time systems, reference frames, extrapolation types and geodetic calibration styles (Supplementary Material). Common to other mass-balance programs, this approach centers on fitting a mass-balance profile to point observations (Fig. 3), then integrating (extrapolating) the profile over the coincident glacier hypsometry to estimate seasonal and annual mass balances. The cumulative glacier-wide mass balance is calculated, then compared with the cumulative geodetic mass balance. If the two series diverge beyond geodetic uncertainties, a calibration is performed. We use the word calibration to describe a least-squares adjustment of the in situ glaciological data to minimize the difference with geodetic data (Zemp and others, 2013). Many of the analysis steps require the use of a mass-balance model (Section 4.2) to resolve changes in mass among observations (e.g. direct measurements and image acquisitions) or to coincide with the floating-date time system.

4.1 Estimating glacier-wide mass balance
Mass-balance profiles (e.g. Fountain and Vecchia, 1999; Cogley and others, 2011) provide an elevation-based parameterization of point observations \( (b(z)) \), through which glacier-wide mass balance \( (B) \) is calculated. For each profile fit, we required a minimum of three distributed point balance observations: one in the ablation zone, another in the vicinity of the ELA and a third in the accumulation zone. We explore sensitivity to various interpolation and extrapolation schemes (Section 4.5) by producing a solution set with multiple forms of balance profiles (e.g. linear, piecewise-linear) and the site-index method historically used at GG and WG (Fig. S5). Ultimately, we use a piecewise-linear balance profile for interpretation, and the other methods to constrain uncertainties arising from the balance profile. This piecewise approach can be applied to almost all seasonal and annual datasets, allows variable profile slopes in the accumulation and ablation zones, and collapses to a linear profile under many sparse data cases. Extenuating circumstances occasionally prohibited field data collection at more than two sites requiring infilling as described in the Supplementary Material.

Resolving glacier-wide mass balance in floating-date time systems requires estimates to be coincident with seasonal glacier mass extrema (Cogley and others, 2011). The first step in resolving extrema is modeling point balance time series at the highest and lowest elevation sites to bracket the timing of the glacier-wide extrema. Then daily glacier-wide solutions through the bracketed and lowest elevation sites to bracket the timing of the glacier-wide extrema. Then daily glacier-wide solutions through the bracketed and lowest elevation sites to bracket the timing of the glacier-wide extrema is modeling point balance time series at the highest mass extrema (Cogley and others, 2011). The first step in resolving extrema is modeling point balance time series at the highest mass extrema (Cogley and others, 2011). The first step in resolving extrema is modeling point balance time series at the highest mass extrema (Cogley and others, 2011). The first step in resolving extrema is modeling point balance time series at the highest mass extrema (Cogley and others, 2011). The first step in resolving extrema is modeling point balance time series at the highest mass extrema (Cogley and others, 2011). The first step in resolving extrema is modeling point balance time series at the highest mass extrema (Cogley and others, 2011). The first step in resolving extrema is modeling point balance time series at the highest mass extrema (Cogley and others, 2011).

4.2 Mass-balance model
Many time systems require resolving the timing mismatch among observations (e.g. stakes, DEMs) using a mass-balance model (e.g. van Beusekom and others, 2010). Model adjustments are generally applied over short-time intervals (i.e. days–weeks). We estimated daily point mass balance \( (b) \) as the sum of ablation \( (a) \) and accumulation \( (c) \):

\[
b = a + c.
\]

Ablation was modeled, using the positive degree-day approach, as:

\[
a = T_{(T > 0)} k_{s,i}
\]

where \( T_{(T > 0)} \) is a positive air temperature estimate for each measurement site on the glacier and \( k_{s,i} \) is an ablation coefficient (degree-day factor) for snow \( (s) \) or ice \( (i) \) determined for each glacier. Site temperatures were calculated from weather station observations using a constant adiabatic lapse rate of \(-6.5 \degree C \text{km}^{-1}\) (Barry, 1992) to account for elevation differences. Ablation coefficients \( (k_{s,i}) \) were solved for each glacier using all data over the entire study period (Table S8) and show general agreement with literature values (e.g. Braithwaite, 1995).

Snow accumulation was estimated as:

\[
c = P_{(T < 1.7)} \gamma
\]

where \( P \) is recorded precipitation when the temperature was lower than 1.7 \degree C and \( \gamma \) a scale factor between the total winter snow accumulation at a mass-balance site and the respective weather station (van Beusekom and others, 2010). For routinely measured sites, site-specific scale factors \( (\gamma) \) were determined, but for sites measured fewer than three times, we applied a glacier-wide average scale factor (Table S9).

4.3 Geodetic mass balance and calibration
Geodetic mass-balance time series were constructed from DEM data (Table 1) to assess and reduce (through calibration) systematic biases in the glaciological cumulative balance (Zemp and others, 2013). The combination of abundant high-quality space-borne imagery and rapid changes to the glaciers prompted us to explore several time-windowed approaches.

