Bias correction and statistical modeling of variable oceanic forcing of Greenland outlet glaciers

Vincent Verjans[1], Alexander Robel[1], Andrew F. Thompson[2], and Helene Seroussi[3]

[1]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[2]Environmental Science and Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[3]Thayer School of Engineering, Dartmouth College, Hanover, NH, USA

Key Points:

- We develop a statistical method to generate ocean forcing boundary conditions for Greenland ice sheet model simulations.
- The method bias-corrects and extrapolates global climate model output using reanalysis products and high-resolution model results.
- Stochastic time series models reproduce the spatiotemporal variability of ocean conditions at negligible computational expense.

Corresponding author: Vincent Verjans, vverjans3@gatech.edu

Abstract

Variability in oceanic conditions directly impacts ice loss from marine outlet glaciers in Greenland, influencing the ice sheet mass balance. Oceanic conditions are available from Atmosphere-Ocean Global Climate Model (AOGCM) output, but these models require extensive computational resources and lack the fine resolution needed to simulate ocean dynamics on the Greenland continental shelf and close to glacier marine termini. Here, we develop a statistical approach to generate ocean forcing for ice sheet model simulations, which incorporates natural spatiotemporal variability and anthropogenic changes. Starting from raw AOGCM ocean heat content, we apply: (1) a bias-correction using ocean reanalysis, (2) an extrapolation accounting for on-shelf ocean dynamics, and (3) stochastic time series models to generate realizations of natural variability. The bias-correction reduces model errors by ∼25 % when compared to independent in-situ measurements. The bias-corrected time series are subsequently extrapolated to fjord mouth locations using relations constrained from available high-resolution regional ocean model results. The stochastic time series models reproduce the spatial correlation, characteristic timescales, and the amplitude of natural variability of bias-corrected AOGCMs, but at negligible computational expense. We demonstrate the efficiency of this method by generating >6000 time series of ocean forcing for >200 Greenland marine-terminating glacier locations until 2100. As our method is computationally efficient and adaptable to any ocean model output and reanalysis product, it provides flexibility in exploring sensitivity to ocean conditions in Greenland ice sheet model simulations. We provide the output and workflow in an open-source repository, and discuss advantages and future developments for our method.

Plain Language Summary

Model simulations of the Greenland ice sheet require knowledge of ocean conditions. The evolution of ocean conditions has a strong impact on ice sheet model predictions, as there are more than 200 glaciers in Greenland flowing directly into the ocean. However, modeling oceanic forcing is difficult. The state-of-the-art approach is to use output from Atmosphere-Ocean Global Climate Models (AOGCMs). But these models cannot accurately capture the ocean dynamics on the Greenland shelf, and they can show strong biases compared to observations. Furthermore, AOGCMs are computationally expensive, meaning that it is impossible to thoroughly characterize the uncertainty associated with the chaotic nature of climate. Here, we propose a procedure to bias-correct and extrapolate oceanic output from AOGCMs. Our method exploits observational datasets, as well as available high-resolution ocean model results. Using statistical models, we reproduce patterns of spatiotemporal ocean variability at low computational expense, and represent internal climate variability and global warming trends. The goal is to provide a scalable procedure to generate ocean forcing for long-term Greenland ice sheet model predictions.

1 Introduction

Since 1992, Greenland ice sheet (GrIS) mass loss has contributed ∼0.4 mm yr^{-1} to global mean sea-level rise (IMBIE, 2020). Mass losses are approximately equally partitioned between increased surface melt runoff and increased ice discharge into the ocean, although variability in the contribution of these two processes is strongly linked to temporal variability in climatic forcing (Mouginot et al., 2019). In particular, increased ice discharge has been linked to warming oceanic conditions (Holland et al., 2008; Walsh et al., 2012; Straneo & Heimbach, 2013; Porter et al., 2018; Wood et al., 2021). Increased ice loss at outlet glacier termini causes glacier thinning and speed-up, thus inducing longer-term dynamic responses in the ice sheet interior (Nick et al., 2009; Felikson et al., 2017). Changes in ocean temperatures, and their link to increased outlet glacier mass loss rates,

are therefore expected to exert a major control on future GrIS mass balance (Wood et al., 2021).

Melting of marine-terminating glaciers is driven by relatively warm deep waters of Atlantic origin (Straneo & Heimbach, 2013; Sutherland et al., 2013). However, heat delivery to outlet glacier termini depends on water circulation across the continental shelf and within fjords. Waters of Atlantic origin are transported onto the shelf mostly via cross-shelf troughs (Rignot et al., 2012; Sutherland et al., 2013). Oceanic heat found at the shelf break can be restricted from reaching glacier termini due to several factors. First, heat can be eroded due to vertical mixing. Second, the presence of sills can block the access of Atlantic waters into fjords (Straneo et al., 2012; Straneo & Cenedese, 2015; Jackson et al., 2018). Finally, close to the glacier termini, several convective processes can influence heat availability for glacier melt, such as subglacial discharge (Slater et al., 2018), wind-driven flow (Jackson et al., 2014; Sutherland et al., 2014; Jackson et al., 2018), and sea-ice formation (Cottier et al., 2010). All of these local processes over the continental shelf contribute to setting the characteristic variability of ocean thermal forcing.

