



The Potential of the Ensemble Kalman Filter to Improve Glacier Modeling

Logan Knudsen¹ · Hannah Park-Kaufmann²  · Emily Corcoran³ · Alexander Robel⁴ · Talea Mayo⁵

Received: 22 May 2023 / Revised: 26 March 2024 / Accepted: 30 April 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Using a simplified two-stage ice sheet model, we explore the potential of statistical data assimilation methods to improve predictions of glacier melt, which has significant implications for reducing uncertainty in projections of sea level rise. Through twin experiments utilizing artificial data, we find that the ensemble Kalman filter improves simulations of glacier evolution initialized with incorrect initial conditions and parameters, providing us with better predictions of future glacier melt. We explore the number of observations necessary to produce an accurate model run. We also explore optimal observation assimilation schemes, and determine that deviations from the true glacier response that stem from having few data points in the pre-satellite era can be corrected with modern observation data. Our results show that statistical data

✉ Hannah Park-Kaufmann
hk9622@bard.edu

Logan Knudsen
loganpknudsen@tamu.edu

Emily Corcoran
efc24@njit.edu

Alexander Robel
alexander.robelt@eas.gatech.edu

Talea Mayo
talea.mayo@emory.edu

¹ Department of Mathematics, Texas A&M University, College Station, Texas, USA

² Department of Mathematics and Conservatory of Music, Bard College, Annandale-On-Hudson, New York, USA

³ Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey, USA

⁴ School of Earth & Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

⁵ Department of Mathematics, Emory University, Atlanta, Georgia, USA

assimilation methods have great potential to improve complex glacier models using real-world observations.

Keywords Glacier modeling · Data assimilation · Kalman filter · Dynamical systems

1 Introduction

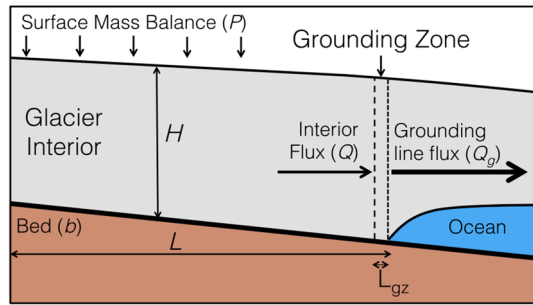
Climate change has and will continue to have a number of widespread impacts, including sea level rise. Sea level rise results in increased coastal flooding, e.g. “sunny day floods” and less frequent but more severe hurricane storm surges [6, 17]. Sea level rise is due to a number of factors, one of which is glacier melt [21]. The future glacier contribution to sea level rise is mainly explored through glacier modeling, and specifically through simulation of marine-terminating glaciers, which flow towards the ocean and then melt or fracture [23]. By the year 2300, the Antarctic ice sheet could cause 2.4–5.8 m of sea level rise globally assuming “business as usual,” i.e. if greenhouse gas and/or aerosol emissions do not decrease [20]. Due to the severe impacts of glacial melting, modeling changes in ice sheets is an important task.

Marine-terminating glaciers are susceptible to rapid retreat through a positive feedback process that can occur when the bed (solid earth) that they rest on decreases in elevation towards the ice sheet interior. In such a scenario, initial retreat causes the grounding line (the point where glacier ice starts floating in the ocean) to become thicker and therefore discharge more ice. This increased sink of ice mass causes further glacier retreat and a continuation of this feedback process [24]. Many marine-terminating glaciers in Greenland have already undergone periods of rapid retreat due to this feedback process [7], and much larger glaciers in Antarctica are thought to be susceptible to this feedback as well [19]. Models of glacier flow and evolution have been developed to simulate these processes and project potential future glacier retreats that could contribute to sea level rise. However, large uncertainties in the initial state and parameters in such models exist due to both the inherent difficulty of measuring glacier properties in remote field locations and the lack of observations of ice sheets state prior to the modern satellite era, i.e. in the 1950s. Data assimilation is one method that may provide a promising way forward to improve glacier models using existing observations. The method may also provide insight into site selection for future observational campaigns.

Data assimilation is a method to move numerical models closer to reality by using real-world observations to readjust the model state at specified times [10]. Data assimilation was initially developed for use in weather models in the 1960s in order to improve weather forecast accuracy on short-range weather models [3]. These models would typically have a run time of 1–6 h, and during that process real-time observations were assimilated to readjust the current model run. Since its introduction, data assimilation has been incorporated into other geoscientific numerical models, including ocean and land surface models [1, 5, 18].

Data assimilation can also play a factor in decision making when it comes to collecting field data. In the field of glaciology, data is largely collected via satellite, aircraft, or in-situ field missions. While these methods collect valuable data for mod-

Fig. 1 Schematic of the two-stage model of a marine-terminating glacier from [22]. The bed geometry shown in schematic is purely illustrative



eling glaciers, they are expensive and time-consuming. Data assimilation can be used to inform the locations and timing of data collection efforts. This can help researchers more efficiently utilize funding and avoid unnecessarily expensive data collection that does not significantly improve glacier models.

Recently, data assimilation has been used to improve glacier modeling. [12] developed methods for transient calibration of a glacier model state and parameters using a model adjoint in conjunction with variational assimilation methods. The use of ensemble-type data assimilation methods with glacier models has been explored in studies with relatively simple models and idealized case studies. [13] and [4] used an ensemble Kalman filter with highly idealized glacier models to show the utility of this method for determining where observations need to be gathered to maximally reduce projection uncertainty. A more recent study by [11] used an ensemble Kalman filter to calibrate glacier state and parameters governing basal topography and friction in an idealized marine-terminating glacier flowline model on the verge of undergoing the marine ice sheet instability. In this work, we implement an ensemble Kalman filter with a simplified marine-terminating glacier model. Our goal is to determine which model parameters and state variables are optimal targets for data assimilation and sufficient ensemble sizes and observation times for improving model accuracy and reducing model uncertainty.

2 Methods

2.1 Glacier Model

We have chosen a simplified two-stage ice sheet model for our exploration (Fig. 1). This model describes the changes in ice mass of a marine-terminating glacier, which is a glacier that is fed by snowfall on land and then flows into the ocean where it may begin to float. Marine-terminating glaciers may change over time due to climate change through increasing melt at the glacier surface or interface with the ocean [22].