4.3.1 Geodetic mass balance
For each glacier, DEMs were produced in a common geospatial reference frame (WGS84) and projection (UTM), and then aligned using a universal co-registration method (Nuth and Kääb, 2011). The co-registration minimizes apparent elevation changes across stable (bedrock), snow-free, off-glacier terrain. The highest quality DEM in each time series is designated as the ‘reference,’ with which all other DEMs were aligned (Table 1). Once co-registered, DEMs were differenced from the reference DEM to yield elevation change across the glacier surface (Shean and others, 2016; Sass and others, 2017). Voxels in the difference map exceeding 5% of the glacier surface were filled using a regression between the change in surface elevation and the reference surface elevation (Larsen and others, 2015). Glacier volume change from the DEM differencing was converted to mass change by specifying a material density of \( 850 \pm 60 \text{ kg m}^{-3} \), which accounts for the unknown changes in firm density and volume (Huss, 2013). We do not explicitly account for internal accumulation or ablation (Oerlemans, 2013; Andreassen and others, 2016), but given representative precipitation rates and glacier elevation ranges we can assume none of the glaciers is subject to significant internal accumulation. We also estimated
the physical limit of internal melting from dissipative heat and found it to range only between 0.01 and 0.06 m w.e. a\(^{-1}\) across the study glaciers.

4.3.2 Geodetic calibration

As the first step in geodetic calibration, we rejected DEMs with >60% snow-covered area to minimize uncertainty related to poor constraints on the density of seasonal snow. Next, we addressed the temporal mismatch between field campaigns and image acquisition, as acquisition rarely coincides with the timing of mass minima (Tables S2–S6). Using the mass-balance model, we adjust the geodetic mass-balance results to account for mass gains or losses between the DEM dates and the timing of mass extrema. These adjustments were on average 25 ± 18 days which amounted to a −0.45 ± 0.51 m w.e. average adjustment on the cumulative geodetic mass balance. Once both datasets were migrated to a common time, calibration of the direct time series occurred.

The most commonly adopted calibration approach involves applying a sequential adjustment to the direct time series for each DEM pair in the geodetic time series (Fig. S7; Zemp and others, 2013). Sequential calibrations minimize the residual difference of cumulative direct and geodetic mass-balance time series for each geodetic interval independently. This approach may be appropriate when a limited number of DEMs exist (e.g. decadal or longer time spans between acquisitions), and offers the advantage of resolving permanent solutions over the majority of the record. However, when differencing over short-time intervals (<5 years), changes in firn structure and the associated density assumption can introduce substantial aliasing errors in the geodetic mass-balance time series.

Alternatively, a global calibration can be developed, with a single weighted least-squares fit over the entire time series (e.g. van Beusekom and others, 2010; O’Neel and others, 2014). Weights were derived from the uncertainties of each DEM. The slope of the least-squares fit describes the annualized rate of systematic
bias, which was then applied to calibrate the direct time series (Fig. S7). This global approach has the benefit of fitting through the maximum number of data points, but the distinct disadvantage that it assumes a uniform bias in the direct method over the entire period of record. A constant bias may exist for constant measurement networks, small changes in glacier geometry or under steady climate conditions, but is highly unlikely to apply uniformly through periods of substantial glacier change. Furthermore, it results in changes (generally small) to the entire calibrated time series each time a new DEM is produced, which may not be warranted.

We developed a third, hybrid approach, where a piecewise-linear calibration is fit over ~10-year intervals (longer if necessary), each of which must contain at least three DEMs (Fig. S7). This breakpoint approach maximizes the incorporation of geometric data while minimizing errors that arise from simple density assumptions when there is substantial interannual variability in glacier mass balance (Huss, 2013). It allows calibration intervals to coincide with changes in glacier hypsometry, measurement networks and/or climate. Regardless of the calibration approach used, annual calibrations are distributed equally between the accumulation and ablation seasons.

4.4 Variability and trend analyses

Mass balance and meteorological time series were analyzed for correlations, variability and the presence of trends using parametric (c.f. Medwedeff and Roe, 2017) and nonparametric methods (e.g. O’Neil and others, 2014). Our parametric approach involved forming linearly – detrended anomalies, evaluating the normality of residuals, then using decade-normalized t-statistics (after Roe, 2011) to evaluate and intercompare the presence of significant trends. To improve confidence in our statistics, a similar process was followed using the nonparametric Mann-Kendall test (Hamed and Ramachandra Rao, 1998) on anomaly time series.

4.5 Sensitivity and uncertainty

Traditional error assessment is commonly done from the bottom up: investigators begin with point balance errors and arrive at error estimates for glacier-wide balance (e.g. Cogley and Adams, 1998; Zemp and others, 2013; Beedle and others, 2014). However, this approach requires assumptions about the relation between point measurements and the glacier-wide mass balance that are largely untestable, notably that point measurements adequately represent mass-balance variability over the entire glacier.