While ice sheet model simulations of Greenland outlet glaciers are sensitive to oceanic forcing, no Greenland-scale ocean model completely captures the range of processes and time scales governing ocean heat transfer to glacier termini (Slater et al., 2020). Even the highest-resolution regional ocean model simulations do not capture all the kilometer to sub-kilometer scale processes at play, and they only extend over periods shorter than 20 yr (e.g., Rignot et al., 2012; Gillard et al., 2016). In this context, Greenland ice sheet model predictions use ocean forcing input provided by Atmosphere-Ocean General Circulation Models (AOGCMs), because they cover periods until 2100 and beyond. However, due to their coarse resolution in the ocean (typically $1^\circ \times 1^\circ$), they cannot simulate fjord processes, and their representation of on-shelf ocean dynamics is incomplete (Slater et al., 2020). Because of the current inability to fully resolve fjord dynamics within large-scale ocean models, Xu et al. (2012) and Rignot et al. (2016) developed a parameterization of glacier melt based on an empirical relation found with water temperatures at the fjord mouth and subglacial discharge. This parameterization has not been calibrated to measurements from outlet glaciers, but to idealized fjord-scale model simulations. Another parameterization, adapted to coarse-resolution AOGCM output, has been developed for use in ice sheet models by linking regionally-averaged ocean temperatures to individual glacier terminus positions (Cowton et al., 2018; Slater et al., 2019). The latest Ice Sheet Model Intercomparison for CMIP6 (ISMIP6) for Greenland has proposed these two types of parameterizations: prescribing either terminus position or glacier frontal melt as a function of far-field ocean temperature averaged over large regional oceanic sectors (Goelzer et al., 2020; Slater et al., 2020). While computationally convenient, the terminus position parameterization neglects feedback effects from ice flow dynamics and bed topography on outlet glacier dynamics, and its empirical parameter is highly uncertain (Slater et al., 2019, 2020). Furthermore, the parameterization relies on far-field ocean temperature, thus neglecting variability associated with shelf processes in heat transport towards the fjords. The direct ice melt rate parameterization, on the other hand, allows ice sheet models to resolve interactions between melt and calving rates, but remains to be validated against large-scale Greenland outlet glaciers observations. It also relies on the assumption that AOGCMs can provide accurate ocean temperature fields at the entry of fjord mouths, which corresponds to areas where fine-scale dynamical and topographical details cannot be resolved in coarse resolution models.

Finally, current ice sheet model predictions neglect internal variability in ocean conditions, as they generally use a single deterministic AOGCM output to represent future oceanic conditions. ISMIP6 accounted for inter-AOGCM uncertainty and greenhouse gas emission-scenario dependence by using six different AOGCMs, one of which included both a high- and low-emission scenario (Goelzer et al., 2020; Slater et al., 2020). However, different runs from the same AOGCM starting with only round-off level errors in initial con-

ditions can exhibit large differences in patterns of climate variability over a range of timescales (Kay et al., 2015; Maher et al., 2019). This internal variability is caused by the chaotic nature of the climate system (Hasselmann, 1976). Due to the computational expense of AOGCM simulations, the number of different simulations from a given AOGCM is limited, making it challenging to directly force ice sheet models with a model ensemble of climatic forcings representative of internal climate variability. To date, the most comprehensive evaluation of internal ocean variability impact on GrIS simulations has been performed by Tsai et al. (2017), as they used the coarse-resolution ocean output of 50 realizations from a same AOGCM. An alternative approach to quantify the impact of internal variability is to calibrate statistical models to a small set of AOGCM runs (e.g., Castruccio & Stein, 2013; Hu & Castruccio, 2021), which can then be used to generate stochastic climatic forcing within ice sheet models (Verjans et al., 2022).

Using AOGCMs to provide long-term ocean forcing for ice sheet models thus faces the limitations of horizontal resolution and characterization of internal variability. Furthermore, bias-correction techniques are needed due to AOGCM disagreement with in-situ observational data (Slater et al., 2019), but currently-used techniques are simplistic compared to methods applied in other climate model applications (e.g., Cui et al., 2012). There is a stark contrast between the level of refinement of atmospheric forcing compared to ocean forcing used in Greenland ice sheet simulations. For atmospheric forcing, there exists a large number of high-resolution models to downscale AOGCM output, which are specifically calibrated to ice sheet processes (Fettweis et al., 2020). Dynamical downscaling allows to resolve small scale processes over Greenland, while being forced by AOGCM large scale fields at the domain boundaries. Some models even associate this dynamical downscaling with a statistical downscaling process (Noël et al., 2016). In contrast, no high-resolution ocean model or downscaling method have been used to generate ocean forcing for Greenland ice sheet model simulations in any prior studies, although such methods have been applied to ocean models for other applications (e.g., Camus et al., 2014; Oliver & Holbrook, 2014; Fagundes et al., 2020).

In this study, we outline a statistical method to compute ocean thermal forcing for outlet glaciers in Greenland ice sheet model simulations. We describe the methodological details, present results from application of the method, and discuss advantages and possible future developments. We provide the code and many different outputs in an open source repository (see *Open Research*).

2 Methods

The overarching objective of this study is to generate the most representative realizations of ocean thermal forcing (TF) at the fjord mouth of Greenland outlet glaciers. We describe a general method to achieve this, starting from the ocean temperature and salinity outputs of any AOGCM. The variable of interest is TF, which is defined as the ocean temperature above the freezing point. To calculate TF, we use the salinity- and depth-dependent empirical equation for the freezing point from Cowton et al. (2015):

$$TF(\mathbf{x}, t) = T_{oc}(\mathbf{x}, t) - (\lambda_1 S_{oc}(\mathbf{x}, t) + \lambda_2 + \lambda_3 z)\,, \tag{1}$$

where T_{oc} is the ocean temperature [°C], S_{oc} is the ocean salinity [psu], z is depth [m, positive upwards], and $\lambda_1, \lambda_2, \lambda_3$ are parameters set to -5.73×10^{-2} °C psu^{-1}, 8.32×10^{-2} °C, and -7.61×10^{-4} °C m^{-1}, respectively. The dependence on space and time is highlighted by $\mathbf{x}$ and t, respectively.

We average TF between the surface and 500 m depth: $TF_{0\text{-}500}$. At gridpoints where the bathymetry is shallower than 500 m, $TF_{0\text{-}500}$ only accounts for TF values ranging between 0 m and the seafloor depth. However, we discard gridpoints where the bathymetry is less than 100 m deep to generate our datasets of $TF_{0\text{-}500}$, as the coarse resolution AOGCMs

do not capture the fine details of bathymetry in these areas. As such, our computation of $TF_{0\text{-}500}$ at any gridpoint i, j with bathymetry $B_{i,j}$ is defined as:

$$TF_{0-500} = \frac{1}{|\max(B_{i,j}, -500)|} \int_{\max(B_{i,j},-500)}^{0} TF(z)dz \qquad \text{if } B_{i,j} \leq -100. \tag{2}$$

Our approach differs from the approach of Slater et al. (2019, 2020), as they averaged TF only between 200 and 500 m depth ($TF_{200\text{-}500}$). The 0-500 m depth range is chosen here to remain consistent with the derivation of the melt parameterization of Xu et al. (2012). Furthermore, on a large part of the Greenland continental shelf, the 0-200 m depth range is an important fraction of the water column (Morlighem et al., 2017, see also Fig.1a). As such, $TF_{0\text{-}200}$ is potentially an important contributor to variability in thermal forcing for Greenland outlet glaciers, although these waters are often colder and fresher than deeper Atlantic waters. However, the method outlined in this study can easily be reproduced choosing any depth range over which TF is averaged, which can serve different applications. In particular, it is possible to process TF separately over different depth ranges to study shallower Arctic waters and deeper Atlantic waters separately. In the following, we drop the subscript 0-500 to simplify the notation, but any TF symbol stands for $TF_{0\text{-}500}$ as defined in Eq. (2).