A marine-terminating glacier can be represented with a simplified two-box model that introduces a nested box into the system to more accurately represent the grounding line, i.e. the point at which the glacier first begins floating in seawater. In this model, the grounding line position is L and the ice thickness at the grounding line is h_g . The

interior of the glacier has a thickness, H . The balance of mass inputs and outputs in the glacier interior is described by the snowfall rate on the surface, P , and the ice flux out of the interior box and across the grounding line, Q . The ice flux across (out of) the grounding line is denoted Q_g . The change in length and height of the glacier with respect to time, t , can then be described with the differential equations:

$$\frac{dH}{dt} = P - \frac{Q_g}{L} - \frac{H}{h_g L} (Q - Q_g) \quad (1)$$

$$\frac{dL}{dt} = \frac{1}{h_g} (Q - Q_g). \quad (2)$$

The following diagnostic equations are also used in this model:

$$h_g = -\lambda b(L) \quad (3)$$

$$Q = \gamma \frac{H^{2n+1}}{L^n} \quad (4)$$

$$Q_g = \Omega h_g^\beta, \quad (5)$$

where γ and Ω are assumed to be constants describing ice flow rate for our purposes, and $\lambda = \frac{\rho_w}{\rho_i}$, where ρ_w is seawater density and ρ_i is glacial ice density. Further, b is the depth of the bed below sea level at the grounding line, β can be derived from asymptotic boundary layer analysis of the grounding line, and $n \in \mathbb{N}$ assumes the value of Glen's flow law exponent $n = 3$. Despite the simplicity of this model, it offers a sufficiently good approximation of glacier flow using Q and Q_g , and has been shown to reproduce the behavior of a more complex glacier flow model [22]. We thus expect that findings determined using this simplified model may be used to inform applications of more complex models.

Important initial conditions and parameters used in the model are smb_o , smb_1 , and smb_f , which are used to describe the surface mass balance at three distinct points in time and define the spatially averaged surface mass balance, P , in Equation 1; H_o and L_o , which define initial conditions for height and length at time (year) 0; b_x , which defines the slope of the Earth underneath the glacier; and $sill_{slope}$, $sill_{min}$, and $sill_{max}$, which define the slope of the glacial sill (region of reverse slope) and its starting ($sill_{min}$) and end ($sill_{max}$) points. For a more in-depth explanation and justification of the model, see [22]. The differential equations describing glacier behavior, Equations 1-5, are solved using nominal values of these initial conditions and parameters and a 4th order Runge–Kutta method to advance the model in time.

2.2 Sensitivity Analysis

As our aim is to improve the model through data assimilation, we first conduct sensitivity analyses to explore which parameters most significantly affect the model output. Sensitivity analyses allow exploration of how various sources of uncertainty in the mathematical model may contribute to the model's overall uncertainty [16]. We vary

nominal values of the model parameters by $\pm 10\%$, with the understanding that if the outputs vary significantly as a result, the output is considered sensitive to the specification of the inputs. This provides insight into the most promising applications of data assimilation. For this analysis, the parameters are grouped together into three categories: initial conditions (initial height, length, and bed slope), sill parameters (sill minimum, maximum, and slope), and surface mass balance parameters (initial, mid, and final surface mass balance). The parameters in each category are perturbed together, as they are related to one another. In section 3.1, we analyze the spread of the glacier heights and lengths produced from the range of perturbed values for each of the three categories and use our findings to direct the data assimilation experiments.

2.3 Data Assimilation

The Kalman filter is a data assimilation technique that uses a linear model, observations, and corresponding error covariance matrices to update model output to better reflect reality [8, 9, 15]. Specifically, the goal of the Kalman filter is to compute an optimal estimate (analysis) of the model state at time t , $\mathbf{x}_t^a$, as a combination of the model output, $\mathbf{x}_t^f$, and observations, $\mathbf{y}_t$ (Equation 6). The key to this is the Kalman Gain matrix, $\mathbf{K}_t^a$, which decides the balance of how much this combination relies on the model output vs. the observations, i.e. it is used to determine the optimal weight of each based on the error covariance matrices. The linear observation operator, $\mathbf{H}_t$, allows the model state to be compared to the observations by projecting it into the observation space. In practice the observation operator can be an identity matrix, e.g. when the locations of observations and model output are the same.

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t^a (\mathbf{y}_t - \mathbf{H}_t \mathbf{x}_t^f) \quad (6)$$

2.3.1 Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) is a version of the Kalman filter better suited for large, nonlinear models such as those used in glacier modeling. Here, the model state is represented by an ensemble, and the model error covariance matrix is represented by the sample covariance of this ensemble. An analysis is performed for each ensemble member.

Consider the following state estimation system:

$$\mathbf{x}_t^{f,(i)} = \mathcal{M} \mathbf{x}_{t-1}^{a,(i)} + w_t^{(i)} \quad (7)$$

$$\mathbf{P}_t^f = \frac{1}{N-1} \sum_{i=0}^N [\mathbf{x}_t^{f,(i)} - \bar{\mathbf{x}}_t^f][\mathbf{x}_t^{f,(i)} - \bar{\mathbf{x}}_t^f]^T. \quad (8)$$

Here, $\mathcal{M}$ represents our nonlinear glacier model and $\mathbf{x}_{t-1}^{a,(1)}, \mathbf{x}_{t-1}^{a,(2)}, \dots, \mathbf{x}_{t-1}^{a,(N)}$, represent the ensemble of N model states at time $t-1$. Unknown model error is denoted by $w_t^{(i)}$. We initialize the state ensemble by choosing values $\mathbf{x}_{t-1}^{a,(i)}$ normally distributed around the estimated value of $\mathbf{x}_{t-1}^a$ with a covariance chosen to reflect the uncertainty

in the state. The forecast of each ensemble member, $\mathbf{x}_t^{f,(i)}$, is output by the model at time t , and used to compute the forecast error covariance, $\mathbf{P}_t^f$, where $\mathbf{x}_t^f$ is the mean of the ensemble of forecasted states.

Given $\mathbf{y}_t$, an observation of the state at time t , we add noise, $\mathbf{v}_t^{(i)}$, $i = 1, \dots, N$, to develop an ensemble of N observations, $\mathbf{y}_t^{(1)}, \mathbf{y}_t^{(2)}, \dots, \mathbf{y}_t^{(N)}$. The noise on our observations is assumed to be normally distributed around 0, and is chosen so that it reflects the known measurement error covariance, $\mathbf{R}_t$.