We endeavor to avoid those assumptions using a top-down approach that partitions the total uncertainty in glacier-wide mass balance into two components. The first component reflects uncertainty in the temporal trend (i.e. cumulative mass balance), and the second component relates to annual balance uncertainty. Trend uncertainty derives from the geometric balance, while uncertainty in annual balance follows from quantifying the magnitude of variance from the different extrapolation and calibration methods. The two terms can be combined to define characteristic uncertainties for the mass balance at each glacier (Thibert and Vincent, 2009).

4.5.1 Geodetic mass-balance uncertainty

Geodetic mass-balance uncertainty arises from errors in computing glacier volume change, and errors in the material density assignment required to convert volume to mass. Volume change error has two sources: elevation error and elevation data gaps.

Elevation error is characterized by off-glacier elevation differences remaining after co-registration. In all cases, spatial autocorrelation areas were smaller than the glaciers, and in some cases the spatial autocorrelation area may be as small as an artifact of the total DEM area, which leads to a potential underestimate of uncertainty. Hence, we use the more conservative normalized median absolute deviation \( N_p \), as a robust estimate of variance in a non-Gaussian distribution,

\[
N_p = 1.4826 \cdot \text{median}(|\Delta z - m_{\Delta z}|),
\]

where \( \Delta z \) are the remaining off-glacier elevation differences between the aligned DEMs, \( m_{\Delta z} \) is the median of the \( \Delta z \) values and the coefficient scales the rank-based statistic to provide a robust estimate of variance compared to the standard deviation (Hölle and Höhle, 2009; Shean and others, 2016).

Error from infilling elevation data gaps is nonzero when elevation data are missing over >5% of the glacier surface (Tables S2–S6). Gaps are infilled using a linear interpolation model across the glacier surface (Larsen and others, 2015). We evaluate uncertainty due to gap interpolation with the mean absolute error statistic \( M \), given by,

\[
M = \sum |r| / n,
\]

where \( r \) is the residual between observed and infilled surface elevation changes, and \( n \) is the number of measured values over the glacier surface.

We combined these two elevation uncertainty components using the fraction of the glacier surface that required interpolation \( f \) and the fraction that did not need interpolation \((1 - f)\) to yield \( \sigma_{\Delta z} \), the total elevation uncertainty, given by,

\[
\sigma_{\Delta z} = N_p (1 - f) + (M + N_p) f.
\]

Finally, we take the total geodetic mass-balance error \( \sigma_g \) (m w.e.) as the quadratic sum of the volume change term and a density term (Beedle and others, 2014),

\[
\sigma_g = \sqrt{(\sigma_{\Delta z} \gamma)^2 + (dh \lambda_s)^2},
\]

where \( dh \) is the area-averaged surface elevation change (m), and \( \gamma \) and \( \lambda_s \) are unitless factors to convert length to water equivalent units. \( \gamma \) takes a value of 0.85 (i.e. 850 kg m\(^{-3}\)/1000 kg m\(^{-3}\)) to convert \( \sigma_{\Delta z} \), and \( \lambda_s \) takes a value of 0.06, i.e. 60 kg m\(^{-3}\)/1000 kg m\(^{-3}\) (Huss, 2013). \( \sigma_g \) represents the total uncertainty over the geodetic differencing interval, but can be expressed in terms of annual error with units m w.e. a\(^{-1}\).

4.5.2 Sensitivity tests

Evaluating different methods of point measurement extrapolation and geodetic calibration offers insight into how the number and distribution of point measurements affect the resolution of glacier-wide mass-balance estimates. We characterize the sensitivity of our methodology by producing a set of nine mass-balance time series, each of which has a different spatial-extrapolation and/or geodetic-calibration approach. The solution set consists of the historically used site-index method (March and Trabant, 1996) and two profile fitting variations (linear and piece-wise linear). Each of these three variants is further subjected to three geodetic calibration approaches (sequential, breakpoint, global; Fig. S7). To characterize uncertainty, we calculated the normalized median absolute deviation \( N_p \), Eqn (4) substituting \( DB \) for \( \Delta z \) across the solution set, with the deviation \( DB \) defined as the difference between any given estimate and the preferred solution. The combined characteristic uncertainty is then estimated as the root sum of squares of the trend and annual balance components for each glacier. Several other parameters such as lapse rates or degree-day factors were explored separately and found to have a much smaller influence.
on solutions (e.g. adjusting the lapse rate by 1 °C only impacted solutions by \(10^{-2}\) ma\(^{-1}\), whereas the differences between geodetic calibrations are \(10^{-1}\) to 100 ma\(^{-1}\)).

