20th century global glacier mass change: an ensemble-based model reconstruction

Jan-Hendrik Malles[1,2] and Ben Marzeion[1,2]

[1]Institute of Geography, Climate Lab, University of Bremen, Bremen, Germany
[2]MARUM - Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

Correspondence: J.-H. Malles (jmalles@uni-bremen.de)

Abstract. Negative glacier mass balances in most of Earth's glacierized regions contribute roughly one quarter to currently observed rates of sea-level rise, and have likely contributed an even larger fraction during the 20th century. The distant past and future of glaciers' mass balances, and hence their contribution to sea-level rise, can only be calculated using numerical models. Since independent of complexity, models always rely on some form of parameterizations and a choice of boundary conditions, a need for optimization arises. In this work, a model for computing monthly mass balances of glaciers on the global scale was forced with nine different data sets of near-surface air temperature and precipitation anomalies, as well as with their mean and median, leading to a total of eleven different forcing data sets. Five global parameters of the model's mass balance equations were varied systematically, within physically plausible ranges, for each forcing data set. We then identified optimal parameter combinations by cross-validating the model results against in-situ mass balance observations, using three criteria: model bias, temporal correlation, and the ratio between the observed and modeled temporal standard deviation of specific mass balances. The goal is to better constrain the glaciers' 20th century sea-level budget contribution and its uncertainty. We find that the disagreement between the different ensemble members is often larger than the uncertainties obtained via cross-validation, particularly in times and places where few or no validation data are available, such as the first half of the 20th century. We show that the reason for this is that the availability of mass balance observations often coincides with less uncertainty in the forcing data, such that the cross-validation procedure does not capture the true out-of-sample uncertainty of the glacier model. Therefore, ensemble spread is introduced as an additional estimate of reconstruction uncertainty, increasing the total uncertainty compared to the model uncertainty obtained in the cross validation. Our ensemble mean estimate indicates a sea-level contribution by global glaciers (excluding Antarctic periphery) for 1901 - 2018 of 76.2 ± 5.9 mm sea-level equivalent (SLE), or 0.65 ± 0.05 mm SLE yr^{-1}.

1 Introduction

Glacier mass loss across most of the world is constituting a major part of the contemporary and projected 21st century sea-level rise (e.g., Slangen et al., 2017; Oppenheimer et al., 2019). Moreover, glaciers constitute important freshwater reservoirs for some regions of the world, and their vanishing is likely to induce seasonal water scarcity in regions depending on those

reservoirs (Cruz et al., 2007; Huss and Hock, 2018; Wijngaard et al., 2018; Kaser et al., 2010; Small and J. Nicholls, 2003).

The future evolution of glaciers' mass balances is usually estimated using numerical models (Hock et al., 2019; Marzeion et al., 2020). This is the case for the more distant past as well (e.g., Goosse et al., 2018; Parkes and Goosse, 2020), since glaciers are mostly situated in remote locations and thus lacking comprehensive in-situ measurement densities, at least before 1950 (WGMS, 2020). It is therefore important to assess and improve glacier mass balance models used to reconstruct or

30   project glacier evolution. An ensemble-based, long-term reconstruction can add to our understanding of the uncertainties in glacier modeling, which might in turn enhance our ability to make more robust projections of glacier mass loss (Hock et al., 2019; Marzeion et al., 2020). The modeling approaches to establishing global estimates for the glaciers' mass balances mostly make use of temperature index melt models to represent the energy available for melting precipitation and ice (e.g., Huss and Hock, 2015; Radić and Hock, 2011; Hirabayashi et al., 2013). As a glacier's mass balance is interrelated with its geometric

35   and hypsometric properties, some kind of length-area-volume scaling relation is often incorporated to account for changes in these properties in the models (Bahr et al., 2015). The model used in this work additionally includes a response time scaling to account for the glacier geometries' response lagging climatic forcing, but is lacking an explicit representation of ice dynamic processes such as deformation, sliding, or calving (Marzeion et al., 2012).

40   Although there are approaches based on solving the energy balance at the ice surface, the models used for this are either yet lacking ice dynamics or geometric scaling (Shannon et al., 2019), can only be applied to a small number of glaciers and depend on upscaling to obtain global numbers (Giesen and Oerlemans, 2013), or do not perform significantly better than a similar model without energy balance implementation (Huss and Hock, 2015). Another difficulty for models resolving the energy balance is the introduction of additional parameters that have to be constrained, which in turn adds complexity to the

45   model optimization.

Due to computational limitations, models solving the full equations of motion and thermodynamics individually for each glacier are generally not applied at the global scale. However, the Open Global Glacier Model (OGGM, Maussion et al., 2019) has been applied to compute ice velocity and thickness for each glacier based on a flowline approach.

50

None of the models resolving the energy balance or explicitly calculating ice dynamics have been applied to globally reconstruct the glacier mass change on a century time scale. This implies that a comprehensive analysis determining which modeling approach might be most appropriate is not yet possible. The need for a robust model evaluation, which can also be used to better understand the glacier model contribution to projection uncertainty (Marzeion et al., 2020), is obvious.

55

Models of (parts of) the Earth system are typically evaluated using observations and/or proxy data, usually with the objective to minimize the model's deviation from observations, e.g. by minimizing the root mean squared error (RMSE, Gleckler et al., 2008; Taylor, 2001). Although in the case of glaciers, direct in-situ mass balance measurements are sparse and very heteroge-

neously distributed, they are essential in assessing the uncertainty of mass balance models, though other evaluation methods
60 exist, e.g. by calibration with a combination of satellite gravimetry, altimetry and glaciological measurement data (Huss and
Hock, 2015). Such combined calibration data usually are not available for individual glaciers, or do not have the temporal
resolution required to assess the model's ability to capture variability.

Uncertainties of numerical models are caused by (i) uncertain boundary and initial conditions, (ii) approximations of the
65 model's equations, and (iii) lack of knowledge about parameters involved in the model set-up (Hourdin et al., 2017). There-
fore, optimization of parameters and/or input data is a standard procedure in glacier modeling (Huss and Hock, 2015; Radić
and Hock, 2011; Marzeion et al., 2012). Often, a single parameter is chosen to be minimized during the calibration (e.g., the
model's RMSE with respect to observed in-situ mass balances). Rye et al. (2012) suggested multi-objective optimization for a
(regional) glacier model, striving for 'Pareto optimality' (Marler and Arora, 2004), to constrain parameters more robustly.
70

Here, we apply a multi-objective optimization, concerning the five global parameters most relevant in the applied model, for
each of nine meteorological forcing data sets (see Table 1), their mean and their median. Since the model is able to hindcast
glacier evolution, the aim of this work is to (i) optimize the model parameters in order to obtain model setups that reproduce
in-situ mass-balance observations as closely as possible, and (ii) to more robustly estimate model uncertainty, taking into
75 account ensemble spread at times and in regions where observations are sparse. We use the model of Marzeion et al. (2012),
but introduce changes to the mass-balance calibration routine (see 2.1.2). Additionally, we incorporate newer boundary and
initial conditions as well as reference data, against which the model is calibrated and evaluated. We show that the ensemble
approach to the reconstruction produces more robust estimates of model uncertainty than taking into account results from a
cross-validation alone.

80 2   Data and Methods

2.1   The global glacier mass balance model

2.1.1   Basic equations and parameters

In this section, those features of the mass-balance model that are relevant to the optimization procedure are described. A more
thorough description is given in Marzeion et al. (2012).
85 The annual mass balance $B(t)$ of each glacier is computed as:

$$B(t) = \left[\sum_{i=1}^{12} [P_i^{solid}(t) - \mu^* \cdot max(T_i^{terminus}(t) - T_{melt}, 0)] \right] - \beta^* \qquad (1)$$

where B is the annual modeled mass balance for an individual glacier in year t, P_i^{solid} the amount of solid precipitation in
month i, μ^* a glacier-specific temperature sensitivity parameter, $T_i^{terminus}$ the mean temperature in month i at the glacier's

terminus elevation, T_{melt} a global threshold temperature for snow and ice melt at the glacier surface, and β^* a model bias
90   correction parameter. Values for μ^* and β^* are obtained by assuming an equilibrium of the glacier in present-day geometry
during a 31-year period centered around year t^*. In contrast to the initial publication of the model, we objectify the selection of
t^*: while Marzeion et al. (2012) argue that t^* is a function of the regional climatological history, it also depends on the glacier's
response time scale, as discussed in Roe et al. (2020, submitted), for which there is no reason to assume spatial coherence. This
means that we now do not spatially interpolate t^* as before, but introduce it as an additional global parameter. In section 2.1.2,
95   we elaborate further on this point.

The inference of the glacier-specific parameters (μ^* and β^*) is assessed in a leave-one-glacier-out cross-validation procedure
to determine the out-of-sample uncertainty. While values for μ^* can be computed for each individual glacier based on t^*, those
for β^* are spatially interpolated from the ten closest glaciers with at least three years of available in-situ observations, using
100   inverse distance weighting.

While one global parameter (T_{melt}) was introduced in Eq. 1, three other ones are associated with the calculation of the
monthly solid precipitation $P_i^{solid}(t)$:

$$P_i^{solid}(t) = (a \cdot P_i^{CRUclim} + P_i^{anom}(t)) \cdot (1 + \gamma_{precip} \cdot (z_{mean} - z_{CRUclim})) \cdot f_i^{solid}(t) \tag{2}$$

105   where a is a precipitation correction factor, $P_i^{CRUclim}$ is the monthly climatological precipitation sum taken from the grid
point of the CRU CL 2.0 data set (New et al., 2002) closest to the respective glacier in month i, $P_i^{anom}(t)$ is the monthly
total precipitation anomaly deduced from the forcing data set, γ_{precip} is a global precipitation lapse rate, z_{mean} is the mean
elevation of the glacier, $z_{CRUclim}$ is the elevation of the grid point in the CRU CL 2.0 data set, and $f_i^{solid}(t)$ is the fraction of
solid precipitation:

$$
110 \quad f_i^{solid}(t) = \begin{cases} 1 \text{ if } T_i^{terminus}(t) \leq T_{prec\ solid} \\[2mm] 0 \text{ if } T_i^{z_{max}}(t) \geq T_{prec\ solid}, \text{with } T_i^{z_{max}}(t) \\[2mm] = T_i^{terminus}(t) + \gamma_{temp} \cdot (z_{max} - z_{terminus}) \\[2mm] 1 + \frac{T_i^{terminus}(t) - T_{prec\ solid}}{\gamma_{temp} \cdot (z_{max} - z_{terminus})} \text{ otherwise} \end{cases} \tag{3}
$$

where $T_{prec\ solid}$ is a global threshold temperature for solid precipitation, γ_{temp} is an empirically derived, local temperature
lapse rate, z_{max} the maximum glacier elevation, and $z_{terminus}$ the elevation of the glaciers' terminus.