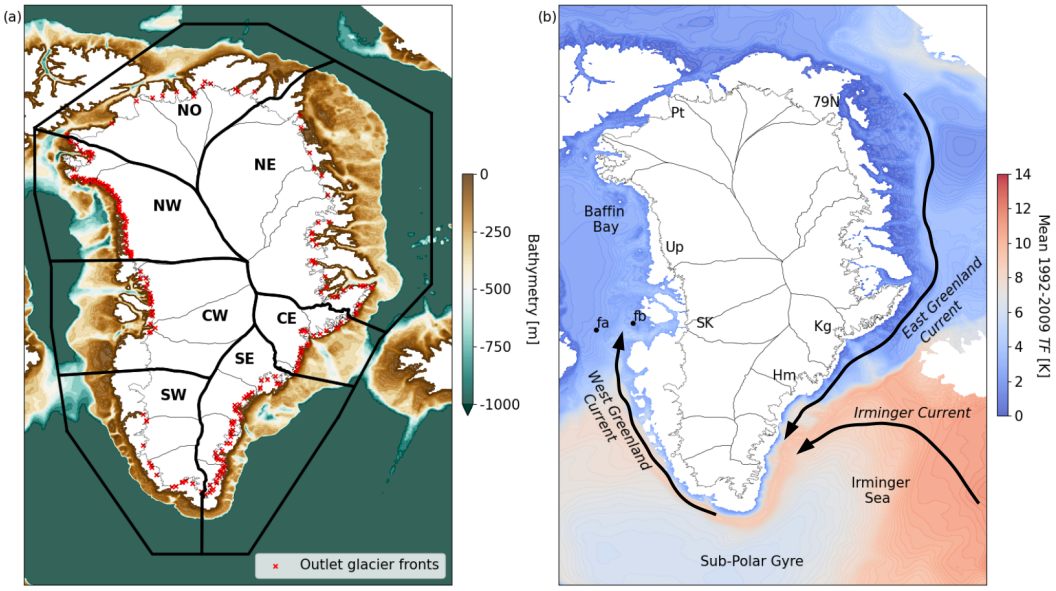

Figure 1. Maps of Greenland with (a) bathymetry and (b) mean (1992-2009) TF from ECCO (Nguyen et al., 2012). The ECCO bathymetry uses a merged product of the blend S2004 (Marks & Smith, 2006) and International Bathymetric Chart of the Arctic Ocean (Jakobsson et al., 2008). In (a), we show the delineation of the 7 oceanic sectors that constitute the ocean domain for this study and the 226 marine-terminating outlet glacier front locations. The sectors are taken from Slater et al. (2019), but the SE and SW sectors are extended 100 km southwards. The outlet glacier fronts are from Wood et al. (2021). In (b), we show major locations mentioned in this study. Glaciers are Kg: Kangerlussuaq, Hm: Helheim, SK: Sermeq Kujalleq, Up: Upernavik, Pt: Petermann, 79N: 79 North. The points fa and fb show locations of gridpoints used in Fig (5a) and (5b), respectively.

Our methodology to generate *TF* time series based on AOGCM *TF* output consists of three separate steps, which are outlined in Figure 2, and multiple data and model products, which are summarized in Table 1. The first step is to use an ocean reanalysis product, which is constrained by observations, to correct the AOGCM bias and to constrain its temporal variability in *TF* (first key step in Fig. 2). This statistical correction is performed via Quantile Delta Mapping, which is a method detailed in Section 2.1. The second step is to extrapolate the corrected AOGCM *TF* from the open ocean to inshore, i.e., at the mouths of fjords on the Greenland shelf (second key step in Fig. 2). This extrapolation process uses constraints derived from output of a high-resolution regional ocean model and accounts for different offshore-inshore relationships at different timescales of variability, as explained in Section 2.2. The final step is to calibrate statistical time series models to the variability of the corrected and extrapolated *TF* obtained after the first two steps (third key step in Fig. 2). This calibration procedure, detailed in Section 2.3, effectively reproduces the stochastic variability in ocean heat content, while preserving deterministic signals such as means, trends, and seasonality patterns. Ultimately, our method constructs statistical models capable of generating large numbers of realizations of *TF* for Greenland glaciers. This study uses specific data and model products, described in Table 1, but allows great flexibility in the particular choice of these products.

Type of product	Used in this study	Used for	Sections	Temporal coverage	Horizontal resolution	Reference
AOGCM	MIROC-ES2L	Raw *TF*	2.1, 2.2, 2.3	1850-2100	1°	Hajima et al. (2020)
AOGCM	IPSL-CM6A	Raw *TF*	2.1, 2.2, 2.3	1850-2100	1°	Boucher et al. (2020)
Reanalysis	EN4 objective analyses	Quantile Delta Mapping (QDM)	2.1	1900-2022	1°	Good et al. (2013)
High-resolution ocean model	ECCO-Arctic	Offshore to inshore extrapolation	2.2	1992-2009	4 km	Nguyen et al. (2012)
In-situ data	CTD	Evaluation of QDM	3.1	2000-2015	/	Good et al. (2013)

Table 1. Model and data products required in our method, and the examples of specific products used for this study (*TF*: thermal forcing).