The three equations that describe the analysis step are as follows:

$$\mathbf{x}_t^{a,(i)} = \mathbf{x}_t^{f,(i)} + \mathbf{K}_t^a (\mathbf{y}_t^{(i)} - \mathbf{H}_t \mathbf{x}_t^{f,(i)}) \quad (9)$$

$$\mathbf{K}_t^a = \mathbf{P}_t^f \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t^T + \mathbf{R}_t)^{-1} \quad (10)$$

$$\mathbf{P}_f^a = \frac{1}{N-1} \sum_{i=0}^N [\mathbf{x}_t^{a,(i)} - \mathbf{x}_t^a][\mathbf{x}_t^{a,(i)} - \mathbf{x}_t^a]^T, \quad (11)$$

where $\mathbf{x}_t^a$ is the mean of the ensemble of analyzed states.

Algorithm 1 outlines the process we use for a model at time $t \in T_{obs} = \{1, \dots, T\}$ when data are available for a set of times, T_{obs} . The algorithm is illustrated in Fig. 2.

Algorithm 1 Ensemble Kalman Filter

```

Generate  $\mathbf{x}_0^{a,(0)}, \mathbf{x}_0^{a,(1)}, \dots, \mathbf{x}_0^{a,(N)}$ 
for  $t = 0, 1, \dots, T$  do
  if  $t \in T_{obs}$  then
    Calculate  $\mathbf{P}_t^f$ 
    Calculate  $\mathbf{R}_t$ 
    Calculate  $\mathbf{K}_t^a = \mathbf{P}_t^f \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_t^f \mathbf{H}_t^T + \mathbf{R}_t)^{-1}$ 
    for  $i = 0, 1, \dots, N$  do
       $\mathbf{x}_t^{f,(i)} = \mathcal{M} \mathbf{x}_{t-1}^{(i)}$ 
       $\mathbf{y}_t^{(i)} = \mathbf{y}_t + \mathbf{v}_t^{(i)}$ 
       $\mathbf{x}_t^{a,(i)} = \mathbf{x}_t^{f,(i)} + \mathbf{K}_t^a (\mathbf{y}_t^{(i)} - \mathbf{H}_t \mathbf{x}_t^{f,(i)})$ 
    end for
    Calculate  $\mathbf{x}_t^a = \frac{1}{N} \sum_{i=0}^N \mathbf{x}_t^{a,(i)}$ 
    Calculate  $\mathbf{P}_f^a = \frac{1}{N-1} \sum_{i=0}^N [\mathbf{x}_t^{f,(i)} - \mathbf{x}_t^a][\mathbf{x}_t^{f,(i)} - \mathbf{x}_t^a]'$ 
  else
    for  $i = 0, 1, \dots, N$  do
       $\mathbf{x}_t^{(i)} = \mathcal{M} \mathbf{x}_{t-1}^{(i)}$ 
    end for
    Calculate  $\mathbf{x}_t^a = \frac{1}{N} \sum_{i=0}^N \mathbf{x}_t^{a,(i)}$ 
    Calculate  $\mathbf{P}_f = \frac{1}{N-1} \sum_{i=0}^N [\mathbf{x}_t^{(i)} - \mathbf{x}_t^a][\mathbf{x}_t^{(i)} - \mathbf{x}_t^a]'$ 
  end if
end for

```

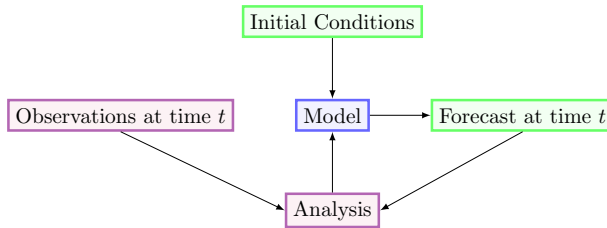


Fig. 2 The code structure diagram, showing the relationship between the aspects of the model, with the arrows depicting the flow of data. The model takes initial conditions as input and outputs a forecast at time t (green). There is a cyclic relationship between the modeling step (blue), the data assimilation and the analysis (red), which allows for continuous refinement and improvement of the model based on the latest observations and efficient use of observational data

2.3.2 Twin Experiments

Like many proof-of-concept data assimilation studies, we conduct observation system simulation experiments (OSSEs), or twin experiments, to assess the potential of the EnKF to improve glacier modeling. This has two main advantages, namely that we are not required to acquire real-world data and can also isolate and address various sources of model error. This provides a mechanism to directly assess the ability of the EnKF to improve the model, serving as a foundational proof of concept for applications to larger glacier models and real-world data.

The method is as follows: We first set initial conditions and parameters of the model to specific values informed by the sensitivity analysis described above. We run the model, and use the model output as synthetic measurements. We perturb the synthetic measurements by adding noise drawn from a random normal distribution to generate noisy “observations.” We then change the initial conditions and parameters of the model to values different than those used in the “true” simulation, i.e. we define an erroneous “initial guess” of the model parameters and state. We first execute this model with no data assimilation to quantify the impact of the model errors. This is known as the baseline case. We then execute the model again, using the EnKF to assimilate the synthetic observations throughout the simulation. We compare the model output to that obtained from the “true” simulation, allowing us to evaluate the performance of the data assimilation method. This is a standard method for evaluating inverse problems [2].

We design our experiments to take into account that more frequent and higher accuracy observations accompanied a significant shift in the technology available for glacier observation around 1980 with the start of the polar-observing satellite era. While modern satellite observations offer precise and timely data on glacier flow, earlier paleo observations contribute essential context for understanding long-term trends and natural variability. Together, these observations offer complementary insights giving a fuller picture of glacier dynamics over time. We thus assume data is initially sparsely available and then available to be assimilated more frequently later in the simulations.

2.3.3 EnKF Parameters

We explore two hyperparameters of the EnKF, the ensemble size and the times of assimilation. For the ensemble size, we explore the impact of sizes ranging from 2 to 75 ensemble members. For the assimilation times, we explore the optimal number and frequency of observations in both the pre-satellite and satellite eras. For the post-satellite era, we simulated observational frequencies of 3 months, 6 months, yearly, 2 years, 4 years, 8 years and 16 years. For the pre-satellite era, we simulated observational frequencies of 9.5 years, 19 years, 38 years, 76 years and 190 years.