115

The four global parameters (T_{melt}, a, γ_{precip}, and $T_{prec.\ solid}$) introduced in Eq. 1 - 3 are at the core of the model's mass balance computations and hence subject to the optimization presented here. Marzeion et al. (2012) used the CRU TS 3.0 data set to obtain $T_i^{terminus}(t)$ and $P_i^{anom}(t)$. Here, we include additional meteorological data sets as well as their mean and median values in the optimization (see Sect. 2.2.1).

120

The monthly mass balances are subsequently translated into volume, area and length changes by geometric scaling and relaxation (details in Marzeion et al., 2012). Initial values for the area of each individual glacier at the start of the model run (e.g., beginning of the 20th century) are found using an iterative approach that minimizes the difference in area between modeled glacier and the Randolph Glacier Inventory (RGI) record in the year of the respective observation. If this iterative procedure is
125 not successful, the glacier is not included in the reconstruction. For these glaciers, a simple upscaling is applied in the computation of regional and global results.

Note that since the CRU CL 2.0 data set used to obtain $P_i^{CRUclim}$ and $T_i^{CRUclim}$ does not cover Antarctica, we do not consider glaciers in the periphery of Antarctica and subantarctic glaciers here (labeled region 19 in RGI, 2017).

130 **2.1.2 Mass-balance calibration**

As explained above, we treat the parameter t^* as a global one, opposed to a glacier-specific estimation in Marzeion et al. (2012). In order to illustrate the reasoning, we need to discuss the mass-balance calibration for an individual glacier in the model in detail. The calibration is based on the idea of inferring a glacier's temperature sensitivity μ^* by finding a climatologcal time period in the forcing data set (centered around t^*) which would result in a zero annual mass balance of the glacier in present-day
135 geometry. Thus, for each center year $\tilde{t}$ of a 31-year period, we can calculate $\mu(\tilde{t})$ by requiring:

$$B(\tilde{t}) = \sum_{i=1}^{12}[P_{i,clim}^{solid}(\tilde{t}) - \mu(\tilde{t}) \cdot max(T_{i,clim}^{terminus}(\tilde{t}) - T_{melt}, 0)] = 0 \tag{4}$$

where $P_{i,clim}^{solid}(\tilde{t})$, and $T_{i,clim}^{terminus}(\tilde{t})$ are climatological averages of $P_i^{solid}(\tilde{t})$ and $T_i^{terminus}(\tilde{t})$. Note that the calculation is based on a smaller number of years when $\tilde{t} < 1916$ or $\tilde{t} > 2003$. For each glacier that has at least three years of in-situ mass-balance observations, we calculate the modeled mass balance (based on Eq. 1) for each $\tilde{t}$. Then, the associated model bias of
140 an individual validation glacier is calculated as

$$\overline{B_M} - \overline{B_o} = \beta(\tilde{t}) \tag{5}$$

where $\overline{B_M}$ is the mean modeled mass balance of the respective glacier for the years of available mass balance measurements, and $\overline{B_o}$ the mean observed mass balance. Marzeion et al. (2012) chose t^* to be that $\tilde{t}$, for which $|\beta(\tilde{t})|$ was minimal. μ^* was then calculated from equation 4 applied to t^*, and β^* taken as $\beta(t^*)$. For glaciers without in-situ observations of mass balances,
145 t^* and β^* were interpolated from the ten closest glaciers with observations, using an inverse distance weighting. Using this

method, Marzeion et al. (2012) were able to identify a suitable parameter set in the leave-one-out cross-validation procedure, applying CRU TS 3.0 as atmospheric boundary conditions. However, this is not generally the case for the meteorological data sets considered here, and there is a conceptual shortcoming in the spatial interpolation of t^*, which we will illustrate for one exemplary model setup.

150

The upper panel of Fig. 1 shows the global average of $\beta(\tilde{t})$, weighted by the length of each glacier's in-situ mass-balance measurement time series (henceforth, all mentioned averages over different glaciers imply such a weighting), using CRU TS 4.03 as boundary condition, applying the optimal parameter set (see section 2.3).

155   The lower panel shows that the distribution of t^* estimated directly is bi-modal, with frequent values either at the beginning or end of the considered period, but the spatial interpolation leads to a more even distribution. This in turn means that, generally speaking, the spatial interpolation moves t^* towards the mid of the considered time period, thereby increasing the value of β^* for glaciers with an early t^*, and decreasing it for those with a late t^* (see upper panel of Fig. 1). It also shows that there are more glaciers with t^* at the beginning of the 20th century than at the end of the 20th century or the beginning of the 21st century.

160

Furthermore, those glaciers with t^* at the beginning of the 20th century tend to have a positive β^*, implying that even with present-day geometry, those glaciers would have lost mass under climatic conditions of the early 20th century. The zero-crossing of the global average $\beta(\tilde{t})$ is thus found at a period when positively and negatively biased glaciers cancel each other. Since moving the median of t^* towards the mid of the of the modeled period generally goes along with an increase of the 165   averaged model bias, using the spatial interpolation of t^* tends to lead to a positively biased model setup, which then becomes apparent in the leave-one-glacier-out cross-validation.

In order to avoid this effect, and taking into account that neighboring glaciers will have different response times, such that even if they experience a very similar evolution of climate anomalies we cannot expect a close spatial coherence of t^*, we no 170   longer spatially interpolate t^* and treat it as a 5th global parameter instead. Note that μ^* is still a glacier-specific parameter following Eq. 4, and that $\beta(t^*)$ is still interpolated from the ten closest glaciers in an inverse distance weighted manner. Also note that the leave-one-glacier-out cross-validation (Sect. 2.3) will reveal any potential new model errors introduced through this change.

2.2   Data

175   ### 2.2.1   Meteorological Data

We conducted the search for an optimal parameter set for the version 4.03 of the CRU TS data (corresponding to an update of Marzeion et al., 2012) and additionally eight reanalysis data sets, as well as the mean and the median of all the data sets (see Table 1). Data sets not extending back to 1901 were filled with CRU TS 4.03 data, exclusively for the purpose of initialization

of glacier areas; the results are only shown (and evaluated) during time periods for which we have forcing data from the re-
180 spective data set.

Anomalies of temperature and precipitation were calculated with respect to the 1961 to 1990 reference period used in CRU
CL 2.0. For those data sets not covering the period 1961 to 1991, they were obtained by calculating the difference between
the 1961 to 1990 and the 1981 to 2010 periods in the CRU TS 4.03 data set, and subsequently subtracting this value from the
185 respective data set's 1981 to 2010 mean.

2.2.2 Glacier Data

The glacier model requires information about location, area, terminus and maximum elevation of each glacier at some point of
time within the modeled time interval. The RGI provides these data. Its most recent version (RGI v6.0) was used in this work.
The RGI relies mostly on Landsat and other satellite imagery. Distinction of individual glaciers within glacier complexes was
190 realized mostly by semi-automatic algorithms for detecting watershed divides (RGI, 2017).

To be able to cross-validate the modeled mass balances, we use in-situ observations of glaciers' mass balances collected by
the World Glacier Monitoring Service (WGMS, 2018). We ignore any uncertainties of these observations (Cogley, 2009) and
treat them as the 'true' annual mass balance of a glacier in the recorded year.

195 ### 2.3 Parameter optimization strategy

For the identification of a optimal parameter set, we applied a 'brute-force' approach, i.e. we varied each parameter other than
t^* (see below) using the following ranges, for each meteorological data set:

- threshold temperature for snow/ice melt (T_{melt}) [°C]: {-2, -1, 0, 1, 2}
200 - threshold temperature for solid phase precipitation ($T_{prec.\ solid}$) [°C]: {-1, 0, 1, 2, 3, 4}
- precipitation lapse rate (γ_{precip}) [%/100 m]: {0, 1, 2, 3, 4, 5}
- precipitation correction factor (a): {1, 1.5, 2, 2.5, 3}

This resulted in 900 model validation runs for each of the eleven forcing data sets (i.e., 9900 runs in total). For all forcing
205 data sets except 20CRV3, all zero-crossings of the global mean $\beta(\tilde{t})$ were found before with $\tilde{t} < 1920$ (for 20CRV3, some
were found in 1962 and 1976). For each forcing data set, we selected the twenty best-performing parameter sets that showed
a zero-crossing of the global mean $\beta(\tilde{t})$. We then performed another cross-validation with those parameter sets to fine-tune t^*,
applying the range 1901 to 1920, except for 20CRV3 where we applied the time ranges 1909 - 1918, 1960 - 1964, and 1974 -
1978. Hence, we performed 400 additional cross-validation runs per data set.

210

From those cross-validations, three characteristic statistical measures of model performance were computed: model bias (i.e., mean model error) with respect to observations, the temporal correlation with observations, and the ratio of standard deviations of interannual variability between modeled and observed mass balances. We do not include the mean squared error (MSE) as a performance measure, since it is simply a (weighted) combination of the three performance measures:

215    $$MSE = \sigma_M^2 + \sigma_o^2 - 2\sigma_M\sigma_o R + (\overline{M} - \overline{O})^2 \qquad (6)$$

where σ_M is the standard deviation of modeled mass balances, σ_o the standard deviation of observed mass balances, R the Pearson correlation coefficient, $\overline{M}$ the mean of modeled mass balances, and $\overline{O}$ the mean of observed mass balances (thus, the last term corresponds to the squared bias).

220   From Eq. 6 it can be inferred that a minimum MSE occurs for a model setup in which the standard deviation ratio equals the correlation coefficient. Hence, in a model setup that is not perfectly (positively) correlated with the observations (i.e., $0 < R < 1$), a more realistic standard deviation ratio (e.g. $1 \geq \frac{\sigma_M}{\sigma_o} > R$) will result in a higher MSE. However, a correlation coefficient equal to one is generally not achievable in complex models such as the one used in this work. Therefore, minimizing the MSE will lead to preference of parameter sets that underestimate variance. This is problematic, since a correct representation of vari-
225   ance is indicative of correct model sensitivity to changes in the forcing. E.g., it is possible to imagine to apply a model setup that yields a low bias and good correlation, but largely underestimates the interannual variation of mass balances. It is therefore beneficial to not only minimize the MSE, but rather to minimize the three statistical coefficients it comprises individually, in order to not trade a realistic model sensitivity for a smaller MSE.