2.1 Statistical correction of thermal forcing

The first step in our method is to bias-correct the AOGCM *TF* to the mean, seasonality, and interannual variability of a reference dataset. To illustrate our method, the reference dataset used is the Hadley Centre EN4.2.1 monthly objective analyses (Good et al., 2013), hereafter referred to as EN4. EN4 is a gridded product at 1° resolution covering 1900-2022. The EN4 method uses a local interpolation of Conductivity-Temperature-Depth (CTD) profile measurements combined to a background persistence from damped anomalies of the previous month; we refer to Good et al. (2013) for the methodological details. While it may be questionable to use a reanalysis product as the ground-truth dataset for a statistical correction, we note several advantages of the EN4 monthly objective analyses. First, EN4 is an interpolated product of oceanographic profile data. In contrast to other reanalysis products that use a dynamical model with data assimilation, EN4 is more strongly constrained by observations, ensuring better agreement with in-situ data. On the other hand, it implies that EN4 interpolates between observations without dynamical constraints, and is more prone to errors in case of observational uncertainties and if some periods and/or regions have sparse observational coverage. Second, the long temporal coverage provides more robust statistics than from most other reanalysis products. The long response timescales of ocean dynamics imply that at least several decades are needed to capture oceanic mean conditions and variability. Still, we note

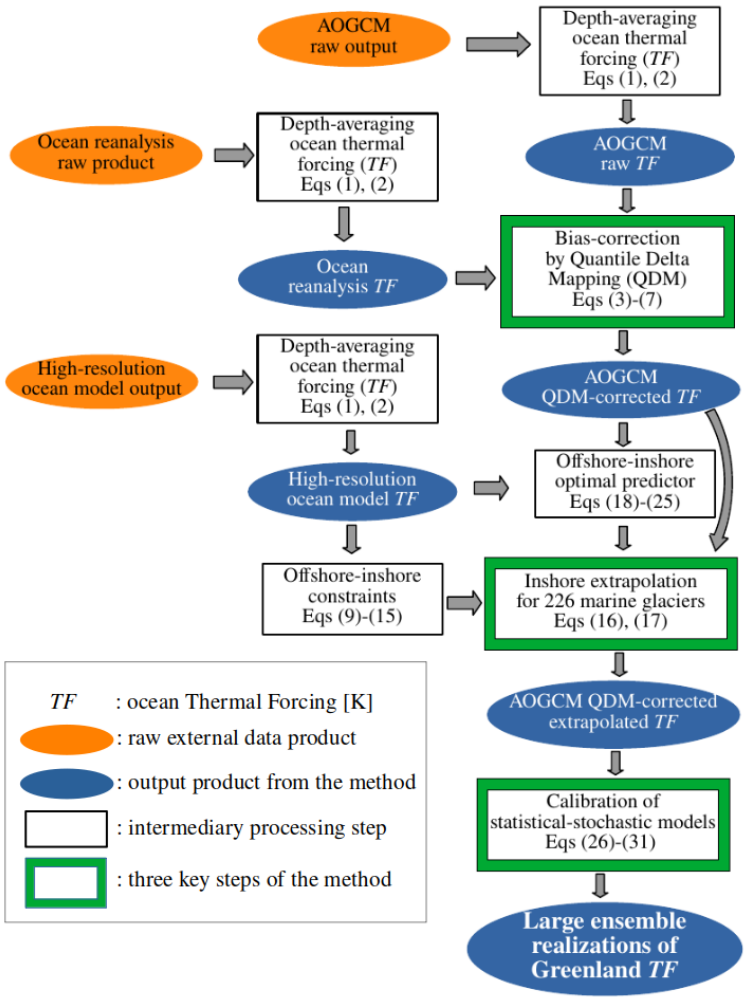

Figure 2. Flowchart of the methodology. The final output is a large ensemble of realizations of Greenland oceanic conditions at all the marine terminating glaciers.

that the bias-correction could equally well be performed with other gridded ocean reanalysis products in place of the EN4 objective analyses.

We use both the temperature and salinity products from EN4 to compute a monthly gridded *TF* field, following Eq. (1). We discard EN4 gridpoints if the *TF* time series has negative values. EN4 also provides an observational weight variable, which varies between 0 and 1 depending on how strongly EN4 is constrained by observations (see Good et al., 2013). We average the observational weights over large areas: the 7 oceanic sectors around Greenland used by Slater et al. (2019): North (NO), North-East (NE), Central-East (CE), South-East (SE), South-West (SW), Central-West (CW), and North-West (NW) (see Fig.1a). We extend the SE and SW sectors 100 km southwards because the original sector delineations only cover two to three gridpoints of EN4 in the meridional direction. In Figure 3, the sector-averaged observational weights clearly show a shift around 1950, with oceanic properties being significantly better constrained from this date. For this reason, we elect 1950 as the starting date for the calibration period, and we discard the 1900-1950 data. Nevertheless, EN4 observational weights are consistently lower in winter months, due to sparser in-situ data collection. We correct *TF* time series of individual AOGCM gridpoints over their period overlapping the 1950-2022 period. In the Coupled Model Intercomparison Project 6 (CMIP6) framework, historical model experiments cover the period 1850-2015, and model results post-2015 depend on the emission scenario assumed. Thus, our period for calibrating AOGCM *TF* is 1950-2015. Any given AOGCM gridpoint is corrected with the nearest neighbor EN4 *TF* time series. The statistical correction requires time series of data, and can therefore not be performed with in-situ data only, which are sparse in time and space.