5. Results

We designed this reanalysis with the goal of comparing results with climate and constraining uncertainties from variations in methods. The approach yielded a suite of solutions from which we quantify the magnitude of variance arising from various extrapolation and calibration methods (i.e. sensitivity testing). We selected a preferred solution in a floating-date (as opposed to a fixed-date system; Cogley and others, 2011) ‘stratigraphic’ time system designed to quantify the duration of the accumulation and ablation seasons and to keep mass changes from any given season within a single mass-balance year. It uses
time-variable glacier hypsometries, yielding conventional mass balances (Elsberg and others, 2001; Cogley and others, 2011). The solution incorporates a piecewise-linear mass-balance profile, and is geodetically calibrated using the breakpoint approach.

### 5.1 Mass Balance

Mass loss at the benchmark glaciers over the study interval is ubiquitous (Fig. 4). Average mass loss rates range from $-0.28$ to $-0.58$ m w.e. a$^{-1}$, with uncertainty ranging between $0.22$ and 0.42 m w.e. a$^{-1}$ (Table 2). Averaged over their entire records, LCG exhibits the highest rate of mass loss and SG the lowest. Average rates calculated before and after 1990 (approximate midpoint) reveal that the rate of mass loss increased in all cases. Continental GG lost mass continuously through the study, whereas the maritime glaciers were initially closer to equilibrium conditions (Fig. 4). Figure 5 shows that annual mass balances are predominantly negative at all glaciers, especially in the latter half of the records. Mass turnover, as assessed via the average of annual balances (low turnover, 1.2 m w.e. a$^{-1}$ for GG) and maritime classes (high turnover, e.g. 3.1 m w.e. a$^{-1}$ for LCG). Annual balance variability, as quantified by the standard deviation (Fig. S8), ranges between 0.6 and 1.1 m w.e. a$^{-1}$, and is lowest at LCG, and highest at WG, both of which are maritime climate settings. No progression in variability between the continental and maritime exists. Summer balance variability is equal to or exceeds winter variability except at WG, where winter balance variability exceeds summer.

Annual balances exhibit field-significant ($\alpha = 0.10$, or 90% confidence) negative trends for all but SG, and trends at GG and LC are 95% confident (Fig. 6a and Table S11). Winter balance trends exhibit variable slopes and sign, and none are significant (Fig. 6b). Statistically significant trends in summer balance exist at GG, WG and SCG (Fig. 6c). Temperature data used in this analysis show long-term summer and annual-averaged warming trends (Fig. 7), offering a mechanistic explanation for mass-ter loss rates. Total reflects the characteristic uncertainty for each glacier through the combination of the two components.

### 5.2 Sensitivity and uncertainty

Total uncertainty consists of a component related to geodetic mass balance and a component related to the direct glaciological mass balance. The uncertainty in the geodetic calibrations varies between $0.1$ and $0.2$ m w.e. a$^{-1}$, and this term is dominated by time-system adjustments (which average 25 days). Uncertainty in the direct approach is estimated using $N_g$, the total uncertainty, which ranges between 0.1 and 0.4 m w.e. a$^{-1}$. This metric reflects solution sensitivity to the various combinations of the extrapolation and calibration methods. Total uncertainty estimates average $-0.2$ m w.e. a$^{-1}$, except at GG, which has a higher uncertainty of 0.44 m w.e. a$^{-1}$ (Table 2).

### 6. Discussion

Long-duration glacier mass-balance records offer insight into long-term changes in climate from forcing and response perspectives (Pepin and others, 2015; Medwedeff and Roe, 2017). They provide valuable observational constraints for climate, hydrological and Earth system models (Beamer and others, 2016; Huss and Hock, 2018). However, loosely standardized computation and/or analysis methods impede interpretations of the often-complex relationship between glaciers and climate. Advances in understanding of mass-balance processes, technology and numerical methods improve our capability to assess glacier mass balance, but also require careful and consistent application to avoid introducing further bias and/or scatter within or among historical records. To address these problems, we reanalyzed USGS Benchmark Glacier datasets, with attention to tightened temporal consistency, solution sensitivity and comparability. Our work reinforces the necessity of geodetic calibration (e.g. Zemp and others, 2013), illustrates the impact of elevation-independent mass-balance drivers and offers insight into broad spatial and temporal patterns of glacier change.

#### 6.1 Glacier-wide solutions

Traditionally, the primary goal of glacier mass-balance programs is to reduce multiple forms of measurement into a single number representing glacier-wide average mass balance. This simple output facilitates subsequent use of the data (e.g. by the sea-level rise...
or decision-making communities), but tends to mask key assumptions and uncertainties required during formulation. This reanalysis was designed to provide exact, data-driven understanding of these aspects of mass balance.

Figure 9a shows that uncalibrated solutions vary broadly at GG, and diverge from geodetic balances in excess of uncertainty, demonstrating the need for calibration prior to evaluating trends. The apparent ability of the site-index method to outperform the profile approach is largely an artifact of the maximally sparse stake networks that were used over the majority of the time series and the way this approach caps (instead of extrapolates) the mass-balance rate at the upper measurement site (Fig. 9a). Annual balance solution variants (Fig. 9b) show similar patterns of interannual variability, which do not appreciably influence the goodness-of-fit in the calibration process (Fig. 9c). Results for the other benchmark glaciers are provided in Figs S9–S12.