230   All three performance measures were calculated for each validated glacier in a respective data set, and then averaged over all these glaciers, weighted by the number of available mass balance observations per glacier.

Standard deviation ratios were brought to represent the deviation from an optimum value (i.e. one) by:

$$SR = \frac{\sigma_M}{\sigma_o} - 1 \qquad (7)$$

235   To determine for each meteorological data set a model parameter set that, on average, shows the highest skill to represent the behavior of observed glaciers, we normalize the performance measures introduced above such that the individual scores s range from 0 for the worst to 1 for the best validation result by the following equations:

$$s_{i,bias} = \frac{max(|bias|) - |bias_i|}{max(|bias|) - min(|bias|)}, s_{i,SR} = \frac{max(|SR|) - |SR_i|}{max(|SR|) - min(|SR_i|)}, s_{i,R} = \frac{|R_i| - min(|R|)}{max(|R|) - min(|R|)} \qquad (8)$$

where i is the individual model setup the score is calculated for. These scores were then added up to identify the 'optimal' model
240   setup as the one with the maximum overall score. If a model setup obtained the single best result for all three performance measures individually, it would thus yield a score of three. Note that the three (or potentially other) performance measures might be weighted differently, based on the objective of the model application. However, as shown below, we do not find

substantial trade-offs between the three performance measures, such that any potential weighting would have a very limited influence on the results.

3 Cross validation and uncertainty assessment

3.1 Performance measures

Table 2 shows the values obtained for performance measures and optimal global parameters. We differentiate between the mean and median of the forcing data input used as individual boundary conditions (mean/median input) and the mean and median of the ensemble output values (mean/median output). For more than half of the validated meteorological data sets, the global mean bias of the optimal parameter set is smaller than 10 mm w.e. yr^{-1}, and the correlation is larger than 0.6, while the amplitude of the interannual variability is estimated correctly within a small margin (ca. 5 %). RMSEs lie roughly between 700 and 800 mm w.e. yr^{-1} for most data sets. Only 20CRV3 shows a significantly higher RMSE, caused by some large outliers. Note that the number of glaciers that cannot be initialized also depends on the meteorological data set used as boundary condition. CERA20C, e.g., not only performs the worst (obtaining an overall score of 1.38 using the optimal parameter set), but leads to only 274 of 299 validation glaciers being initialized in the cross validation, and 180,799 of the 212,795 glaciers in the global reconstruction run, representing 84 % of today's global glacier area. In contrast, the best performing model setup that covers the whole model period (CRU TS 4.03) is able to initialize 298 validation glaciers and 201,004 glaciers in the global reconstruction run, representing ca. 98 % of the global glacier area. Following our scoring system, we find that the statistically best performing single data set covering the whole model period is CRU TS 4.03, and the overall best performing data set, but only covering 1979 - 2018, is ERA5. Our best estimate for the whole model period is the mean model output.

Independent of the time period considered, the mean output of the ensemble shows the best performance, exceeding not only the best individual ensemble member, but also the result obtained by the mean and median input. The statistically best-performing individual ensemble members vary with the time periods that are covered by the meteorological data sets. E.g., for the period 1958 to 2018, JRA55 leads to the best performance; from 1979 onward, it is ERA5. Table 2 also shows that the performance measures attain better values if the averages are weighted be length of the observation time series than than in the non-weighted average, illustrating the need for long-term observations for reliable model validation.

In order to assess the consistency of validation results among the ensemble members, two-sample Kolmogorov-Smirnov-Tests for the similarity of distributions were conducted for all 55 unique pairs of modeled mass-balance and model deviation distributions. Model deviation here refers to the respective differences between each modeled and observed mass-balance value in the cross-validation procedure; its average thus corresponds to the average of the bias weighted by the number of available mass-balance observations per validation glacier. The confidence level we require for rejecting the similarity of distributions is at 95 %. Regarding the distributions of modeled mass-balances, only 10 (18 %) of the tested pairs are not significantly different; all involving the six best-scored model setups (see Table 2). Model deviation distribution pairs do not significantly

differ in 27 (49 %) cases, of which only 1 (2 %) involved 20CRV3, CERA20C, or ERA20C. We conducted Welch's t-test for the similarity of means in the same manner. Here, only the three lowest-scored model setups' means of modeled mass-balances are significantly distinguishable from other ensemble members. Concerning the mean model deviation, only that of CERA20C was significantly different from the others. Hence, the similarity tests indicate that the results of model setups with higher scores

280 tend to be more consistent among each other and to differ from lower scored ones statistically. Model deviation distributions significantly different from those of other ensemble members are to a large degree produced by low-scored model setups, while the mean is only significantly different for CERA20C. The significantly high bias and low score of CERA20C indicate particular issues with this forcing data set and lead us to exclude it from the following ensemble calculations. In the subsequent section we will explore where these issues stem from and in doing so explain why the temporal and spatial constraints of the

285 validation data hinder us to make assertions over which individual model setup is the most reliable one over the whole temporal and spatial model domain.

3.2 Differences between ensemble members inconsistent with uncertainty estimates

The leave-one-glacier-out cross-validation procedure applied here is designed to estimate the uncertainty of model results for glaciers that have no in-situ mass balance observations, and for times where there are no in-situ observations. Therefore, in

290 principle, the results of the individual ensemble members should agree within their respective uncertainty estimates. However, there is a strong spatial bias in in-situ mass balance observations towards certain RGI regions, mostly locations where also the past state of the atmosphere is well constrained, since both atmospheric and glaciological observations are denser in easily accessible regions. The majority of glaciers, however, is situated in remote locations where observations of the state of the atmosphere were very sparse, particularly in the first half of the 20th century. Thus, the cross-validation is biased towards

295 times and places where the state of the atmosphere, i.e., the boundary conditions of the glacier model, can be assumed to be exceptionally well constrained.

Figure 2 shows that 66% of the validation data originate from only four RGI regions: Western Canada and USA, Scandinavia, Central Europe, and Central Asia. The lower panel shows the fraction of mean annual ensemble variance of global mass

300 change rates in the modeled period attributable to each RGI region. Most of the ensemble spread is due to disagreement in sparsely observed regions that contain much glacier ice. Of the mean ensemble spread, 33 % can be attributed to the disagreement in estimates for the Greenland periphery. That value increases to 67 % if we included CERA20C in the calculation. This indicates that peripheral glaciers in Greenland are responsible for a large amount of the ensemble spread as well as for the large divergence of CERA20C from the other ensemble members.

305

In the upper left panel (a) of Fig. 3, the issue of temporally biased validation data (all are from the second half of the 20th century or the beginning of the 21st) can be recognized. Mean mass loss rates calculated with forcing data sets that have complete data coverage over the whole model period for the four previously mentioned well-observed regions are shown. Comparing results for the four best-observed regions to global results (right part (B and D) of Fig. 3), it can be seen that the disagreement

310 on the global scale is larger than in the well-observed regions, and that the global reconstruction forced by CERA20C is far off the three other ensemble members while it is not so in the well-observed regions. This behavior can be explained by the much more pronounced warming of glacier locations at the global scale in CERA20C until ca. 1960 (part D of 3): during the calibration, lower temperatures at t^* will lead to higher temperature sensitivities (see Eq. 4). Similarly, the greater increase of temperature will result in higher mass loss rates.

315

Concerning these issues with CERA20C, it is striking that in spite of its large positive specific mass-balance bias in the cross-validation, global mass change estimates obtained with it are strongly more negative than those of the other ensemble members. This underlines the fact that even though the cross-validation is crucial in the optimization process, we cannot entirely rely on it for assessing global and long-term reconstruction performance of individual data sets. Therefore, and because,

320 as stated in the previous section, the best-performing data sets do produce statistically quite similar results for the validation glaciers, we will only use estimates based on the ensemble – i.e., not individual members – in the following.

In both the well-observed regions (panel (a) in Fig. 3) and the global scale (panel (b)), the different model setups disagree stronger in the first half of the 20th century, reflecting that uncertainty in the atmospheric conditions during that time is also

325 greater.

All in all, we find that the ensemble spread tends to be larger than uncertainty estimates obtained via the cross-validation, and that this is caused by the majority of glacier observations coming from places and times where the uncertainty of the state of the atmosphere is smaller than what can typically be expected in glacierized regions. Additionally, we assume that the individual glaciers' error estimates are uncorrelated with each other and random, for we do not have direct model error estimates for

330 every glacier and can thus not account for correlations of individual glaciers' errors. However, the ensemble approach allows to explore if, and to which degree, the cross-validation underestimates the true uncertainty of the reconstruction.

3.3 Combining model and ensemble uncertainty

To account for both the model error, as calculated in the cross-validation procedure (RMSE), and for and the ensemble spread, the total uncertainty of the ensemble average is calculated as follows. First, we calculate the model error of the ensemble

335 average solely determined by the means of the cross-validation error:

$$\epsilon_{model}(t) = \frac{\sqrt{\sqrt{\sum_{i}^{n} \epsilon_i(t)^2}}}{n} \tag{9}$$

where $\epsilon_i(t)$ is the model uncertainty computed in the cross-validation procedure for an individual ensemble member i for year t. Then we add the ensemble spread as a further uncertainty measure to the model error of the ensemble average:

$$\epsilon_{ensemble}(t) = \sqrt{\epsilon_{model}(t)^2 + \sigma(t)^2} \tag{10}$$

340   where $\sigma(t)$ is the ensemble standard deviation in year t.

Figure 4 shows the temporal evolution of total uncertainty ($\epsilon_{ensemble}$) as well as the aggregated model uncertainty (ϵ_{model}) and ensemble spread ($\sigma_{ensemble}$) of the ensemble mean mass change rate estimate. The total uncertainty of the ensemble mean estimate grows as we go back in time, with a sharp increase in the first twenty-five years. This is due to the increase in the

345   model error of the ensemble average, especially in the first decade of the 20th century, which is produced by very high mass losses for a few glaciers in some model setups during that period. The ensemble spread is also greater during the first half of the 20th century compared to later years, which can be attributed to less agreement between meteorological data sets in earlier years. Note that also the number of ensemble members shrinks going back in time, since not all reanalysis products provide data for the whole period.