In all of our analyses, we use the mean square difference, $d = (x_t - x_t^a)^2$, to measure the distance of the analyzed state from the true state at time t , x_t . We calculate the square difference at each time step and take the square root of the mean over all the time steps to assess the best ensemble size and observation scheme.

2.4 Model Experiments

2.4.1 Glacier Model Forecasting

We conduct the twin experiments varying the ensemble size and the times of assimilation as described in Sect. 2.3.3. We implement the EnKF until observations “terminate” in 2022 and then use the analyzed state as an initial condition for the model to simulate H and L into the future, up to the year 2300. This “model forecast experiment” is designed to explore the impact of historical data on projections of glaciers in the future, which can inform current practices in glacier modeling.

3 Results

3.1 Sensitivity Analysis

Through our study of the effect of varying the model parameters by $\pm 10\%$, we find that the glacier height decreases gradually and then is essentially constant until 1950, at which point it decreases more rapidly until 2300 for all parameter values simulated (Fig. 3a–3c). Despite this general pattern, we see that the implementation of 10% variation can cause both a great deal of variation in model behavior and also very little depending on which set of parameters is perturbed.

As we can see in Fig. 3a, varying the surface mass balance parameters causes little variation in the modeled glacier height, and all simulations produce a generally similar shape. This is likely because thick marine-terminating glaciers, like the one simulated in this model, can take many millennia to respond to a change in surface mass balance. On the other hand, varying either the initial conditions (i.e. the group of parameters including bed slope) or the sill variables causes a great deal of variation in the modeled glacier height (Fig. 3b and 3c), with a larger spread observed for the variable initial conditions. These variations have a more direct influence on the behavior of the model at the grounding line, which responds on time scales of decades to centuries, i.e. within

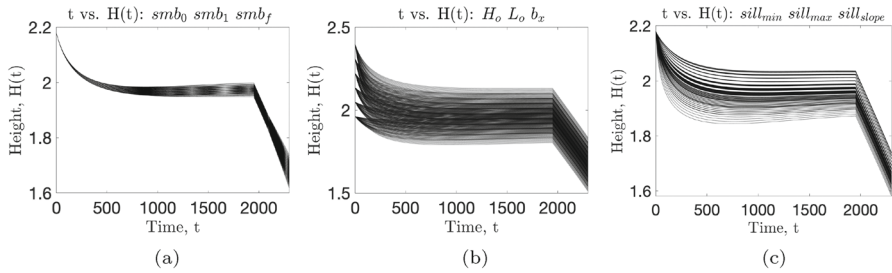
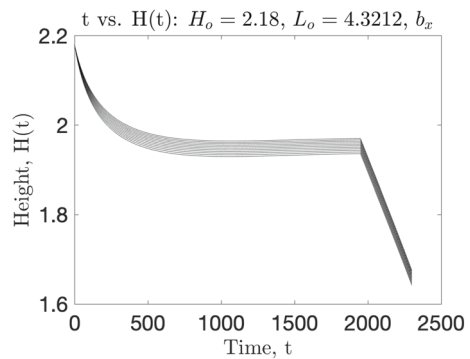


Fig. 3 Glacier height $H(t)$ over time t with $\pm 10\%$ variability in nominal parameters defining **a** surface mass balance, **b** initial conditions, and **c** sill geometry. The sudden change of behavior at 1950 is due to more frequent and higher accuracy observations that came with the start of the satellite era

Fig. 4 Single initial condition with slope varied $\pm 10\%$



the time scale of the simulations considered here. However, it should be noted that much of the variation in Fig. 3b is actually due to changing H_o and L_o , as we find there is little variation in the glacier height if we hold these fixed and only vary the slope, b_x (Fig. 4).

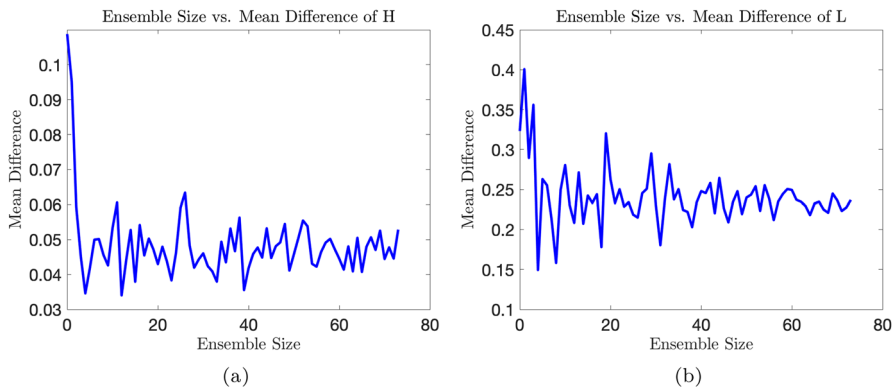
When deciding which model parameters are best suited for applications of data assimilation, our sensitivity analysis reveals that variable sill parameters will produce greater variation in model output. In the real world, uncertainty in the bed topography (and thus sill parameters) is also quite large, as measurements of the ice sheet base require the use of sparse ice-penetrating radar or borehole data. Parts of the bed topography falling in between existing radar measurement lines or boreholes are inferred using interpolation which can introduce substantial errors into these parameters. This motivates our choice of a sill parameter, and the sill slope in particular, as one model parameter to vary for our experiments (Table 1). We also vary smb_o , H_o , and L_o .

3.2 EnKF Parameters

The mean square differences computed when the twin experiments were conducted using various ensemble sizes are shown in Fig. 5a and 5b. We find that ensemble sizes of 7–10 were ideal, as the mean square difference hovers around a relatively constant value when the ensemble sizes is increased beyond this. The relatively small

Table 1 Model parameters and initial conditions used for the baseline simulation

	Parameters and Initial Conditions		
	True	Inaccurate	% variation
smb_o	0.3	0.35	$+\approx 16.67$
smb_l	0.15	0.15	n/a
smb_f	0.0	0.0	n/a
H_o	2.18	2.3	$+\approx 5.5$
L_o	4.44	4.6	$+\approx 3.3$
b_x	-0.001	-0.001	n/a
$sill_{min}$	415	415	n/a
$sill_{max}$	425	425	n/a
$sill_{slope}$	0.01	0.008	-20

**Fig. 5** Mean Square Difference in glacier height **H** **a** and length **L** **b** for varying ensemble sizes. The optimal ensemble size that best balances computational cost and model accuracy lies around 7–10

ensemble size is sensible given that this model has few degrees of freedom (two prognostic variables and few model parameters).