Given the lack of observational constraints over the upper regions of balance profiles and the greatly reduced uncertainty of historic DEMs following SfM re-processing, we chose not to adjust (beyond fitting) balance profiles (e.g. clipping of the upper tail) to improve agreement with geodetic results. Rather, we apply systematic fits to point balance data and allow the geodetic calibration process to resolve biases between the time series (Cox and March, 2004; Thibert and Vincent, 2009).

Although the increased availability of high-resolution geodetic data fuels a desire to difference DEMs over shorter and shorter
intervals, confidence in geodetic mass balance decreases proportionally to the differencing interval. Errors arising from co-registration relate solely to the quality of DEM alignment but are amortized over the difference interval, such that long intervals minimize uncertainty. Errors arising from the volume-to-mass conversion (i.e. glacier thinning/thickening) relate to limits in our understanding of near-surface material density variability across the glacier. Material density is commonly specified as a constant \(850 \pm 60 \text{ kg m}^{-3}\) (Huss, 2013), thus this uncertainty scales with volume change. Realistically, this uncertainty should scale with the fraction of the observed volume change attributable to firn or snow. In many cases this fraction will be inversely proportional to the DEM difference interval, suggesting that over short intervals, this formulation is likely to underestimate true uncertainty (e.g. Klug and others, 2018). In contrast, material density is better constrained for large volume changes (generally longer intervals), and the uncertainty is likely to be overestimated (Huss, 2013). Improved understanding of temperate glacier firnification processes is essential to constrain volume conversion to mass, especially as differencing intervals continue to shorten.

Until a comprehensive firn-densification model is developed, we suggest that these errors can be reduced using fitted calibrations applied over optimal (decadal) time intervals. For example, given the nominal case of a 50-year mass-balance record with 10 DEMs acquired more frequently in the last 20 years, a global calibration

---

**Fig. 7.** Meteorological trends at weather stations used in the reanalysis. (a) Mean annual temperature (°C). (b) Total winter (November–March) precipitation (mm) and (c) ablation season (May–September) temperature (°C). Data representing South Cascade Glacier are from Diablo Dam weather station; Lemon Creek Glacier is represented by Juneau Airport; Sperry Glacier by Flattop Snotel. Solid lines are indicative of significant trends, as evaluated with the modified Mann-Kendall test.
distributes a single correction uniformly across the time series, regardless if measurement networks, balance rates or glacier geometry undergo significant change. Sequential calibration represents an opposite end-member, distributing corrections stepwise for every available geodetic mass balance, where large density errors may alias geodetic balances over the recent sub-decadal differencing intervals (Fig. 9d and Fig. S10d). The breakpoint calibration offers a compromise to these two end-member approaches, maximizing the incorporation of geodetic data, while also minimizing impacts of the evolving relationship between glaciological and geodetic balances.

All of the USGS Benchmark Glaciers required geodetic calibration, which resulted in small changes to recently published values (e.g. O’Neel and others, 2014; Le Bris and Paul, 2015; Clark and others, 2017; WGMS, 2017; Florentine and others, 2018). At four of the five glaciers in this study, geodetic balances are more negative than their direct glaciological counterparts (Fig. 9 and Figs S9–S12), suggesting a consistent tendency for direct observation networks to underestimate ablation (or overestimate accumulation) across the suite of glaciers. These results strongly support the concept that uncalibrated mass-balance solutions will possess bias unless an independent geodetic assessment shows otherwise (Krimmel, 1999; Cox and March, 2004; Huss and others, 2009; Zemp and others, 2013).

The mismatch between direct and geodetic time series is common but remains poorly understood. Perhaps the most commonly suggested driver for the mismatch is internal or basal mass balances (Cogley and Adams, 1998; Zemp and others, 2013; Andreassen and others, 2016). However, at the benchmark glaciers, estimates for these processes are <0.06 m w.e. a$^{-1}$, and constitute only ~15% of the average calibration (Table S10), suggesting alternative processes drive the mismatch.

### 6.2 Elevation dependence

The balance profile method has become standard for extrapolating point observations over the glacier area (Beedle and others, 2014; Barandun and others, 2015; WGMS, 2017). The central tenet of this method is that elevation provides a higher-order control than local effects such as snow redistribution, topographic shading or surface features in glacier-wide mass-balance estimates. Local effects are unlikely to be captured by logistically feasible measurement networks, and are often assumed to be negligible. However, the frequent divergence between geodetic and direct mass balances suggests that important processes elude the elevation-based parameterization.

At WG, ∼20 m amplitude drift and scour patterns characterize ∼20% of the accumulation zone (Fig. 10a; McGrath and others, 2018). Safety and logistical practicalities prohibit the routine collection of a distributed set of measurements from this region, precluding an estimate of average mass within this region of extreme local variability. Sub-annual DEM analyses on glaciers may help to reveal and resolve the complex spatial patterns in snow accumulation, but require extensive, typically unavailable, data to constrain sub-annual changes in surface elevation and its partitioning into components driven by ice flow, changes in material density and mass-balance processes (Sold and others, 2013).