350   ## 4   Global Glacier Mass Loss

Figure 5 shows the temporally accumulated mass loss estimates, relative to 1980 (the year from which onward all meteorological data sets have data coverage), and their uncertainties. The upper panel shows the estimates for each individual ensemble member as well as their model uncertainties ϵ_{model}. Especially in the first half of the 20th century, ensemble members are diverging, with CRU TS 4.03 showing the lowest and ERA20C, next to CERA20C, the highest mass loss during that period. The

355   ensemble average mass change estimate over the whole model period is -76.2 $\pm$ 5.9 mm SLE, which translates to an average mass change rate of -0.65 $\pm$ 0.05 mm SLE yr^{-1}. Table 3 displays the regional and global mass loss rates for different reference periods. Mass change rates estimates for more recent periods are increasingly negative across most regions, reaching -1.00 $\pm$ 0.06 mm SLE yr^{-1} accumulated globally in the most recent period. The only time and region for which an increase in glacier mass is estimated are the Southern Andes in earlier years, although with a relatively high uncertainty due to ensemble spread

360   (see Fig. 2). To explore the period of decelerated mass loss during the 20th century shown in Fig. 6, the periods 1901 to 1940 and 1941 to 1980 are shown in Table 3. For most regions, the mass loss change rate estimates are substantially less negative in the latter period; only New Zealand exhibits a significantly larger mass loss. Regarding the global estimate, most of the mass loss deceleration took place in Greenland and the North American continent (i.e. regions 1 to 5). Thus, after increasing mass loss rates until around 1930 (see Fig. 6), glaciers started to lose less mass until around 1980, possibly caused by atmospheric

365   cooling induced by increasing aerosol concentrations (Ohmura, 2006; Ohmura et al., 2007; Wild, 2012). From then on, the glaciers' contribution to sea-level rise accelerated again until the end of the modeled period. Figure 7 shows the drivers of this behavior: the global ensemble mean temperature (lower panel) and precipitation anomalies as well as total amount of solid precipitation (upper panel; see Eq. 2 and 3; all weighted by glacier area). From ca. 1980 on, heat available for ice and snow melt, i.e. the temperature anomaly, increased monotonously. While precipitation at the glacier locations tended to increase over

370   time, the amount of solid precipitation at glacier locations decreases from roughly 1980, implying that not only ablation is increased, but also accumulation is decreased. In contrast to that, the increase in precipitation between ca. 1930 to 1950 was accompanied by a similar increase in solid precipitation, indicating that the warm anomaly at the same period was too weak

reduce accumulation.

375   Concerning uncertainty estimates, Table 3 shows that most of the uncertainty stems from the regions Alaska, Arctic Canada
(North), and Greenland in the more recent periods. In the earliest period, the Russian Arctic region exhibits the highest uncer-
tainty, which is more than double the value of the central regional estimate for that period, indicating that the large model error
in the early 20th century (see Fig. 4) is mostly produced in this region, and in Greenland.

5   Discussion

380   Table 4 shows our global results compared to previously published estimates for mass loss rates over certain periods. Overall,
there is good agreement within the respective uncertainty ranges. Only for the periods 2003 to 2009 and 2002 to 2016, there
is a significant disagreement between literature values and our model results. The disagreeing values for 2002 to 2016 from
Wouters et al. (2019) were derived from gravimetry (GRACE) data. Estimates for 2003 to 2009 from (Gardner et al., 2013)
also involve GRACE data. Only the disagreeing values from Cogley (2009) do not involve gravimetry measurements. Part of

385   these disagreements might be explained by the storage of meltwater in glacial lakes (Shugar et al., 2020), which (because of the
close proximity to the glaciers) cannot be separated from the ice mass in gravimetry data. GRACE will therefore observe lower
mass change values than in-situ or geodetic observations. However, since these lower values are more correct concerning the
glaciers' contribution to sea-level rise, the issue points to the larger problem of distinguishing between glacier mass change,
and the corresponding sea-level change, which are not exactly equal. However, Shugar et al. (2020) also point out that glacial

390   lake storage accounts for only about 1 % of glacier melt volume (excluding Greenland and Antarctica), which indicates that
this inconsistency is of limited relevance. Gardner et al. (2013) point to discrepancies between satellite-derived and in-situ
estimates of glacier mass losses, alleging a negative bias in in-situ observations. Zemp et al. (2019) addressed this issue by
combining glaciological and geodetic measurements. Although our model is calibrated solely using in-situ observations, its
estimates are still close to Zemp et al. (2019), in which the uncertainty for the longer period is admittedly large (Table 4).

395   Finally, estimates of the global glacier mass change contribution to sea-level rise, excluding Greenland and Antarctic periphery
and not given in Table 4, of Frederikse et al. (2020) agree well with ours for the more recent time intervals they specify (1957 -
2018 and 1993 - 2018), while our estimates lie at the very low end of the confidence interval given for the whole time interval
they studied (1900 - 2018). This is presumably due to the modeling approach that their estimates in early years rely on, which
includes estimations of disappeared and missing glaciers that are not included in the RGI. The increase of global glacier mass

400   loss estimates this causes declines throughout the 20th century (Parkes and Marzeion, 2018).
Regarding regional values, Table 3 shows that roughly two-thirds of our global mass loss estimate, during the most recent
time, occurred in Greenland and the North American continent. A large amount of the global uncertainty originates from these
regions as well. Comparing our regional mass change estimates for recent years to those in the literature (Ciracì et al., 2020;
Wouters et al., 2019; Zemp et al., 2019), the most obvious discrepancy can be found in estimates for the Southern Andes,

405   where our ensemble mean is substantially less negative and even positive in earlier periods, caused by the model setup forced

with 20CRV3 reanalysis data. The opposite is true for the regions Arctic Canada (North) and Svalbard, where our estimate is more negative than those previously published. This might be caused by the relatively large portion of area draining into marine-terminating glaciers in those regions, since glacier-ocean interactions are not included in the model we applied and the calibration applying solely atmospheric forcing might thus be problematic. Finally, our regional estimate is significantly

410   more negative for Greenland than for Alaska in the most recent period, while it is not so in Zemp et al. (2019). Thus, while we find a good agreement of our global mass change estimates with previously published ones, there are significant differences in regional estimates.

   Although the largest potential of reducing the global uncertainty, relevant to e.g. sea-level rise estimates, is in largely
415   glaciated but less observed regions, reducing it in smaller regions (e.g. Southern Andes) could still be valuable concerning hydrological changes and hence water availability.

6   Conclusions

A multi-objective optimization of a global glacier mass balance model forced with an ensemble of meteorological data sets was presented. We demonstrated that it is possible to find statistically well performing model setups of model parameters for
420   each forcing data set, but that we cannot robustly identify which model setup is the most reliable when applied outside of the temporal and spatial domain of validation data. However, one data set (CERA20C) can be identified as performing worse that the others. Disagreement between ensemble members is to a large degree attributable to differences in the forcing data in times and at locations where few validation data are available. The differences in the forcing data result in diverging glacier mass loss estimates, especially in the first half of the 20th century. Regionally, the largest ensemble disagreement is found
425   regarding Greenland's peripheral glaciers. Although our estimates lie within the uncertainty range to most of the previously published global estimates, they seem to agree less with those derived from GRACE data. Finally, all ensemble members agree that around the 1930s mass loss rates from glaciers were comparable to those of today. They were followed by a phase of deceleration roughly between 1940 and 1980, and have been accelerating since then.