When exploring the optimal observation scheme, i.e. the time frames and frequency of assimilation that produce a sufficiently small mean difference over the course of the model run with a small number of observations, we find that before 1950 the best observation frequency is approximately every 19 years for a total of 100 observations (Fig. 6a and 6b). For the time frame of 1950–2300 we found that yearly observations for a total of 350 observations is the best frequency (Fig. 7).

3.3 Twin Experiments

Given our results from Sect. 3.2, we conduct the twin experiments using an ensemble size of 7–10, and assimilation frequency of every 19 years prior to 1950 and yearly observations after this.

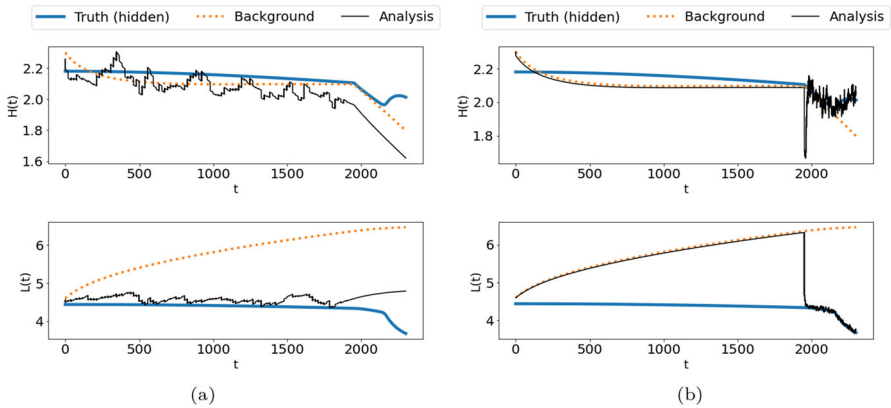


Fig. 6 Background, truth, and analysis for pre-1950 data assimilated every 19 years **a** and assimilated yearly **b**

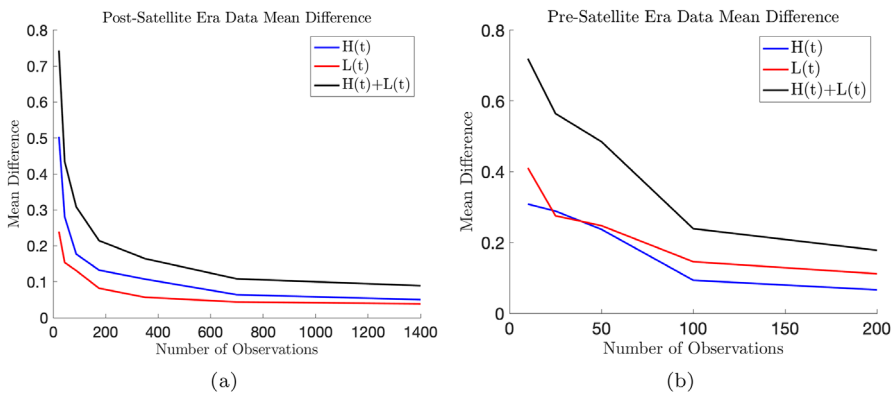


Fig. 7 Mean squared difference between analysis and truth for pre-satellite era data **a** and post-satellite era **b** data assimilation with different frequencies

The results of the “model forecasting” experiments are shown in Fig. 8. We see that after the last observation in 2020, the $H(t)$ analysis diverges from the truth but predicts a similar value at the end of the simulation. Similarly, we see that $L(t)$ analysis diverges slightly from the truth but also ends with a similar value. The analysis of both $H(t)$ and $L(t)$ is significantly closer to the truth than the baseline cases.

4 Discussion

4.1 Data Assimilation Impact on Glacier Studies

Our results show that implementing an EnKF improves the output of the two-stage ice sheet model. It improves model forecasts of glacier length and thickness, moving simulations initialized with erroneous or uncertain parameters closer to the “truth”

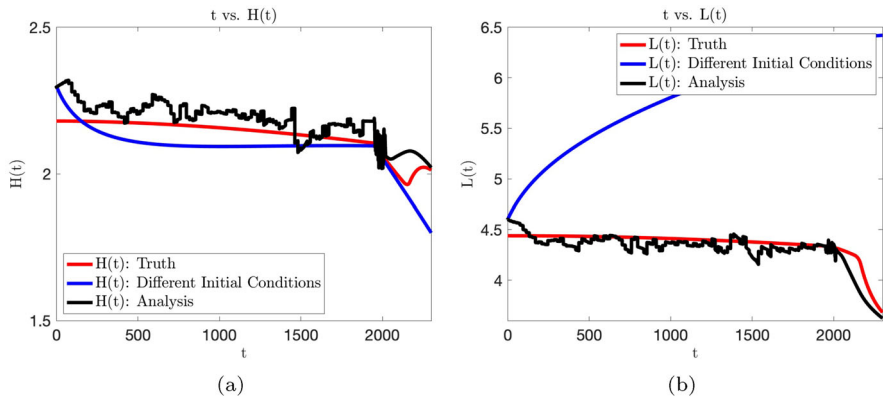


Fig. 8 Model run of **H a** and **L b** using best observation scheme and ensemble size

once data is assimilated. While this model is simplified, this suggests that data assimilation has strong potential to improve more complex glacier models. These methods can also be used to inform glacier studies, e.g., they can help determine the quantity of measurements needed to most accurately model ice sheets. Specifically, our results show that the increased costs of including more frequent observations does not always increase accuracy of the model. The improvements to glacier models shown in this study are important as they can help glacier modelers determine the frequency of measurements that researchers should gather in order to improve model output. For example, in Sect. 3.2 we found that there should be observations every 19 years before 1950 and yearly after 1950 in order to obtain model forecasts with minimal error, with constraints on observational frequency. Using a greater number of assimilation cycles is computationally more expensive, making it worthwhile to find an optimal assimilation scheme. More significant is the fact that it can be very expensive and time-intensive for researchers to collect glacier field data. We are not suggesting that researchers stop collecting data more frequently than once a year, as this data can still provide insight for different areas of research relating to glaciers and the systems they affect. However, assimilation frequency should be further explored using more complicated models and/or using real observation data.