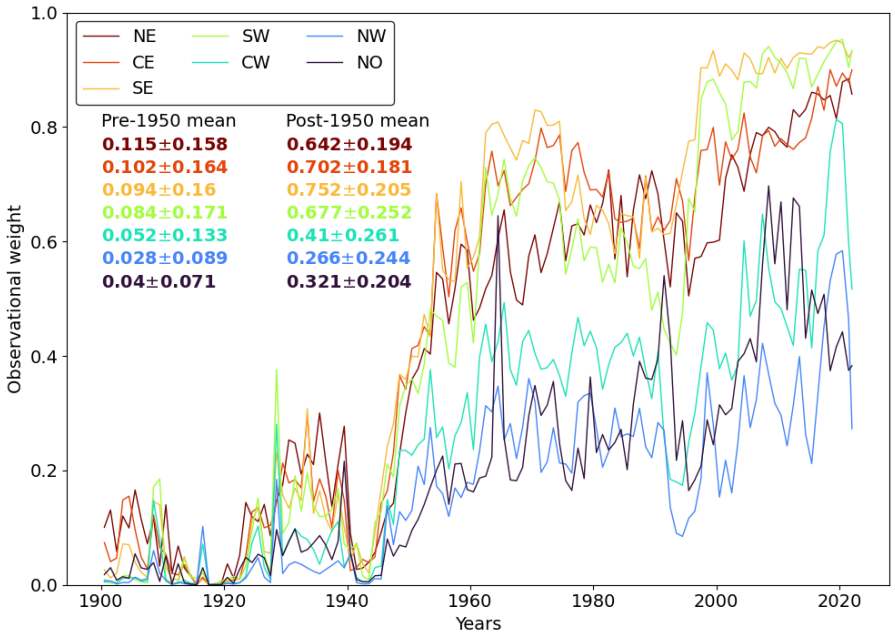

Figure 3. Observational weight variable of the EN4 objective analyses (Good et al., 2013) averaged yearly and by oceanic sector. Higher values of observational weight denote periods better constrained with in-situ measurements. Sector-averaged pre- and post-1950 mean values of the observational weight are displayed, with ± 1 standard deviation.

The motivation for correcting AOGCM outputs is that they may misrepresent the mean and/or variability in *TF*. Figure 4 shows biases between EN4 and two AOGCMs

that participated to ISMIP6. On the Greenland shelf, the biases in mean *TF* and *TF* standard deviation typically range between [-4;4] K and [-1.5;-0.5] K, respectively (Fig. 4). While the bias in the mean varies geographically and between the two AOGCMs, both of them generally underestimate *TF* variability when compared to EN4.

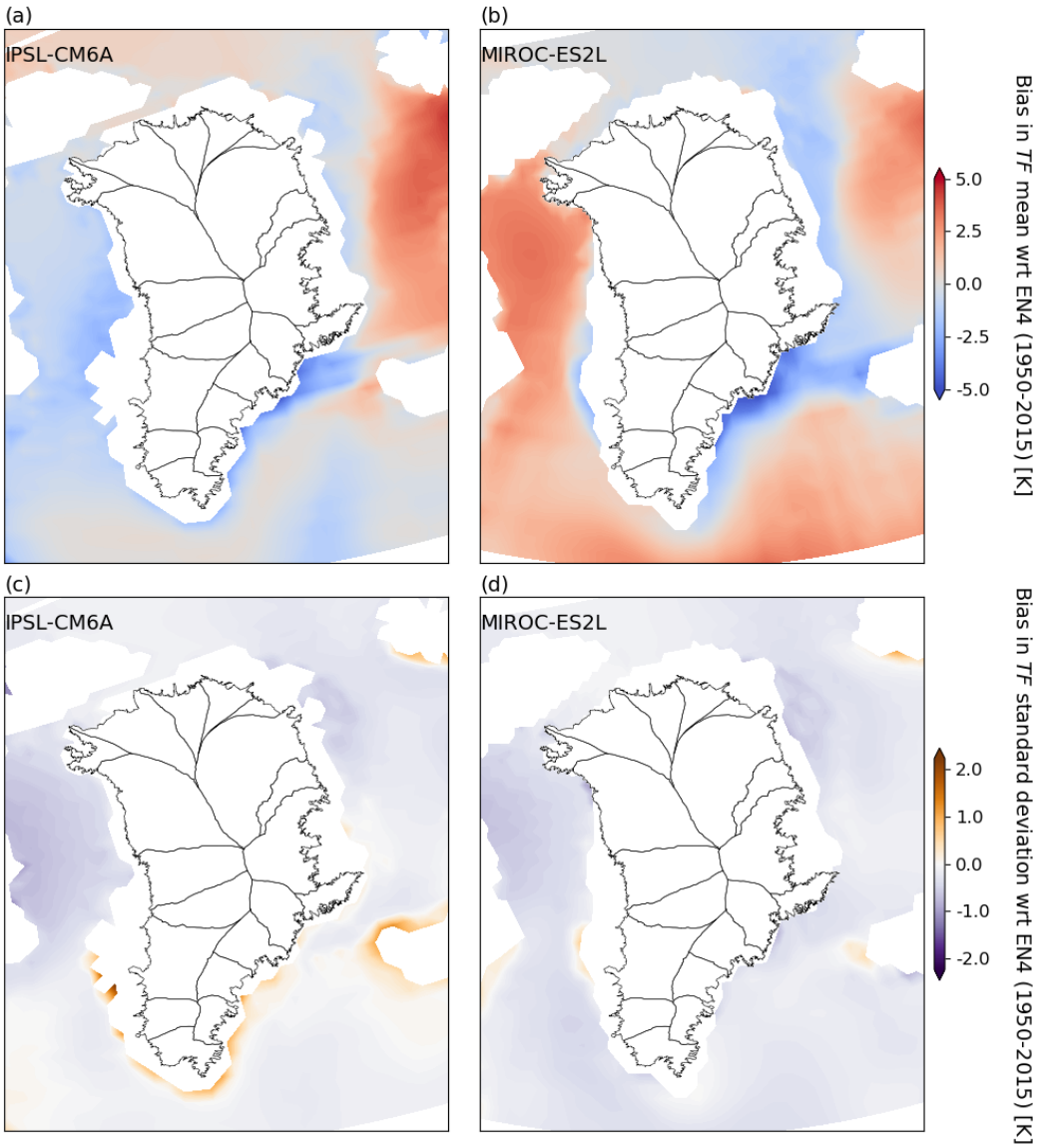

Figure 4. Bias in (a,b) mean *TF* and (c,d) standard deviation in *TF* of two AOGCMs with respect to the EN4 objective analyses (Good et al., 2013). Maps show biases of IPSL-CM6A (a,c, Boucher et al., 2020) and MIROC-ES2L (b,d, Hajima et al., 2020)

We use quantile mapping to correct AOGCM output, which is a method extensively used in climate and hydrological sciences (e.g., Gudmundsson et al., 2012; Themeßl et al., 2012; Cannon et al., 2015). In quantile mapping, the cumulative distribution function (CDF) of a variable from model output (any AOGCM in our case) is corrected to be equal to the CDF of the same variable from a reference dataset (EN4 in our case) over a given calibration period (1950-2015 in our case). The approach can be further extended to correct projected modeled output beyond the calibration period, using the quantile

delta mapping (QDM) technique developed by Cannon et al. (2015). QDM adjusts future model output by calibrating the model CDF to the reference data CDF, and by superimposing model-projected changes in the future period. We give here an overview of the QDM methodology, but refer to Cannon et al. (2015) for all the details. If readers need some visual intuition, we provide example results of QDM applied on time series in Figure 5.