At SG, avalanching constitutes a primary snow accumulation process, yet avalanche input is irregularly recorded by the measurement network, which can alter winter balance estimates in both directions (Clark and others, 2017). The steep headwalls serving as avalanche source regions during winter also provide topographic shading during the ablation season, resulting in an inverted summer balance profile and a larger mass turnover rate than expected from regional climate forcing (Figs 3, 5 and Fig. S6). Roughly 80% of the local forcing can be attributed to augmentation of the accumulation budget and 20% to the ablation budget by shading (Florentine and others, 2018). As shown in Figure 10b, SCG displays similar local controls; spatial distributions of firn align with shading patterns along the glacier margin rather than with elevation. Finally, crevasse fields and icefells increase surface roughness and area-volume ratios, and in turn melt rates via changes to turbulent and latent heat fluxes (Colgan and others, 2016). One such example is found in the icefall between sites AU and N at WG (Fig. 1c) where ablation rates measured within a 15 m elevation range differ by more than 3 m w.e. (Fig. S5). Although hazards present in icefells tend to dissuade direct observations, the enhanced melt occurring in such areas may be essential to include in glacier-wide balance estimates.

Measurement networks rarely span the entire elevation range of a glacier, which can markedly influence the slope of the balance profile (Fig. 3 and Fig. S4), and may consequently explain a substantial fraction of the divergence between direct and geodetic cumulative balances. For example, Figure 9a shows the 1967–2018 uncalibrated cumulative balance at GG reaches +20 m w.e. whereas the calibrated value is ∼30 m w.e. Efforts can be made to constrain the profile, either by imposing limits on mass-balance rate (e.g. Sold and others, 2016), or by using short-term (1–3 years) expanded measurement networks (i.e. stakes, pits, GPR, DEMs) to sample logistically challenging regions of the glacier. However, incorporation of short-term results into long-term programmatic efforts can introduce additional bias over earlier intervals of the record, instead of reducing overall modern biases. Expanded measurement networks help to understand processes driving mass-balance variability and trends (e.g. McGrath and others, 2018), which can subsequently be used to adjust measurement networks, especially if calibration breakpoints are established at a similar time (e.g. GG, 1974; Fig. 9b). Periods of overlap between old and new measurement networks are essential to fully understand changes in bias resulting from changing the network geometry.

For example, expanded measurement networks and GPR analyses at GG and WG suggest that the long-term measurements underestimate accumulation (McGrath and others, 2018). Yet the same expanded networks and the geodetic calibrations suggest that the historically sparse networks underestimate ablation. Indeed, negative geodetic calibrations prevail in existing reanalyses worldwide, with roughly two-thirds of reanalyses producing more negative geodetic balances (Zemp and others, 2013; Andreassen and others, 2015; Barandun and others, 2015; WGMS, 2017). The central tenet of this method is that elevation provides a higher-order control than local effects such as snow redistribution, topographic shading or surface features in glacier-wide mass-balance estimates. Local effects are unlikely to be captured by logistically feasible measurement networks, and are often assumed to be negligible. However, the frequent divergence between geodetic and direct mass balances suggests that important processes elude the elevation-based parameterization.

At WG, ∼20 m amplitude drift and scour patterns characterize ∼20% of the accumulation zone (Fig. 10a; McGrath and others, 2018). Safety and logistical practicalities prohibit the routine collection of a distributed set of measurements from this region, precluding an estimate of average mass within this region of extreme local variability. Sub-annual DEM analyses on glaciers may help to reveal and resolve the complex spatial patterns in snow accumulation, but require extensive, typically unavailable, data to constrain sub-annual changes in surface elevation and its partitioning into components driven by ice flow, changes in material density and mass-balance processes (Sold and others, 2013).

At SG, avalanching constitutes a primary snow accumulation process, yet avalanche input is irregularly recorded by the measurement network, which can alter winter balance estimates in both directions (Clark and others, 2017). The steep headwalls serving as avalanche source regions during winter also provide topographic shading during the ablation season, resulting in an inverted summer balance profile and a larger mass turnover rate than expected from regional climate forcing (Figs 3, 5 and Fig. S6). Roughly 80% of the local forcing can be attributed to augmentation of the accumulation budget and 20% to the ablation budget by shading (Florentine and others, 2018). As shown in Figure 10b, SCG displays similar local controls; spatial distributions of firn align with shading patterns along the glacier margin rather than with elevation. Finally, crevasse fields and icefells increase surface roughness and area-volume ratios, and in turn melt rates via changes to turbulent and latent heat fluxes (Colgan and others, 2016). One such example is found in the icefall between sites AU and N at WG (Fig. 1c) where ablation rates measured within a 15 m elevation range differ by more than 3 m w.e. (Fig. S5). Although hazards present in icefells tend to dissuade direct observations, the enhanced melt occurring in such areas may be essential to include in glacier-wide balance estimates.