Data availability.   The reconstructed, optimized time series will be made available as a supplement to the publication once it is accepted.
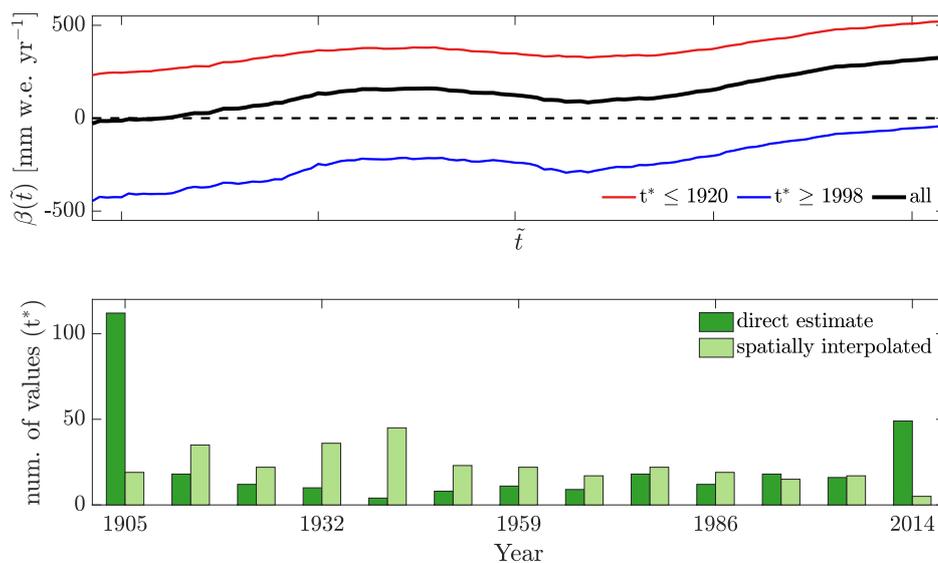
Figure 1. Upper panel: β as a function of $\tilde{t}$ for validation glaciers with $t^* \leq 1920$ (red, n = 132) and ≥ 1998 (blue, n = 72) as well as the weighted average of all validation glaciers (black, n = 297). Lower Panel: Distributions of ideal (green) and spatially interpolated (light green) t^*. Values in both panels are derived from the cross-validation procedure with the optimized CRU TS 4.03 model setup.
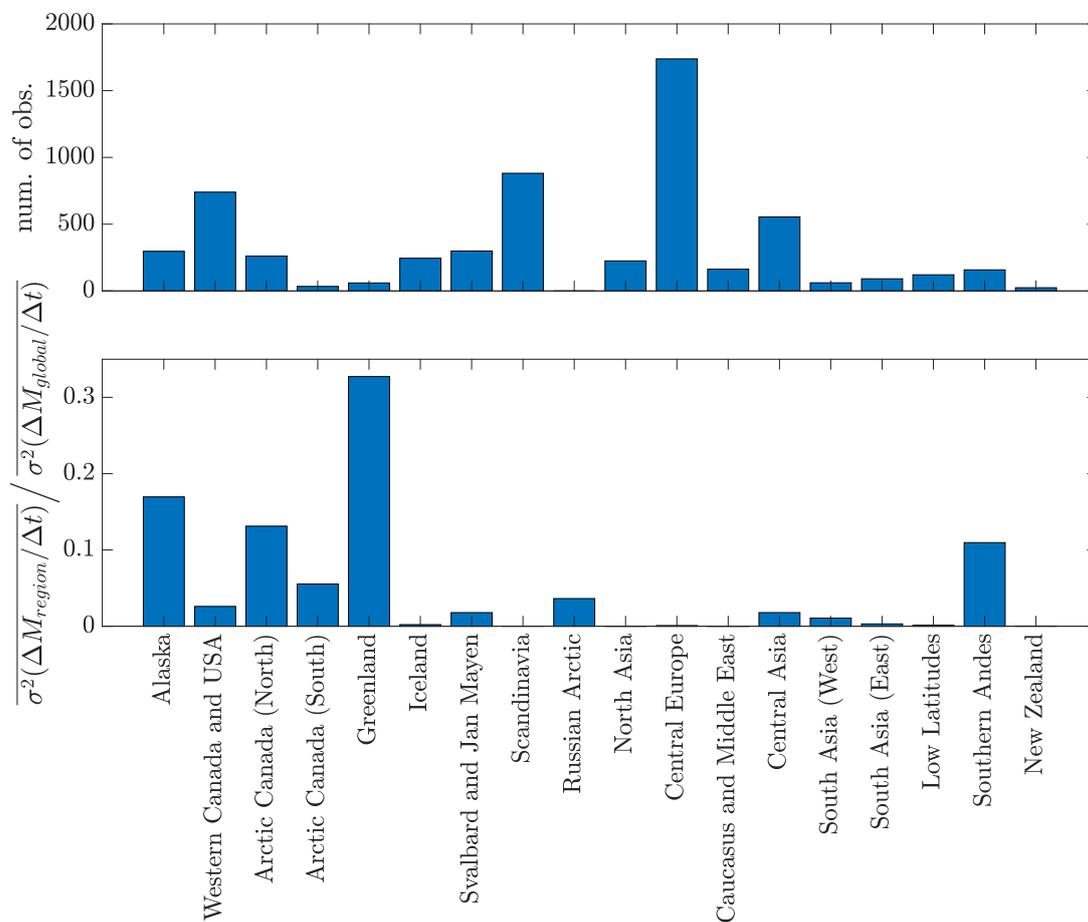
Figure 2. Upper panel: Number of mass-balance observations available for validation in each RGI region. Lower panel: Fraction of ensemble variance of global mean mass change rate (ΔM / Δt) in the modeled period attributable to each RGI region.
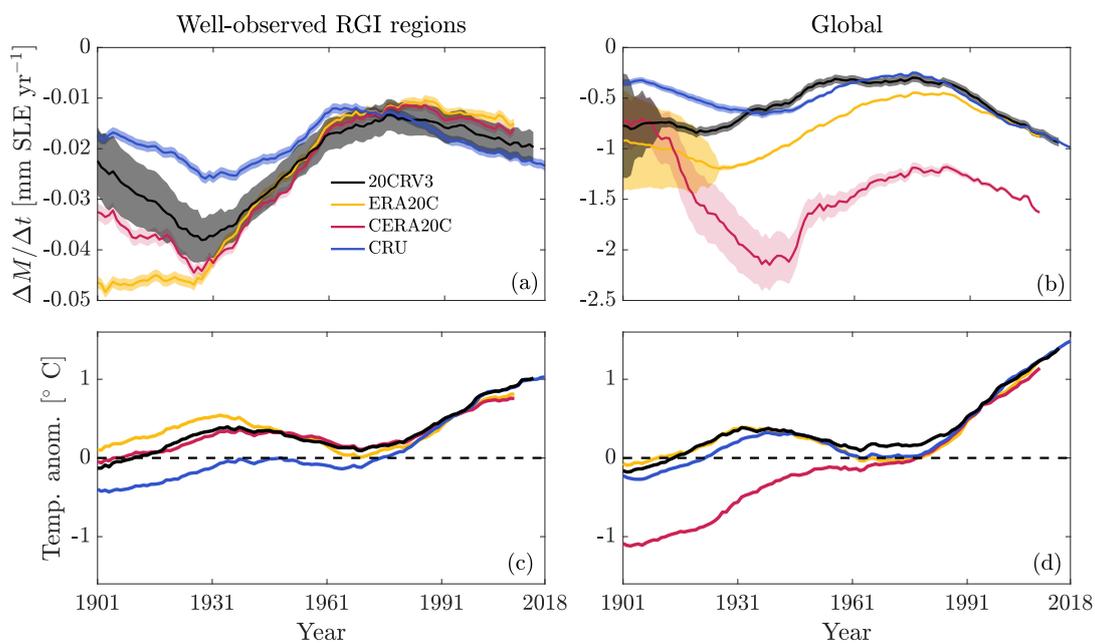
Figure 3. (a) Mass loss rate estimates for individual forcing data sets with whole 20th century coverage, averaged over well-observed regions (Western Canada and USA, Scandinavia, Central Europe, and Central Asia). (b) same as (a) but global estimates. (c) Mean temperature anomalies at glacier locations in well-observed regions. (d) same as (c), but global. In all graphs, 31-year moving averages are shown for clarity.
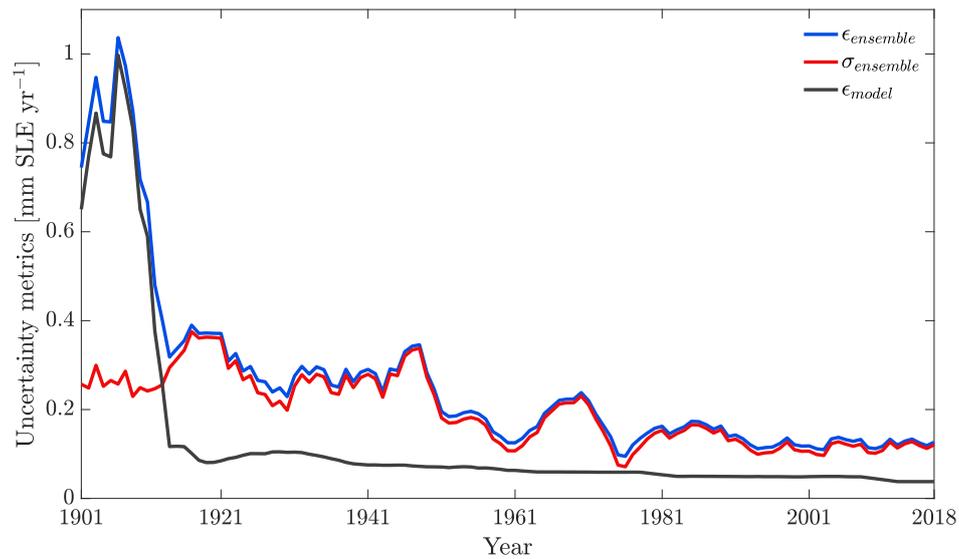
Figure 4. 5-year moving average of the temporal evolution of model uncertainty metrics for annual global mass change rates. $\epsilon_{ensemble}$ is the total uncertainty, i.e. combined model uncertainty (ϵ_{model}) and ensemble spread ($\sigma_{ensemble}$; see Eq. 10).
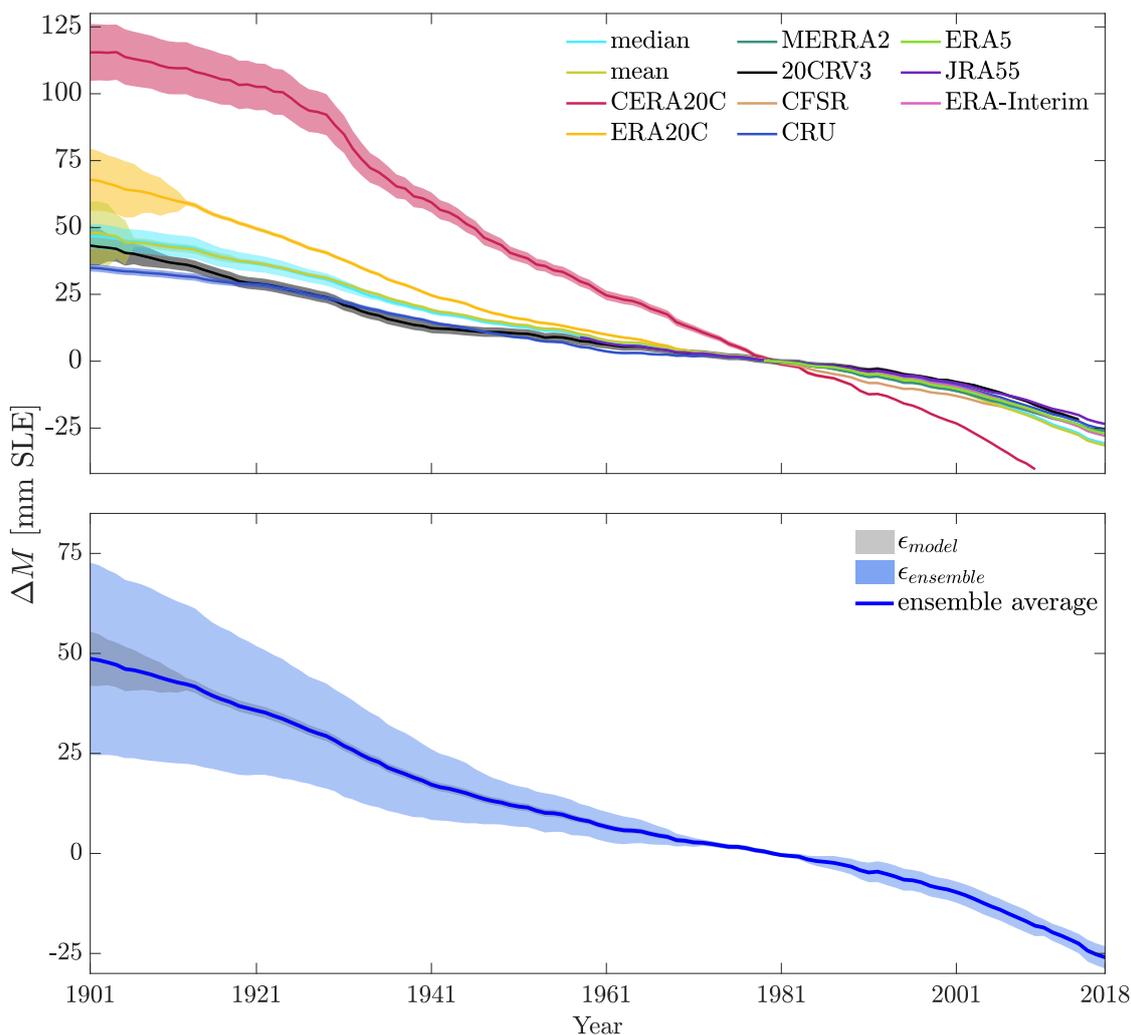
Figure 5. Upper panel: Estimates of temporally accumulated global sea-level contribution relative to 1980 for all forcing data sets. Shaded areas are model uncertainties calculated for individual model setups. Lower panel: Ensemble mean output estimate. Shaded area are the mean model uncertainty (grey, ϵ_{model}) and total ensemble uncertainty (blue, $\epsilon_{ensemble}$; see Eq. 10). Uncertainties shown at the 90% confidence level. Note the different vertical scales of the panels.
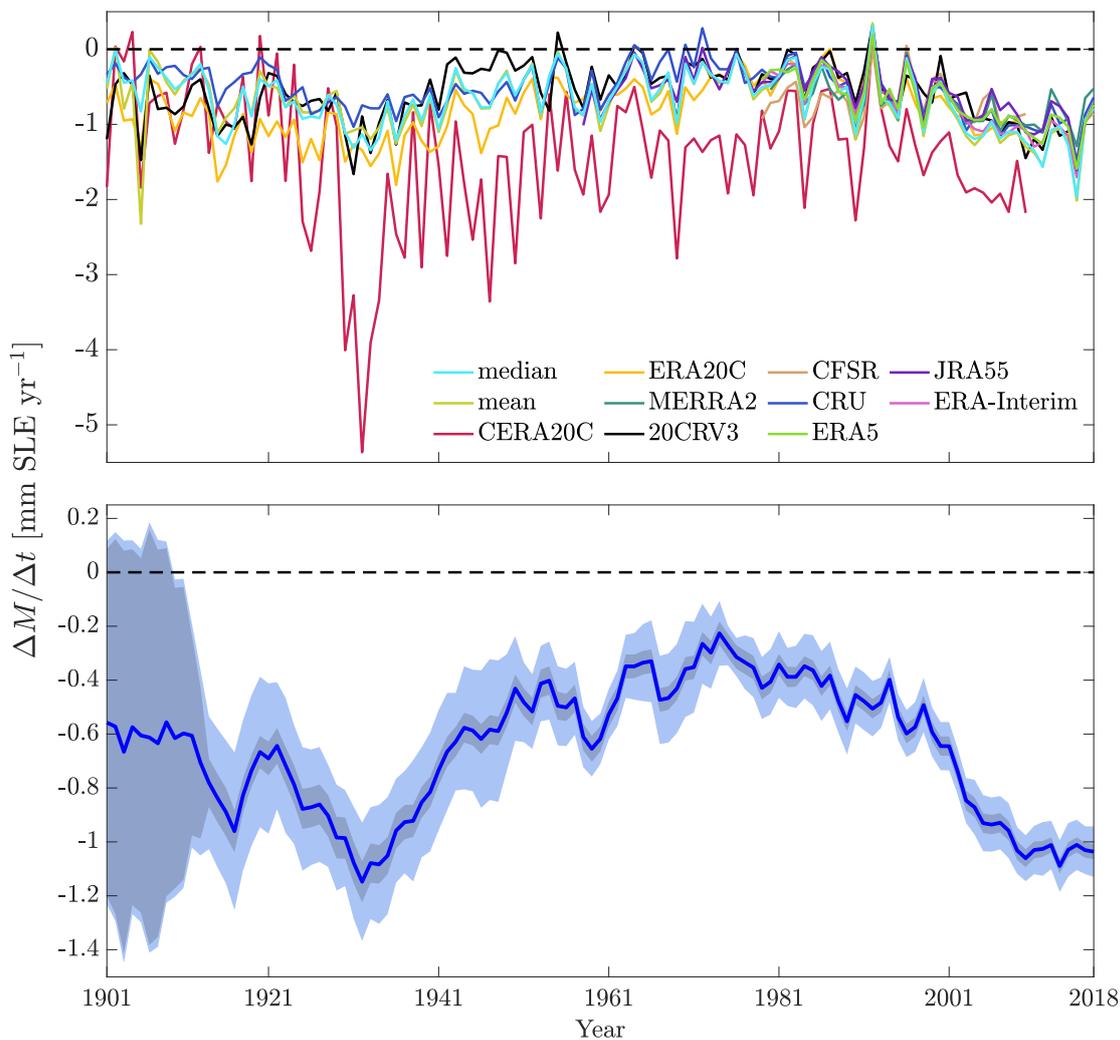
Figure 6. Upper panel: Annual glacier mass change rates for all forcing data sets. Lower panel: Mean of ensemble output mass change rates. A 5-year moving average is shown for clarity. Shaded areas are the mean model uncertainty (grey, ϵ_{model}) and total ensemble uncertainty (blue, $\epsilon_{ensemble}$; see Eq. 10). Uncertainties shown at the 90% confidence level. Note the different vertical scales of the panels.
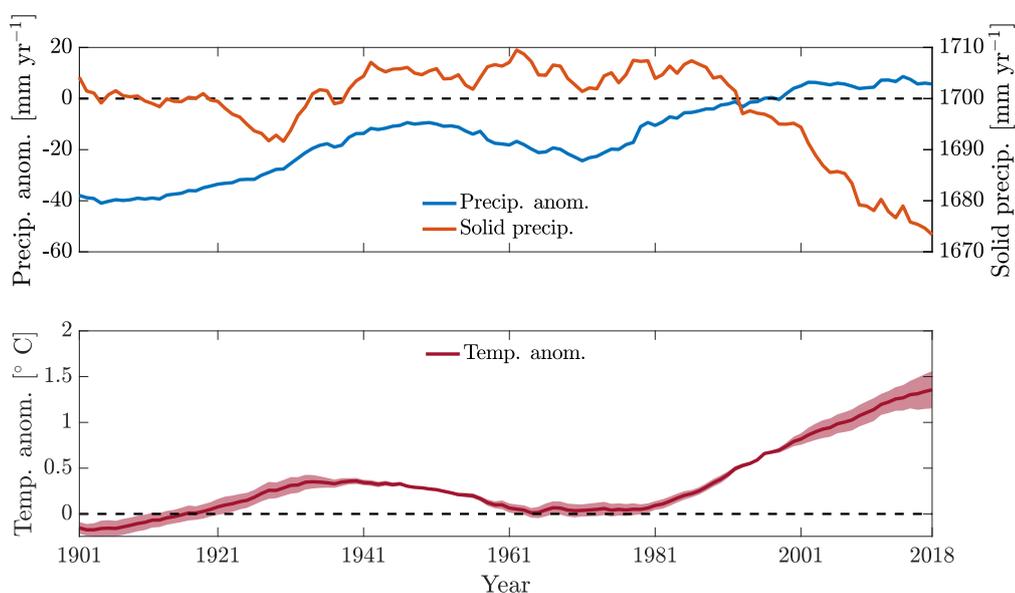
Figure 7. Upper panel: global mean annual precipitation anomaly relative to 1961 to 1990 and amount of total solid precipitation. Lower panel: global mean annual temperature anomaly relative to 1961 to 1990. The shading shows $\pm 1\ \sigma$. Values in both panels are 31-year moving averages of the ensemble mean at glacier tongue locations and weighted by glacier area, except for the graph of solid precipitation, which is based on the median forcing input data, since scales of computed solid precipitation might widely vary between ensemble members depending on model parameters (see Eq. 2 and 3), making the computation of an average, especially with a temporally varying number of ensemble members, less meaningful.