We denote the TF from the observational dataset EN4 over the calibration period as TF_{obs}^{cal} with CDF F_{obs}^{cal}. The modeled TF from an AOGCM over the calibration period is denoted as TF_{mod}^{cal} with CDF F_{mod}^{cal}. Equalizing the model CDF to the observational CDF is achieved by:

$$\widehat{TF}_{mod}^{cal} = \left[F_{obs}^{cal}\right]^{-1}\left[F_{mod}^{cal}\left(TF_{mod}^{cal}\right)\right], \tag{3}$$

where $[F]^{-1}$ denotes the inverse CDF. The resulting $\widehat{TF}_{mod}^{cal}$ follows the same distribution as F_{obs}^{cal} but preserves the relative changes as modeled by the AOGCM. In QDM, we also correct projected model output TF_{mod}^{proj} with CDF F_{mod}^{proj}. In this procedure, the first step is to calculate the nonexceedance probabilities, τ, during the projection period of the modeled values at each time step t:

$$\tau_{mod}^{proj}(t) = F_{mod}^{proj}\left[TF_{mod}^{proj}(t)\right], \tag{4}$$

with $\tau_{mod}^{proj}(t)$ being the probability that, when considering the full projection period, TF_{mod}^{proj} is less than or equal to its value at t, i.e., $TF_{mod}^{proj}(t)$. Therefore, $\tau_{mod}^{proj}(t)$ ranges between 0 and 1. Similar to Cannon et al. (2015) we calculate the projection period CDF F_{mod}^{proj} over 30-year sliding windows. This approach presents the advantages of not representing the entire projection as a single CDF, and avoiding any abrupt distributional changes within the projection period. The τ_{mod}^{proj} values from Eq. (4) are passed as argument to the inverse CDF of the calibration period, $\left[F_{mod}^{cal}\right]^{-1}$, to estimate the calibration period value of TF associated to them. Taking the difference between the projected model values and their corresponding estimated calibration period values gives the estimated absolute changes in quantiles between the calibration period and the 30-year window in the projection period:

$$\Delta_{mod}(t) = TF_{mod}^{proj}(t) - \left[F_{mod}^{cal}\right]^{-1}\left[\tau_{mod}^{proj}(t)\right]. \tag{5}$$

The nonexceedance probability is also passed to $\left[F_{obs}^{cal}\right]^{-1}$ in order to calculate its corresponding bias-corrected TF value under the observational distribution of the calibration period:

$$\widehat{TF}_{obs:mod}^{cal:proj}(t) = \left[F_{obs}^{cal}\right]^{-1}\left[\tau_{mod}^{proj}(t)\right]. \tag{6}$$

In this manner, $\widehat{TF}_{obs:mod}^{cal:proj}(t)$ takes the statistical characteristics of the reference data during the calibration period. We use the physical constraint that TF cannot be below 0 K (i.e., ocean temperature at freezing point). To enforce this constraint, we find the observation CDF limit that corresponds to $TF = 0$ K by linearly extrapolating $\left[F_{obs}^{cal}\right]^{-1}$ until the limit at which it yields 0 K. If $\tau_{mod}^{proj}(t)$ is below that limit, $\widehat{TF}_{obs:mod}^{cal:proj}$ is set to 0 K. Finally, the modeled absolute change in quantiles is added back to $\widehat{TF}_{obs:mod}^{cal:proj}(t)$ to yield the projected bias-corrected value:

$$\widehat{TF}_{mod}^{proj}(t) = \widehat{TF}_{obs:mod}^{cal:proj}(t) + \Delta_{mod}(t). \tag{7}$$

Eq. (7) reintroduces the TF change signal projected by the AOGCM, $\Delta_{mod}(t)$. As such, the QDM procedure ensures that both statistical characteristics of the reference dataset, via Eq. (6), and projected changes in quantiles, via Eq. (5), are preserved. If a negative Δ_{mod} pushes $\widehat{TF}_{mod}^{proj}$ below 0 K, we set $\widehat{TF}_{mod}^{proj} = 0$ K for physical consistency.

In summary, the QDM technique maps the AOGCM CDF to the CDF of EN4 over the calibration period. This has the effect of reducing the biases shown in Fig. 4 to zero. However, QDM preserves the relative changes in time of *TF* as modeled by the AOGCMs. Finally, it allows to extend the calibration to the projection period by super-imposing the relative changes in the *TF* distribution that are projected by the AOGCM to the corrected *TF* time series.

2.2 Extrapolation to fjord mouths

The horizontal resolution of AOGCMs is insufficient to capture the dynamical processes governing ocean heat transfer from the open ocean on to the continental shelf (Slater et al., 2020; Wood et al., 2021), as these models use a typical ocean mesh of 1° (∼75 km around Greenland). In this context, we make use of a high-resolution Arctic forward model run from the Estimating the Circulation and Climate of the Ocean (ECCO) consortium (Nguyen et al., 2012; Rignot et al., 2012). This ECCO run spans the period 1992-2009, and it has a horizontal resolution of 4 km. We use ECCO to statistically constrain the extrapolation from open ocean *TF* to on-the-shelf *TF*, close to the fjord mouths of the Greenland outlet glaciers. Note here that other high-resolution ocean model products can be used for the extrapolation (e.g., Gillard et al., 2016). Hereafter, we refer to open ocean gridpoints as offshore, and to fjord mouth gridpoints as inshore. We consider inshore gridpoints as ECCO gridpoints closest to glacier fjord mouths, while offshore gridpoints can be any other gridpoint within the 7 oceanic sectors considered (see Fig. 1a). We note that some ECCO gridpoints need to be rejected due to unphysical variability in *TF*, caused in general by an initialization shock (Balmaseda et al., 2009). We conservatively discard 2% of the gridpoints, mostly located in Eastern fjords and in the far North. The statistical extrapolation method is based on, and adapted from the work of Oliver and Holbrook (2014), as detailed in this section. A visualization of an example result of the extrapolation methodology is given in Figure 8.