A more complicated model will likely require larger ensemble sizes due to complexities introduced by an increase in the degrees of freedom, something that can be computationally prohibitive to implement. Therefore, we conjecture that it would be worthwhile to explore a more computationally efficient data assimilation method such as the singular evolutive interpolated Kalman filter in more complex models. See e.g., [14, 18] for more details.

The results of the sensitivity analysis show that the model itself is quite sensitive to changes in various parameters, which is significant as real world measurements of such parameters often have a lot of uncertainty and/or error. Knowledge of the range of accuracy in measurements that is necessary for model accuracy may help guide instrumentation development. This also underscores the importance of accurate measurements in simplified (and likely complex) glacier models, and the important

role that data assimilation may play in adjusting the model simulations to reduce the uncertainties created by input errors.

4.2 Study Limitations

It is worth noting that these calculations were performed using one specific set of parameters, so the findings are likely not universal. When the EnKF or similar data assimilation methods are implemented into more complicated glacier models, it will be worthwhile to similarly investigate optimal observation schemes. In this case, precision in determining the optimal observation scheme will be crucial, as small differences in the number and frequency of observations may have significant impacts due to the non-linearities in the system.

5 Conclusions

In this work, we conducted sensitivity analyses, and investigated and compared the impacts of data assimilation, specifically the ensemble Kalman filter (EnKF), on a two-stage ice sheet model. We explored the model output variables, glacier height and length, in the context of glacier melt due to climate change. To the best of our knowledge, this is one of the only studies in which EnKF techniques are incorporated with ice sheet modeling. Our results showed that incorporating observational data with the EnKF technique improves the accuracy of glacier height and length forecasts, and we determined that an ensemble size of 10 best minimizes computational cost while also minimizing error. The assimilation results also suggest the time frame in which observational data is most impactful to the output of the model, and that an ice sheet model may perform reasonably well even when only receiving modern observational data once yearly, without any historical data.

From our ensemble size results (Sect. 3.2), we hypothesise that there is a relationship between the number of variables as well as the parameter sensitivity in the two-stage glacier model and the ideal ensemble size required. When this method is implemented with more complex glacier models, the ideal ensemble size will likely be larger. We further postulate that a model supported by an observation scheme with a sufficient number of recent observations does not require much pre-satellite era observation data to make good predictions. “Sufficient” may be as low as one observation per year. However, if the model trajectory is quite far from the mean observation trajectory, long-term predictions may not be reliable using an EnKF alone. This work provides foundational exploration and motivates implementation of a non-linear data assimilation scheme such as the EnKF with a more complicated glacier model and real-world observations. This has promising implications for better understanding sea level rise.

6 A Data Availability

The full data file with the data we generated, as well as related code and further graphs of data assimilation schemes for this two-stage ice sheet model can be found at <https://github.com/hakuupi/StormSurge>.

The code used to run model and perform assimilation is in the Jupyter_Notebook_Code folder; we ran Python in Jupyter Notebook to generate data and run model.

Our data is stored in Experiment_Data and is saved in csv files. Note that the parameters and initial conditions used are in the name of the file.

The MATLAB code used to generate graphics used in our paper is found in Matlab_Plot_Code.

B Figures

See Appendix Figs. 9, 10 and 11.

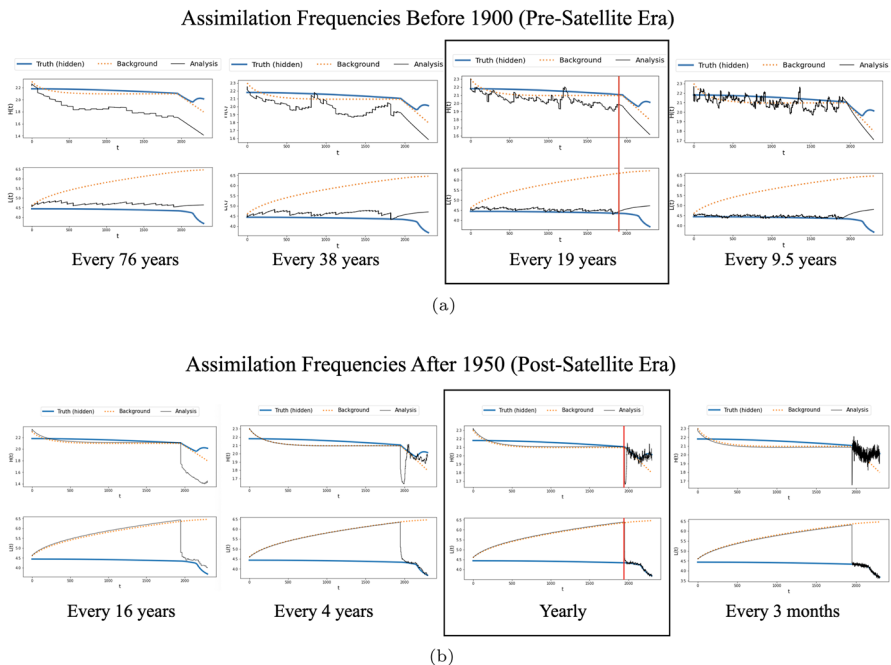


Fig. 9 Assimilation dates and frequencies. **a** Assimilation Frequencies Before 1900 in the pre-satellite era. The assimilation schemes with under 5 percent in the assimilation time period are every 19 years or more frequent. The scheme that assimilates data every 19 years is boxed. The red vertical line marks the year 1900, the end of data assimilation in this experiment. **b** Assimilation Frequencies After 1950 in the post-satellite era. The assimilation schemes with under 5 percent in the assimilation time period are yearly or more frequent. The yearly scheme is boxed. The red vertical line marks the year 1950, the start of data assimilation in this experiment.