Table 1. Resolution and time range of the meteorological data sets used as boundary conditions.

Label used in text & figures	Resolution	Time range	Publication
20CRV3	2 x 2	1871 - 2014	Slivinski et al. (2019)
CFSR	0.5 x 0.5	1979 - 2010	Saha et al. (2010)
CRU CL 2.0	10' x 10'	1961-1991 (climatology)	New et al. (1999)
CRU TS 4.03	0.5 x 0.5	1901 - 2018	Harris and Jones (2020), Harris et al. (2014)
CERA20C	0.28 x 0.28	1900 - 2010	Laloyaux et al. (2018)
ERA5	0.5 x 0.5	1979 - 2018	Copernicus Climate Change Service (C3S) (2019)
ERA20C	1.13 x 1.13	1900 - 2010	Poli et al. (2016)
ERA-Interim	∼0.7 x 0.7	1979 - 2018	Dee et al. (2011)
JRA55	1.25 x 1.25	1958 - 2018	Kobayashi et al. (2015)
MERRA2	0.63 x 0.63	1980 - 2018	Gelaro et al. (2017)

Table 2. Values of the performance measures for each tested data set, applying the optimal parameter set. Values behind coefficients in brackets display non-weighted averages (see text). For the mean and median model output, the score with/without CERA20C is displayed. The total number of cross-validated glaciers used for the respective data set is n and A_M the percentage of glacier area (as recorded in the RGI) the model was able to initialize.

	Bias [mm w.e. yr^{-1}]	R	SR	RMSE [mm w.e. yr^{-1}]	score	n	A_M [%]	t*	T_{melt} [°C]	$T_{p.s.}$ [°C]	$\gamma_{p.}$ [%/100 m]	a
20CRV3	14.1 (51.4)	0.61 (0.55)	-0.02 (-0.13)	978.0 (816.9)	2.30	295	86.8	1978	2	2	1	2.5
CERA20C	79.0 (2.4)	0.56 (0.54)	0.07 (0.08)	747.1 (715.3)	1.38	274	83.8	1902	2	0	4	3
CFSR	0.0 (2.4)	0.60 (0.54)	0.13 (0.08)	804.2 (740.9)	2.13	276	93.4	1917	1	4	3	2
CRU TS 4.03	0.6 (11.6)	0.63 (0.59)	0.01 (-0.03)	739.6 (695.9)	2.62	298	97.7	1917	0	4	4	3
ERA-Interim	1.7 (18.9)	0.64 (0.62)	0.02 (0.00)	715.1 (674.0)	2.70	297	97.5	1907	0	4	3	3
ERA5	0.0 (12.1)	0.67 (0.64)	0.02 (-0.04)	705.83 (674.5)	2.96	299	97.7	1919	0	4	5	2.5
ERA20C	4.6 (19.4)	0.58 (0.56)	0.05 (0.05)	791.0 (735.1)	2.04	281	97.1	1902	0	4	1	3
JRA55	-2.0 (18.3)	0.64 (0.61)	0.00 (-0.01)	701.1 (670.5)	2.66	298	98.6	1915	-1	4	5	3
MERRA2	0.2 (16.8)	0.64 (0.61)	0.00 (-0.02)	719.7 (685.4)	2.72	299	97.5	1908	0	4	1	3
mean in.	9.1 (-4.2)	0.66 (0.63)	0.13 (0.07)	767.5 (714.5)	2.59	299	93.2	1901	1	4	3	3
median in.	15.3 (19.6)	0.66 (0.63)	0.02 (-0.03)	725.9 (679.7)	2.75	299	93.7	1903	1	4	1	3
mean out.	6.9 (11.0)	0.67 (0.64)	-0.04 (-0.10)	707.1 (654.8)	2.90/2.95	-	-	-	-	-	-	-
median out.	22.3 (17.7)	0.67 (0.64)	-0.04 (-0.10)	687.1 (646.1)	2.79/2.85	-	-	-	-	-	-	-

Table 3. Regional mass change rate estimates (in µm SLE yr^{-1}) for different time periods.