2.2.1 Statistical relations between offshore and inshore

We seek a representation of inshore *TF* based on available offshore *TF* in the QDM-corrected AOGCM output. We derive offshore-inshore relationships using the high-resolution ECCO output. Inshore locations are specific to each glacier of the dataset of Wood et al. (2021) (Fig. 1a), and defined here as the 4 closest ECCO gridpoints that have a bathymetry of at least 100 m depth. Other choices could be made by applying knowledge of the regional dynamics, for example by selecting gridpoints along the orientation of a fjord or along contours of constant depth. The offshore locations serve as predictors in the offshore-inshore relationships, and we provide a method to optimize the choice of predictor gridpoints in Section 2.2.2.

We decompose all *TF* (i.e., offshore, inshore, from the high-resolution ECCO, and from a coarse resolution AOGCM) in four different components: a mean, a trend, a seasonal cycle, and residual variability. This is expressed as:

$$TF = \overline{TF} + \dot{TF} + TF^S + TF', \tag{8}$$

where $\overline{TF}$, $\dot{TF}$, TF^S, and TF' denote the long-term mean, the long-term trend, the seasonal component, and the residual variability, respectively. We develop statistical extrapolation relationships for the three components $\overline{TF}$, TF^S, and TF'. We preserve $\dot{TF}$ unchanged from the offshore to the inshore for two reasons. First, it is questionable whether an offshore-inshore relation for the trend calculated over the period of the ECCO run (1992-2009) can be applied to AOGCM trends beyond that period. Second, while linear functions represent the trend over the short 1992-2009 period well, we find evidence of quadratic trends in AOGCM simulations extending until 2100 (see Sect. 2.3). As such, any extrapolation relationship derived over 1992-2009 would not be transferable to the entire AOGCM simulation period.

First, the time-mean $\overline{TF}$ is a scalar, for both offshore and inshore locations. As such, we can relate them linearly as:

$$\overline{TF}_{\text{in}} = \alpha \overline{TF}_{\text{off}}, \tag{9}$$

where α is the only parameter of the regression, and it is glacier-specific. The subscripts in and off denote inshore and offshore, respectively. We derive the α parameter from the ECCO relation between inshore and offshore time-mean TF:

$$\alpha = \frac{\overline{TF}_{\text{hr,in}}}{\overline{TF}_{\text{hr,off}}}, \tag{10}$$

where the subscript hr denotes the high-resolution ECCO model.

Second, we represent the seasonal cycle as a sum of 12 monthly effects. The monthly effect is calculated as the mean, over multiple years, of the TF anomalies, defined as the difference between the monthly value and the annual mean. For example, we compute the difference between the January TF and the annual mean TF for each year of the time series, and the January monthly effect is the mean of these anomalies. The seasonal cycle is thus represented as:

$$TF^S(t) = \sum_{i=1}^{12} M_i \delta\left(t, i\right), \tag{11}$$

where $M_{1,\ldots,12}$ are the 12 monthly effects, and $\delta\left(t, i\right)$ is 1 if the time step t of the time series falls in month i, and 0 otherwise. Note that, by definition, the 12 M_i average to zero. We favor using monthly effects rather than a Fourier series, as done by Oliver and Holbrook (2014), because the latter approach does not capture well the 0 K lower bound on TF, which can be important at high latitudes. Similarly to our approach for $\overline{TF}$, we relate the inshore and offshore monthly effects linearly for each month:

$$M_{i,\text{in}} = \gamma M_{i,\text{off}}. \tag{12}$$

The linear scaling factor γ is calibrated with the high-resolution ECCO product:

$$\gamma = \frac{\sum_{i=1}^{12} |M_{i,\text{hr,in}}|}{\sum_{i=1}^{12} |M_{i,\text{hr,off}}|}, \tag{13}$$

and γ thus captures how strongly seasonality is amplified or reduced inshore compared to offshore. We calculate a single γ with the absolute values of all 12 monthly effects rather than having a specific γ for each individual month, because individual M_i values close to zero could make the ratio unphysically large.

Finally, after subtracting the mean, the trend, and seasonality components from the TF time series, only the non-seasonal residual variability TF' remains (Eq. (8)). We apply a statistical relationship between offshore and inshore TF'. We relate the standard deviations of the inshore and offshore variability linearly:

$$\sigma\left[TF'_{\text{in}}(t)\right] = \beta \sigma\left[TF'_{\text{off}}(t)\right], \tag{14}$$

where $\sigma[\,]$ denotes the standard deviation of the time series in the brackets. We estimate the β parameter from the ratio of ECCO standard deviations in inshore and offshore residual variability:

$$\beta = \frac{\sigma\left[TF'_{\text{hr,in}}(t)\right]}{\sigma\left[TF'_{\text{hr,off}}(t)\right]}. \tag{15}$$

If the trend $\dot{TF}$ is not removed, it would be included in TF', causing β to be inflated or deflated if offshore and inshore trends differ. By detrending, we thus ensure that β captures the ratio of residual inter-annual variability in TF instead of longer-term tendencies.