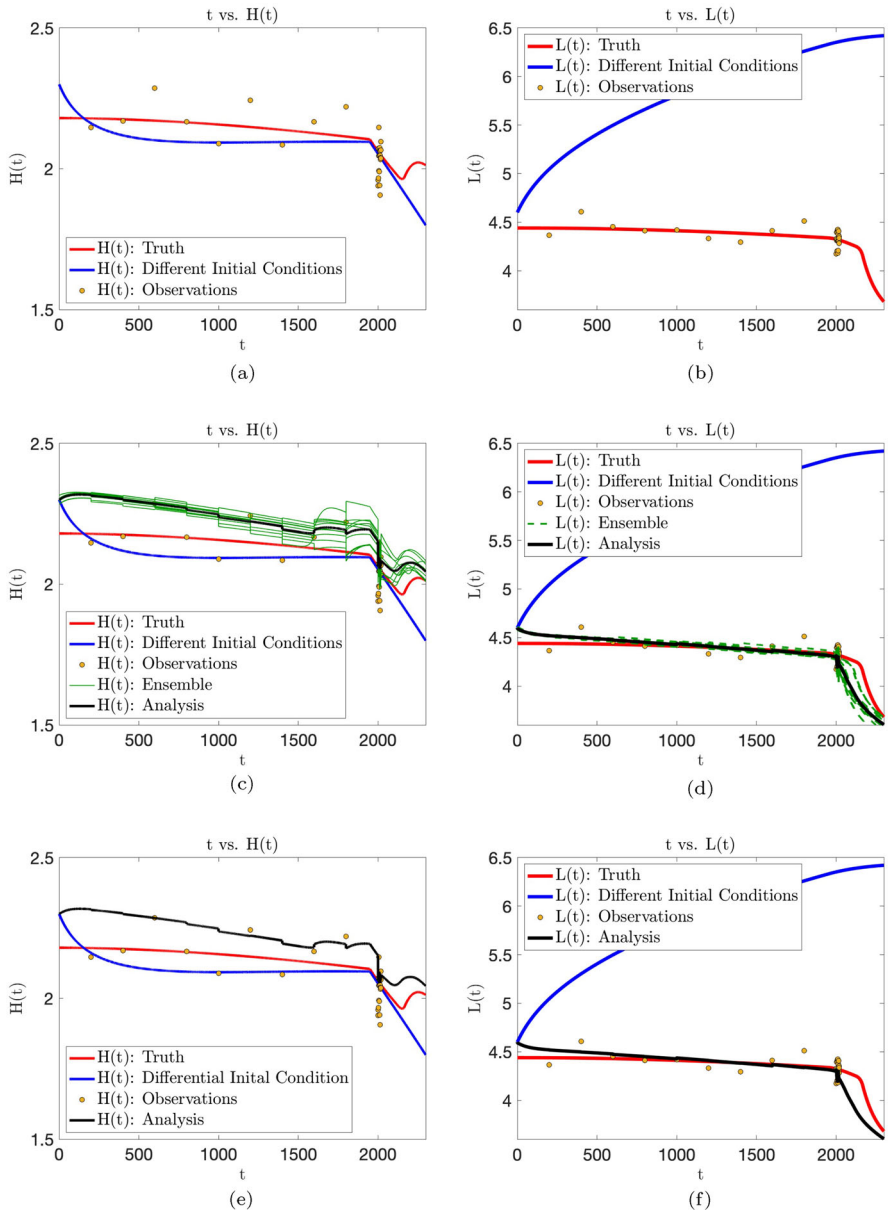


Fig. 10 Model runs with worse observation and ensemble Scheme with mean square difference for H of 0.00982 and mean square difference for L of 0.00941. Note: mean square difference is rounded to 3 significant digits. **a** Observations and Truth Simulation for H . **b** Observations and Truth Simulation for L . **c** Analysis, Ensemble, Observations and Truth Simulation for H . **d** Analysis, Ensemble, Observations and Truth Simulation for L . **e** Analysis, Observations and Truth Simulation for H . **f** Analysis, Observations and Truth Simulation for L

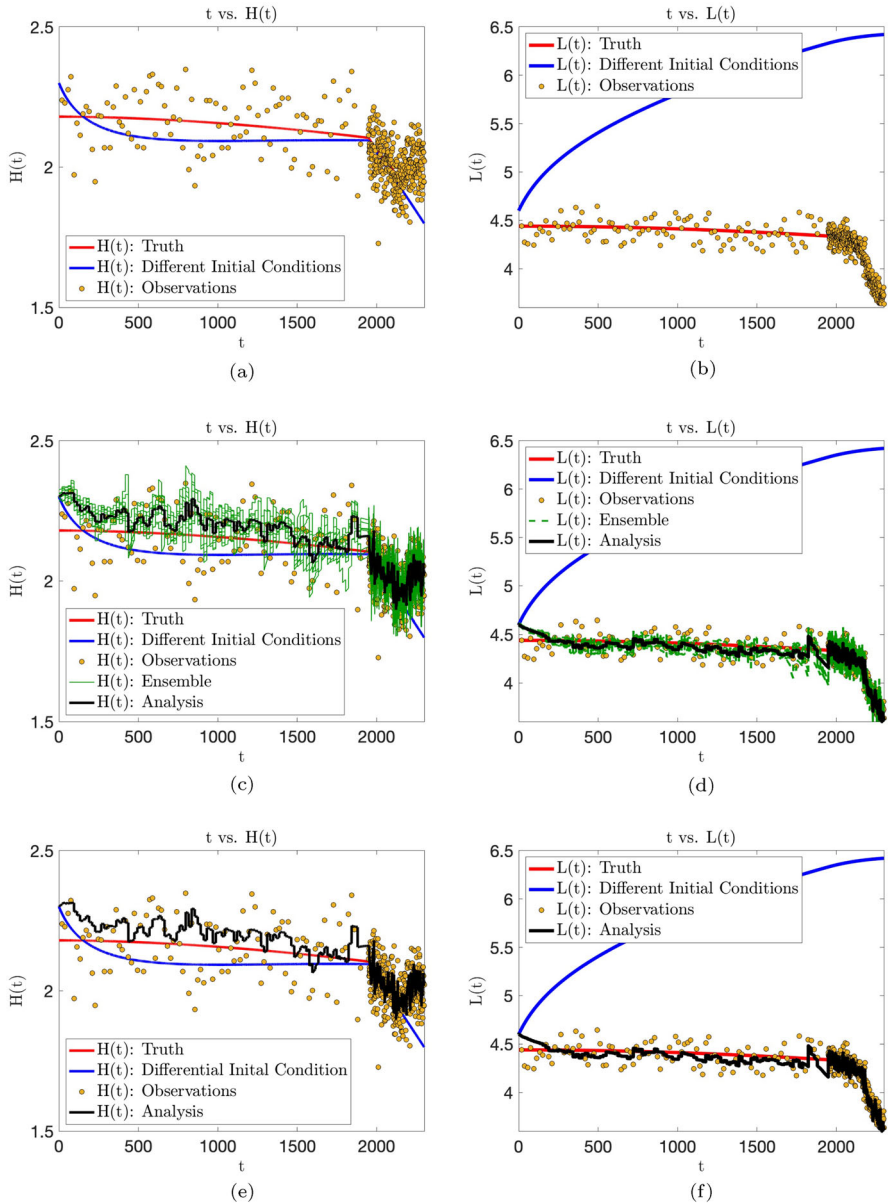


Fig. 11 Model Runs with best observation and ensemble scheme, with mean square difference for H of 0.00353 and mean square difference for L of 0.00328. Note: mean square difference is rounded to 3 significant digits. **a** Observations and Truth Simulation for H . **b** Observations and Truth Simulation for L . **c** Analysis, Ensemble, Observations and Truth Simulation for H . **d** Analysis, Ensemble, Observations and Truth Simulation for L . **e** Analysis, Observations and Truth Simulation for H . **f** Analysis, Observations and Truth Simulation for L