	1901 - 1940	1941 - 1980	1901 - 1990	1971 - 2018	1993 - 2018	2006 - 2018
1 Alaska	-0.08 ± 0.03	-0.05 ± 0.01	-0.06 ± 0.01	-0.09 ± 0.01	-0.13 ± 0.01	-0.14 ± 0.02
2 Western Canada and US	-0.07 ± 0.02	-0.020 ± 0.003	-0.040 ± 0.007	-0.017 ± 0.002	-0.020 ± 0.002	-0.023 ± 0.003
3 Arctic Canada (North)	-0.10 ± 0.02	-0.06 ± 0.02	-0.08 ± 0.01	-0.09 ± 0.01	-0.13 ± 0.02	-0.18 ± 0.02
4 Arctic Canada (South)	-0.08 ± 0.02	-0.034 ± 0.007	-0.055 ± 0.008	-0.036 ± 0.004	-0.050 ± 0.007	-0.07 ± 0.01
5 Greenland	-0.24 ± 0.07	-0.09 ± 0.03	-0.15 ± 0.03	-0.13 ± 0.02	-0.20 ± 0.03	-0.24 ± 0.04
6 Iceland	-0.010 ± 0.003	-0.012 ± 0.002	-0.010 ± 0.002	-0.012 ± 0.002	-0.020 ± 0.002	-0.023 ± 0.003
7 Svalbard	-0.047 ± 0.007	-0.033 ± 0.005	-0.041 ± 0.004	-0.064 ± 0.004	-0.078 ± 0.006	-0.094 ± 0.009
8 Scandinavia	-0.0034 ± 0.0007	-0.0026 ± 0.0004	-0.0026 ± 0.0004	-0.0024 ± 0.0003	-0.0044 ± 0.0004	-0.0057 ± 0.0005
9 Russian Arctic	-0.06 ± 0.10	-0.046 ± 0.007	-0.05 ± -0.04	-0.062 ± 0.006	-0.072 ± 0.007	-0.086 ± 0.0095
10 North Asia	-0.0019 ± 0.0004	-0.0004 ± 0.0002	-0.0012 ± 0.0002	-0.0016 ± 0.0002	-0.0025 ± 0.0002	-0.030 ± 0.003
11 Central Europe	-0.001 ± 0.001	-0.0007 ± 0.0010	-0.0015 ± 0.0008	-0.0040 ± 0.0006	-0.0059 ± 0.0006	-0.0056 ± 0.0006
12 Caucasus and Middle East	-0.0021 ± 0.0004	-0.0001 ± 0.0002	-0.0009 ± 0.0002	-0.0008 ± 0.0002	-0.0016 ± 0.0002	-0.0021 ± 0.0003
13 Central Asia (North)	-0.057 ± 0.007	-0.053 ± 0.004	-0.054 ± 0.004	-0.049 ± 0.003	-0.052 ± 0.004	-0.057 ± 0.005
14 Central Asia (West)	-0.044 ± 0.006	-0.041 ± 0.004	-0.041 ± 0.003	-0.034 ± 0.002	-0.034 ± 0.003	-0.037 ± 0.004
15 Central Asia (South)	-0.021 ± 0.003	-0.017 ± 0.002	-0.019 ± 0.002	-0.015 ± 0.001	-0.017 ± 0.002	-0.018 ± 0.002
16 Low Latitudes	-0.005 ± 0.010	-0.006 ± 0.001	-0.007 ± 0.002	-0.0030 ± 0.0007	-0.0039 ± 0.0008	-0.0035 ± 0.0008
17 Southern Andes	0.03 ± 0.01	0.007 ± 0.011	0.014 ± 0.008	-0.011 ± 0.007	-0.012 ± 0.008	-0.012 ± 0.008
18 New Zealand	-0.002 ± 0.001	-0.0034 ± 0.0009	-0.0025 ± 0.0007	-0.0016 ± 0.0003	-0.0013 ± 0.0004	-0.016 ± 0.0005
Global (without peripheral Antarctic and Subantarctic)	-0.80 ± 0.14	-0.46 ± 0.06	-0.61 ± 0.07	-0.62 ± 0.03	-0.84 ± 0.04	1.00 ± 0.06

Table 4. Different estimates for mean annual glacier mass losses (in mm SLE yr^{-1}) over different time periods. Uncertainties at the 90% level for all values except for Wouters et al. (2019) and Zemp et al. (2019), which are at the 95% level. Estimates exclude Greenland and Antarctic periphery; columns marked with * include Greenland periphery. Cells highlighted in bold font are published estimates disagreeing significantly with ours as discussed in section 5.

	2002 - 2018	1961 - 2016	1992 - 2016	2002 - 2016	2006 - 2016	2003 - 2009*	1961 - 2010*	1902 - 2010*
Ensemble mean	-0.72 ± 0.03	-0.45 ± 0.02	0.60 ± 0.03	-0.72 ± 0.03	-0.76 ± 0.04	-0.93 ± 0.08	-0.51 ± 0.04	-0.62 ± 0.05
Ciracì et al. (2020)	-0.78 ± 0.08	-	-	-	-	-	-	-
Zemp et al. (2019)	-	-0.4 ± 0.3	-	-	-0.74 ± 0.05	-	-	-
Bamber et al. (2018)	-	-	-0.48 ± 0.22	-	-	-	-	-
Wouters et al. (2019)	-	-	-	**-0.55 ± 0.09**	-	-	-	-
WGMS (2015) (direct)	-	-	-	-	-	-1.12	-0.57	-
WGMS (2015) (geodetic)	-	-	-	-	-	-0.85	-1.05	-
Marzeion et al. (2012)	-	-	-	-	-	-0.96 ± 0.12	-	-
Marzeion et al. (2015)	-	-	-	-	-	-0.78 ± 0.15	-0.49 ± 0.05	-0.62 ± 0.05
Cogley (2009) [+]	-	-	-	-	-	**-0.75 ± 0.07**	-0.54 ± 0.05	-
Leclercq et al. (2011) [+]	-	-	-	-	-	-0.87 ± 0.64	-0.58 ± 0.15	-0.78 ± 0.19
Gardner et al. (2013)	-	-	-	-	-	**-0.70 ± 0.07**	-	-

[+] Updated in Marzeion et al. (2017)

430 *Author contributions.* B. Marzeion designed the research and contributed to the manuscript. J.-H. Malles contributed to designing the research, conducted the simulations and statistical evaluation, and wrote the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. This work was supported through the Sea Level Budget Closure Project (SLBC_cci) as part of the ESA Climate Change Initiative (CCI) Programme. BM was supported by the Deutsche Forschungsgemeinschaft (DFG), grant no. MA6966/4-1. JHM was sup-
435 ported by the DFG through the International Research Training Group "Processes and impacts of climate change in the North Atlantic Ocean and the Canadian Arctic" (IRTG 1904/3 ArcTrain).

References

Bahr, D. B., Pfeffer, W. T., and Kaser, G.: A review of volume-area scaling of glaciers, Reviews of Geophysics, 53, 95–140, https://doi.org/10.1002/2014RG000470, 2015.

440 Bamber, J. L., Westaway, R. M., Marzeion, B., and Wouters, B.: The land ice contribution to sea level during the satellite era, Environmental Research Letters, 13, 063 008, https://doi.org/10.1088/1748-9326/aac2f0, https://doi.org/10.1088%2F1748-9326%2Faac2f0, 2018.

Ciracì, E., Velicogna, I., and Swenson, S.: Continuity of the Mass Loss of the World's Glaciers and Ice Caps From the GRACE and GRACE Follow-On Missions, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL086926, 2020.

Cogley, J. G.: Geodetic and direct mass-balance measurements: comparison and joint analysis, Annals of Glaciology, 50, 96–100, 445 https://doi.org/10.3189/172756409787769744, 2009.

Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service Climate Data Store (CDS), 2019.

Cruz, R., Harasawa, H., Lal, M., Wu, S., Anokhin, Y., Punsalmaa, B., Honda, Y., Jafari, M., Li, C., and Huu Ninh, N.: Asia. Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental 450 Panel of Climate Change, Journal of Environment Quality, https://doi.org/10.2134/jeq2008.0015br, 2007.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: 455 configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Frederikse, T., Landerer, F., Caron, L., Adhikari, S., Parkes, D., Humphrey, V. W., Dangendorf, S., Hogarth, P., Zanna, L., Cheng, L., and Wu, Y.-H.: The causes of sea-level rise since 1900, Nature, 584, 393–397, https://doi.org/10.1038/s41586-020-2591-3, 2020.

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., Ligtenberg, 460 S. R. M., Bolch, T., Sharp, M. J., Hagen, J. O., van den Broeke, M. R., and Paul, F.: A Reconciled Estimate of Glacier Contributions to Sea Level Rise: 2003 to 2009, Science, 340, 852–857, https://doi.org/10.1126/science.1234532, 2013.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, 465 B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Giesen, R. H. and Oerlemans, J.: Climate-model induced differences in the 21st century global and regional glacier contributions to sea-level rise, Climate Dynamics, 41, 3283–3300, https://doi.org/10.1007/s00382-013-1743-7, 2013.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research: Atmospheres, 470 113, https://doi.org/10.1029/2007JD008972, 2008.

Goosse, H., Barriat, P.-Y., Dalaiden, Q., Klein, F., Marzeion, B., Maussion, F., Pelucchi, P., and Vlug, A.: Testing the consistency between changes in simulated climate and Alpine glacier length over the past millennium, Climate of the Past, 14, 1119–1133, https://doi.org/10.5194/cp-14-1119-2018, https://cp.copernicus.org/articles/14/1119/2018/, 2018.

Harris, I. and Jones, P.: CRU TS4.03: Climatic Research Unit (CRU) Time-Series (TS) version 4.03 of high-resolution
475   gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2018), Centre for Environmental Data Analysis,
       https://doi.org/10.5285/10d3e3640f004c578403419aac167d82, 2020.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Upyeard high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset,
       International Journal of Climatology, 34, 623–642, 2014.