Measurement networks rarely span the entire elevation range of a glacier, which can markedly influence the slope of the balance profile (Fig. 3 and Fig. S4), and may consequently explain a substantial fraction of the divergence between direct and geodetic cumulative balances. For example, Figure 9a shows the 1967–2018 uncalibrated cumulative balance at GG reaches +20 m w.e. whereas the calibrated value is ∼30 m w.e. Efforts can be made to constrain the profile, either by imposing limits on mass-balance rate (e.g. Sold and others, 2016), or by using short-term (1–3 years) expanded measurement networks (i.e. stakes, pits, GPR, DEMs) to sample logistically challenging regions of the glacier. However, incorporation of short-term results into long-term programmatic efforts can introduce additional bias over earlier intervals of the record, instead of reducing overall modern biases. Expanded measurement networks help to understand processes driving mass-balance variability and trends (e.g. McGrath and others, 2018), which can subsequently be used to adjust measurement networks, especially if calibration breakpoints are established at a similar time (e.g. GG, 1974; Fig. 9b). Periods of overlap between old and new measurement networks are essential to fully understand changes in bias resulting from changing the network geometry.

For example, expanded measurement networks and GPR analyses at GG and WG suggest that the long-term measurements underestimate accumulation (McGrath and others, 2018). Yet the same expanded networks and the geodetic calibrations suggest that the historically sparse networks underestimate ablation. Indeed, negative geodetic calibrations prevail in existing reanalyses worldwide, with roughly two-thirds of reanalyses producing more negative geodetic balances (Zemp and others, 2013; Andreassen and others, 2015; Barandun and others, 2015; WGMS, 2017). The central tenet of this method is that elevation provides a higher-order control than local effects such as snow redistribution, topographic shading or surface features in glacier-wide mass-balance estimates. Local effects are unlikely to be captured by logistically feasible measurement networks, and are often assumed to be negligible. However, the frequent divergence between geodetic

![Fig. 8.](image) (a) Fractional area change, expressed as a percentage of the initial glacier area. Initial year varies by glacier. (b) Absolute mass change, as determined from mass-balance estimates. Local effects are unlikely to be captured by logistically feasible measurement networks, and are often assumed to be negligible. However, the frequent divergence between geodetic
and others, 2016; Klug and others, 2018), suggesting widespread neglect of important ablation processes. Our calibration coefficients (9 of 13 are negative) corroborate this tendency for glaciological mass-balance estimates to have positive bias (Table S10).

These examples highlight processes that result in mass-balance anomalies where safety or logistical factors challenge measurement via direct observation networks. This problem is not unique to the USGS Benchmark Glacier reanalysis; many of the long-term mass-balance records globally are drawn from glaciers with similar logistical challenges and hazards (e.g. WGMS, 2017). A measurement network with a truly random and adequate spatial distribution to accurately capture mass balance directly without a geodetic calibration would be cost-prohibitive at most glaciers (e.g. Cogley and Adams, 1998), underscoring the limitations of measurement networks and the absolute necessity for geodetic calibration.

6.3. Glacier–climate interactions

Glaciers help to inform many aspects of Earth’s changing climate system by integrating climate processes over basin scales and filling high-elevation knowledge gaps (McAfee and others, 2013; Pepin and others, 2015; Medwedeff and Roe, 2017). The Benchmark Glacier reanalysis provides an opportunity to assess how traditional glacier-climate paradigms are evolving under ongoing atmospheric warming. For example, do mass changes at maritime glaciers remain linked to winter precipitation and continental glaciers to summer temperature forcings (e.g. McCabe and Fountain, 1995; Hodge and others, 1998)?

Our results do not show latitudinal patterns (e.g. south-to-north gradient) in balance rates. Rather continentality exerts a stronger influence on mass balance (Fig. 5). Historically, the maritime (WG, LCG, SCG) glaciers, buffered by abundant snow and cooler summers, exhibited reduced mass loss rates. Compared to earlier analyses, we found increasingly strong correlations between summer and annual balances at all the glaciers (Table S12), regardless of continentality (Hodge and others, 1998; Josberger and others, 2007; O’Neel and others, 2014). Meteorology and mass-balance time series show significant trends for ablation season processes (Figs 6 and 7). Trends in precipitation are less robust due to snow undercatch issues, especially in the windy conditions that characterize the alpine environment (Kochendorfer and others, 2017), but suggest a non-uniform role for precipitation forcing. Precipitation trends are negative...
value decreases because more of the surface is bare ice and therefore darker, which drives a positive feedback with the mass balance (Yde and Paasche, 2010; Naegeli and Huss, 2017). Debris can further complicate this feedback, depending on whether it is thin and increases the ablation rate, or if it becomes thick enough to insulate the ice as it does near the terminus of GG (e.g. Pratap and others, 2015).