Acknowledgements This work is supported in part by the US National Science Foundation awards DMS-2051019 and DMS-1751636. We would like to thank Dr. Lars Ruthotto, Dr. Bree Ettinger, and all of the faculty involved in the Emory University Computational Mathematics for Data Science REU for their invaluable support.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no Conflict of interest.

References

1. Alvarez-Cuesta, M., Toimil, A., Losada, I.: Which data assimilation method to use and when: unlocking the potential of observations in shoreline modelling. *Environ. Res. Lett.* **19**(4), 044023 (2024)
2. Asch, M., Bocquet, M., Nodet, M.: *Fundamentals of Algorithms Data Assimilation*. SIAM, Philadelphia (2009)
3. Barker, D., Huang, X.-Y., Liu, Z., Auligné, T., Zhang, X., Rugg, S., Ajjaji, R., Bourgeois, A., Bray, J., Chen, Y., Demirtas, M., Guo, Y.-R., Henderson, T., Huang, W., Lin, H.-C., Michalakes, J., Rizvi, S., Zhang, X.: The weather research and forecasting model's community variational/ensemble data assimilation system: WRFDA. *Bull. Am. Meteorol. Soc.* **93**(6), 831–843 (2012)
4. Bonan, B., Nodet, M., Ritz, C., Peyaud, V.: An ETKF approach for initial state and parameter estimation in ice sheet modelling. *Nonlinear Process. Geophys.* **21**(2), 569–582 (2014)
5. Butler, T., Altaf, M.U., Dawson, C., Hoteit, I., Luo, X., Mayo, T.: Data assimilation within the advanced circulation (adcirc) modeling framework for hurricane storm surge forecasting. *Month. Weather Rev.* **140**(7), 2215–2231 (2012)
6. Camelo, J., Mayo, T.L., Gutmann, E.D.: Projected climate change impacts on hurricane storm surge inundation in the coastal united states. *Front. Built Environ.* (2020)
7. Catania, G., Stearns, L., Sutherland, D., Fried, M., Bartholomaeus, T., Morlighem, M., Shroyer, E., Nash, J.: Geometric controls on tidewater glacier retreat in central western greenland. *J. Geophys. Res. Earth Surf.* **123**(8), 2024–2038 (2018)
8. ECMWF: Data assimilation (2022)
9. Evensen, G., et al.: *Data Assimilation: the Ensemble Kalman Filter*, vol. 2. Springer, New York (2009)
10. Fletcher, S.J.: *Data Assimilation for the Geosciences: From Theory to Application*. Elsevier, Amsterdam (2017)
11. Gillet-Chaufet, F.: Assimilation of surface observations in a transient marine ice sheet model using an ensemble Kalman filter. *Cryosphere* **14**(3), 811–832 (2020)
12. Goldberg, D., Heimbach, P.: Parameter and state estimation with a time-dependent adjoint marine ice sheet model. *Cryosphere* **7**(6), 1659–1678 (2013)
13. Higdon, D., Gattiker, J., Lawrence, E., Jackson, C., Tobis, M., Pratola, M., Habib, S., Heitmann, K., Price, S.: Computer model calibration using the ensemble Kalman filter. *Technometrics* **55**(4), 488–500 (2013)
14. Hoteit, I., Pham, D.-T., Blum, J.: A simplified reduced order kalman filtering and application to altimetric data assimilation in tropical pacific. *J. Marine Syst.* **36**, 101–127 (2002)
15. Katzfuss, M., Stroud, J.R., Wikle, C.K.: Understanding the ensemble kalman filter. *Am. Stat.* **70**(4), 350–357 (2016)
16. Loucks, D.P., Beek, E.v., Stedinger, J.R., Dijkman, J.P.M., Villars, M.T.: *Model Sensitivity and Uncertainty Analysis*. Water Resources Systems Planning and Management, pp. 255–286. UNESCO, Paris (2005)
17. Mayo, T.L., Lin, N.: Climate change impacts to the coastal flood hazard in the northeastern united states. *Weather Clim. Extrem.* 100453 (2022)
18. Mayo, T., Butler, T., Dawson, C., Hoteit, I.: Data assimilation within the advanced circulation (adcirc) modeling framework for the estimation of manning's friction coefficient. *Ocean Model.* **76**, 43–58 (2014)

19. Morlighem, M., Rignot, E., Binder, T., Blankenship, D., Drews, R., Eagles, G., Eisen, O., Ferraccioli, F., Forsberg, R., Fretwell, P., et al.: Deep glacial troughs and stabilizing ridges unveiled beneath the margins of the antarctic ice sheet. *Nat. Geosci.* **13**(2), 132–137 (2020)
20. Oppenheimer, M., Glavovic, B.C., Hinkel, J., Wal, R., Magnan, A.K., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., DeConto, R.M., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., 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, pp. 321–445 (2019). <https://doi.org/10.1017/9781009157964.006>
21. Robel, A.: The long future of antarctic melting. *Nature* **526**(7573), 327–328 (2015)
22. Robel, A.A., Roe, G.H., Haseloff, M.: Response of marine-terminating glaciers to forcing: time scales, sensitivities, instabilities, and stochastic dynamics. *J. Geophys. Res. Earth Surf.* **123**(9), 2205–2227 (2018)
23. Robel, A., Seroussi, H., Roe, G.H.: Marine ice sheet instability amplifies and skews uncertainty in projections of future sea-level rise. *PNAS* **116**(30), 14887–14892 (2019)
24. Weertman, J.: Stability of the junction of an ice sheet and an ice shelf. *J. Glaciol.* **13**(67), 3–11 (1974)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.