Hirabayashi, Y., Zang, Y., Watanabe, S., Koirala, S., and Kanae, S.: Projection of glacier mass changes under a high-emission climate scenario
480   using the global glacier model HYOGA2, Hydrological Research Letters, 7, 6–11, https://doi.org/10.3178/hrl.7.6, 2013.

Hock, R., Bliss, A., Marzeion, B., Giesen, R. H., Hirabayashi, Y., Huss, M., Radić, V., and Slangen, A. B.: GlacierMIP–A model intercom-
       parison of global-scale glacier mass-balance models and projections, Journal of Glaciology, 65, 453–467, 2019.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C.,
       Tomassini, L., Watanabe, M., and Williamson, D.: The Art and Science of Climate Model Tuning, Bulletin of the American Meteorological
485   Society, 98, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1, 2017.

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Frontiers in Earth Science, 3, 54,
       https://doi.org/10.3389/feart.2015.00054, 2015.

Huss, M. and Hock, R.: Global-scale hydrological response to future glacier mass loss, Nature Climate Change, 8, 135–140,
       https://doi.org/10.1038/s41558-017-0049-x, 2018.

490 Kaser, G., Großhauser, M., and Marzeion, B.: Contribution potential of glaciers to water availability in different climate regimes, Proceedings
       of the National Academy of Sciences, 107, 20 223–20 227, https://doi.org/10.1073/pnas.1008162107, 2010.

Kobayashi, S., Ota, Y. Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and
       Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan.
       Ser. II, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

495 Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Broennimann, S., Buizza, R., Dalhgren, P., Dee, D., Haimberger, L., Hersbach,
       H., Kosaka, Y., Martin, M., Poli, P., Rayner, N., Rustemeier, E., and Schepers, D.: CERA-20C: A Coupled Reanalysis of the Twentieth
       Century, Journal of Advances in Modeling Earth Systems, 10, 1172–1195, https://doi.org/10.1029/2018MS001273, 2018.

Leclercq, P. W., Oerlemans, J., and Cogley, J. G.: Estimating the Glacier Contribution to Sea-Level Rise for the Period 1800–2005, Surveys
       in Geophysics, 32, 19, https://doi.org/10.1007/s10712-011-9121-7, 2011.

500 Marler, R. and Arora, J.: Survey of multi-objective optimization methods for engineering, Structural and Multidisciplinary Optimization, 26,
       369–395, https://doi.org/10.1007/s00158-003-0368-6, 2004.

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

Marzeion, B., Leclercq, P. W., Cogley, J. G., and Jarosch, A. H.: Brief Communication: Global reconstructions of glacier mass change during
505   the 20th century are consistent, The Cryosphere, 9, 2399–2404, https://doi.org/10.5194/tc-9-2399-2015, 2015.

Marzeion, B., Champollion, N., Haeberli, W., Langley, K., Leclercq, P., and Paul, F.: Observation-Based Estimates of Global Glacier Mass
       Change and Its Contribution to Sea-Level Change, Surveys in Geophysics, 38, 105–130, https://doi.org/10.1007/s10712-016-9394-y,
       2017.

Marzeion, B., Hock, R., Anderson, B., Bliss, A., Champollion, N., Fujita, K., Huss, M., Immerzeel, W. W., Kraaijenbrink, P., Malles, J.-H.,
510   Maussion, F., Radić, V., Rounce, D. R., Sakai, A., Shannon, S., van de Wal, R., and Zekollari, H.: Partitioning the Uncertainty of Ensemble

Projections of Global Glacier Mass Change, Earth's Future, 8, https://doi.org/10.1029/2019EF001470, https://agupubs.onlinelibrary.wiley. com/doi/abs/10.1029/2019EF001470, e2019EF001470 10.1029/2019EF001470, 2020.

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T., and Marzeion, B.: The Open Global Glacier Model (OGGM) v1.1, Geoscientific Model Development, 12, 909–931, https://doi.org/10.5194/gmd-12-909-2019, https://gmd.copernicus.org/articles/12/909/2019/, 2019.

New, M., Hulme, M., and Jones, P.: Representing Twentieth-Century Space–Time Climate Variability. Part I: Development of a 1961–90 Mean Monthly Terrestrial Climatology, Journal of Climate, 12, 829–856, https://doi.org/10.1175/1520-0442(1999)012<0829:RTCSTC>2.0.CO;2, 1999.

New, M., Lister, D., Hulme, M., and Ian Makin, I.: A high-resolution data set of surface climate over global land areas, Climate Research, 21, 1–25, 2002.

Ohmura, A.: Changes in mountain glaciers and ice caps during the 20th century, Annals of Glaciology, 43, 361–368, https://doi.org/10.3189/172756406781812212, 2006.

Ohmura, A., Bauder, A., Müller, H., and Kappenberger, G.: Long-term change of mass balance and the role of radiation, Annals of Glaciology, 46, 367–374, https://doi.org/10.3189/172756407782871297, 2007.

Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A., Abd-Elgawad, A., Cai, R., CifuentesJara, M., DeConto, R., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., and Sebesvari, Z.: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)], In press, 2019.

Parkes, D. and Goosse, H.: Modelling regional glacier length changes over the last millennium using the Open Global Glacier Model, The Cryosphere, 14, 3135–3153, https://doi.org/10.5194/tc-14-3135-2020, https://tc.copernicus.org/articles/14/3135/2020/, 2020.

Parkes, D. and Marzeion, B.: Twentieth-century contribution to sea-level rise from uncharted glaciers, Nature, 563, 551–554, https://doi.org/10.1038/s41586-018-0687-9, 2018.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C: An Atmospheric Reanalysis of the Twentieth Century, Journal of Climate, 29, 4083–4097, https://doi.org/10.1175/JCLI-D-15-0556.1, 2016.

Radić, V. and Hock, R.: Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise, Nature Geoscience, 4, 91–94, 2011.

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

Roe, G. H., Christian, J. E., and Marzeion, B.: On the attribution of industrial-era glacier mass loss to anthropogenic climate change, The Cryosphere Discussions, 2020, 1–27, https://doi.org/10.5194/tc-2020-265, https://tc.copernicus.org/preprints/tc-2020-265/, 2020.

Rye, C. J., Willis, I. C., Arnold, N. S., and Kohler, J.: On the need for automated multiobjective optimization and uncertainty estimation of glacier mass balance models, Journal of Geophysical Research: Earth Surface, 117, https://doi.org/10.1029/2011JF002184, 2012.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y.,

Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, Bulletin of the American
550   Meteorological Society, 91, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1, 2010.

Shannon, S., Smith, R., Wiltshire, A., Payne, T., Huss, M., Betts, R., Caesar, J., Koutroulis, A., Jones, D., and Harrisson, S.: Global glacier
    volume projections under high end climate change scenarios, The Cryosphere, 1, 325–350, https://doi.org/10.5194/tc-13-325-2019, 2019.

Shugar, D. H., Burr, A., Haritashya, U. K., Kargel, J. S., C, S. W., Kennedy, M. C., Bevington, A. R., Betts, R. A., Harrison, S., and Strattman,
    K.: Rapid worldwide growth of glacial lakes since 1990, Nature Climate Change, pp. 1–7, https://doi.org/10.1038/s41558-020-0855-4,
555   2020.

Slangen, A. B. A., Meyssignac, B., Agosta, C., Champollion, N., Church, J. A., Fettweis, X., Ligtenberg, S. R. M., Marzeion, B., Melet, A.,
    Palmer, M. D., Richter, K., Roberts, C. D., and Spada, G.: Evaluating Model Simulations of Twentieth-Century Sea Level Rise. Part I:
    Global Mean Sea Level Change, Journal of Climate, 30, 8539–8563, https://doi.org/10.1175/JCLI-D-17-0110.1, 2017.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H.,
560   Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-
    Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Blancq, F. L., Lee, T.-C., Lorrey, A.,
    Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz,
    B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: Towards a more reliable historical reanalysis:
    Improvements for version 3 of the Twentieth Century Reanalysis system, Quarterly Journal of the Royal Meteorological Society, 145,
565   2876–2908, https://doi.org/10.1002/qj.3598, 2019.

Small, C. and J. Nicholls, R.: A Global Analysis of Human Settlement in Coastal Zones, Journal of Coastal Research, 19, 2003.

Taylor, K.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research, 106, 7183–7192,
    2001.

WGMS: Global Glacier Change Bulletin No. 1 (2012-2013), Zemp, M., Gärtner-Roer, I., Nussbaumer, S.U., Hüsler, F., Machguth, H.,
570   Mölg, N., Paul F., and Hoelzle, M. (eds.), ICSU(WDS)/IUGG(IACS)/UNEP/UNESCO/WMO, World Glacier Monitoring Service, Zurich,
    Switzerland, p. 230, 2015.

WGMS: Fluctuations of Glaciers Database, World Glacier Monitoring Service, Zurich, Switzerland, https://doi.org/10.5904/wgms-fog-2018-
    11, 2018.

WGMS: Global Glacier Change Bulletin No. 3 (2016-2017), Zemp, M., Gärtner-Roer, I., Nussbaumer, S. U., Bannwart, J., Rastner, P., Paul,
575   F., and Hoelzle, M. (eds.), ISC(WDS)/IUGG(IACS)/UNEP/UNESCO/WMO, World Glacier Monitoring Service, Zurich, Switzerland,
    https://doi.org/10.5904/wgms-fog-2020-08, 2020.

Wijngaard, R. R., Biemans, H., Lutz, H., Shrestha, A. B., Wester, P., and Immerzeel, W.: Climate change vs. socio-
    economic development:understanding the future South Asian water gap, Hydrology and Earth System Sciences, 22, 6297–6321,
    https://doi.org/10.5194/hess-22-6297-2018, 2018.

580 Wild, M.: Enlightening Global Dimming and Brightening, Bulletin of the American Meteorological Society, 93, 27–37,
    https://doi.org/10.1175/BAMS-D-11-00074.1, 2012.

Wouters, B., Gardner, A. S., and Moholdt, G.: Global Glacier Mass Loss During the GRACE Satellite Mission (2002-2016), Frontiers in
    Earth Science, 7, 96, https://doi.org/10.3389/feart.2019.00096, 2019.

Zemp, M., Huss, M., Thibert, E., Eckert, N., McNabb, R., Huber, J., Barandun, M., Machguth, H., Nussbaumer, S. U., Gärtner-Roer, I.,
585   Thomson, L., Paul, F., Maussion, F., Kutuzov, S., and Cogley, J. G.: Global glacier mass changes and their contributions to sea-level rise
    from 1961 to 2016, Nature, 568, 382–386, https://doi.org/10.1038/s41586-019-1071-0, 2019.