Glacier thickness influences the rate at which glacier geometry can evolve, which determines the response time for the geometry to achieve equilibrium with climate (e.g. Johansson and others, 1989; Christian and others, 2018). GG is in the process of dividing into two independent glaciers with different thickness profiles (Farinotti and others, 2019) and thereby different response characteristics (Paul and others, 2004; Jiskoot and Mueller, 2012). The terminus area fed by the west branch (sites X, W, T) retreated more than 1 km since 1993 and is in the process of fragmenting and disconnecting from the main trunk of the glacier. In contrast, the main trunk (sites B, D, V), characterized by lower surface slopes and overdeepened ice, has thinned but retreated only ∼650 m. We expect continued retreat from the disconnecting lobe, which will likely contrast the overdeepened trunk that is more susceptible to altitude-mass-balance feedbacks, as observed across other low-sloping Alaska glaciers (Sass and others, 2017). This divergent response to the same climate highlights the complex nature of linkages among mass balance, glacier geometry, response time and regional climate forcing.

7. Conclusions
A consistent and comprehensive reanalysis of data from the USGS Benchmark Glaciers informed uncertainties, facilitated intercomparison and allowed insight into climate forcing of five North American glaciers. In agreement with other studies, our efforts underscore the importance of geodetic calibration to accurately assess mass-balance trends, and the irreplaceable ability of direct measurements to resolve processes driving glacier change.

Mass loss from the USGS Benchmark Glaciers is ubiquitous since measurements began, with an increase in the rate of mass loss since 1990 at four of the five glaciers. The exception is SG, where local controls, including avalanching and topographic shading, now influence mass balance more than regional

Fig. 10. Orthophotos produced from Worldview imagery acquired on (a) Wolverine Glacier, 12 September 2018 and (b) South Cascade Glacier, 14 October 2015 with insets showing the full glacier extent. Both photos illustrate elevation-independent accumulation variability.
climate. We observed no evidence of latitudinal influence on mass-balance rate but found an increasingly important role of summer balance in governing annual balance. The reanalysis revealed that elevation-independent processes (crevasses, shading, snow redistribution) provide important, often neglected, influence on glacier-wide mass balance. Reanalysis reduced uncertainty in annual glacier mass balance, increasing our confidence in deducing the rate at which glaciers are losing mass. Given current projections of continued summer warming (e.g. IPCC, 2013) and the decadal response times for mountain glaciers (Harrison and others, 2001; Christian and others, 2018), we expect continued mass loss with potential for further acceleration in loss rates from the USGS Benchmark Glaciers, as well as the thousands of other glaciers in Western North America and Alaska. Improved resolution of mass-balance processes documented here is essential for developing physically accurate models and to better resolve linkages among glaciers, climate, and ecosystems. As our confidence in our ability to resolve mass changes at glaciers increases, it opens a new and important layer of questions that address the wellbeing and function of our ecosystems and planet in a warmer future.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/jog.2019.66.

Acknowledgements. Countless individuals helped with fieldwork and historical data reduction. We are particularly appreciative of program maintenance from Mark Meier, Bob Krimmel, Larry Mayo, Dennis Trabant and Rod March. Recently Adam Clark, Erin Whorton and Zach Miller provided invaluable assistance with field and data preparation. Thanks to the Juneau Icefield Research Program for data sharing, access and collaboration, in particular to Mauri Pelto for guiding us through the history of data collection by JIRP. Comments from Jack Kohler, Liam Colgan, Mike Loso and an anonymous reviewer greatly improved the manuscript. Funding for the work came from USGS under the guise of many programs throughout the years.

Author contribution. The following authors Shad O’Neil (SO), Christopher McNeil (CM), Louis Sans (LS), Cynthia Florentine (CF), Emily Baker (EB), Erick Petitzsch (EP), Dan McGrath (DM), Andrew Fountain (AF), Dan Fagre (DF), contributed to this work. Concept and design were done by SO, LS, CM; formal analyses by LS, CM, EF, EB, DM; visualization was done by SO, LS, EB, CM; funding acquisition and project administration/supervision were done by SO, CM; programming was done by CM, LS, CF, EB, CM; visualization was done by SO, LS, CM; edit and review by LS, CF, AF, CM; contributed to this work. This work was supported by the US Government.

References


Mayo LR, Trabant DC and March RS
McCabe GJ and Fountain AG
McAfee SA, Walsh J and Rupp TS
Marzeion B and 5 others
Krimmel RM
Johnson A
Josberger EG, Bidlake WR, March RS and Kennedy BW
Larsen CF and 5 others
(2013) Summary for policymakers. In Stocker TF Qin D Plattner GK IPCC
Krimmel RM
Höhle J and Höhle M
Journal of Glaciology
(2016)
46
Glacier, Alaska
3946–0459.00093.
Cascade Glacier, Washington, 1986
1468-0459.00093.
Annals of Glaciology
western Alaska.
Cryosphere
1468–0459.