Using the calibrated statistical relationships, we construct an estimate of inshore TF of a coarse-resolution AOGCM using its offshore TF as a predictor. In other words,

we can use $TF_{\mathrm{cr,off}}$ to predict $TF_{\mathrm{cr,in}}$, where the subscript cr denotes a coarse-resolution AOGCM. We predict the three extrapolated components of $TF_{\mathrm{cr,in}}$ separately:

$$\begin{cases} \overline{TF}_{\mathrm{cr,in}} = \alpha \overline{TF}_{\mathrm{cr,off}} \\ TF^{S}_{\mathrm{cr,in}}(t) = \gamma \sum_{i=1}^{12} M_{i,\mathrm{cr,off}} \delta\,(t,i) \\ TF'_{\mathrm{cr,in}}(t) = \beta\, TF'_{\mathrm{cr,off}}(t) \end{cases} . \tag{16}$$

By multiplying $TF'_{\mathrm{cr,off}}$ by β, we use the fact that the residuals have mean zero by construction, and we assume that they follow a normal distribution. As such, $TF'_{\mathrm{cr,in}}$ follows a normal distribution with mean zero and its standard deviation is $\beta\sigma\,(TF'_{\mathrm{cr,off}})$. Our Eqs. (14) and (15) differ from the approach of Oliver and Holbrook (2014), as they related $TF'_{\mathrm{hr,off}}$ and $TF'_{\mathrm{hr,in}}$ by linear regression to estimate β. By using the ratio of the standard deviations instead, we avoid issues of decreased variability in $TF'_{\mathrm{cr,in}}$ compared to $TF'_{\mathrm{cr,off}}$ when $TF'_{\mathrm{hr,in}}$ and $TF'_{\mathrm{hr,off}}$ show similar variability, but are not well-correlated in time. However, our approach relies on the assumption of normally-distributed residuals. To verify this assumption, we draw 1000 random bootstrap samples of size 50 from all the monthly values of residual variability, and compute the Shapiro-Wilk normality test (Shapiro & Wilk, 1965). We find that the normality assumption is validated for 87% of these bootstrap samples at the 5% significance level.

Using Eqs. (8) and (16), the full estimated time series of $TF_{\mathrm{cr,in}}$ is reconstructed as:

$$TF_{\mathrm{cr,in}} = \overline{TF}_{\mathrm{cr,in}} + \dot{TF}_{\mathrm{cr,off}} + TF^{S}_{\mathrm{cr,in}} + TF'_{\mathrm{cr,in}}. \tag{17}$$

Note that the trend is preserved from the offshore AOGCM gridpoint. Because the extrapolation is based on the short period of ECCO output (1992-2009), we simply remove a linear trend for the extrapolation process, as a higher-order polynomial would likely include part of the residual variability which we aim to extrapolate.

For a given glacier front (i.e., a given inshore location), we need to determine an optimal offshore predictor gridpoint of the coarse-resolution AOGCM (see Sect. 2.2.2). However, the parameters α, γ, and β are constrained with relations from ECCO. Thus, once the AOGCM offshore gridpoint predictor is determined, we find its nearest-neighbor ECCO gridpoint. We relate the $TF_{\mathrm{hr,off}}$ of this ECCO gridpoint to the ECCO $TF_{\mathrm{hr,in}}$ of the glacier front to constrain the parameters. As a reminder, $TF_{\mathrm{hr,in}}$ for a given glacier front is computed as the average of the 4 ECCO gridpoints closest to the glacier front.

2.2.2 Determining the offshore predictor locations

In this section, we describe a procedure, modified from Oliver and Holbrook (2014), to determine an optimal offshore gridpoint as a predictor for the inshore AOGCM TF time series of a given glacier front. We limit the possible domain for an offshore predictor to our 7 oceanic sectors around Greenland (Fig.1a). To determine the optimal predictors, we use three quantitative criteria for each TF component.

The first criterion is a quality function, Q. The quality function is defined as the agreement between the QDM-corrected coarse-resolution AOGCM and the high-resolution ECCO. The quality is defined separately for each component:

$$\overline{Q} = \frac{\left|\overline{TF}_{\mathrm{cr,off}} - \overline{TF}_{\mathrm{hr,off}}\right|^{-1}}{1 + \left|\overline{TF}_{\mathrm{cr,off}} - \overline{TF}_{\mathrm{hr,off}}\right|^{-1}}, \tag{18}$$

$$Q^{S} = \left[\frac{1 + r\left(TF^{S}_{\mathrm{cr,off}}, TF^{S}_{\mathrm{hr,off}}\right)}{2} \times \min\left(\frac{\sum_{i=1}^{12}|M_{i,\mathrm{cr,off}}|}{\sum_{i=1}^{12}|M_{i,\mathrm{hr,off}}|}, \frac{\sum_{i=1}^{12}|M_{i,\mathrm{hr,off}}|}{\sum_{i=1}^{12}|M_{i,\mathrm{cr,off}}|}\right)\right]^{1/2}, \tag{19}$$

$$Q' = \left[\frac{1 + r\left(TF'_{\mathrm{cr,off}}, TF'_{\mathrm{hr,off}}\right)}{2} \times \min\left(\frac{\sigma\left(TF'_{\mathrm{cr,off}}\right)}{\sigma\left(TF'_{\mathrm{hr,off}}\right)}, \frac{\sigma\left(TF'_{\mathrm{hr,off}}\right)}{\sigma\left(TF'_{\mathrm{cr,off}}\right)}\right)\right]^{1/2}, \tag{20}$$

where $r(\,)$ denotes the correlation coefficient. Our expressions for the quality functions slightly differ from those of Oliver and Holbrook (2014). Note that Eq. (18) avoids extreme values of $\overline{Q}$ in the case of $\overline{TF}$ being close to zero, while the expressions of Eqs. (18, 19, 20) all ensure that $\overline{Q}$, Q^S, and Q' remain bounded between 0 and 1. By using the quality functions as a criterion for selecting the offshore predictor gridpoint of the coarse-resolution AOGCM, we favor offshore gridpoints where the QDM-corrected AOGCM agrees well with ECCO. The two terms in Eq. (19) favor strong temporal correlation between the seasonality patterns of ECCO and the AOGCM and similar total seasonal amplitude, respectively. Similarly, the two terms in Eq. (20) favor temporal correlation between and similar variability amplitude of the residuals of ECCO and the AOGCM. In Eqs. (19, 20), we give equal weights to the correlation and amplitude terms by taking their geometric mean.

The second criterion is a strength function, S. The strength functions measure the similarity between the offshore TF components to their corresponding components of the inshore gridpoint of interest in the high-resolution model ECCO. We compute the strength functions at all the ECCO gridpoints that are a nearest neighbor of an existing AOGCM gridpoint. In this manner, each nearest neighbor is attributed a strength criterion value. In other words, the values of the strength functions of a given coarse-resolution AOGCM gridpoint depend on the TF time series of the nearest-neighbor high-resolution ECCO model gridpoint. The expressions for the strength functions are similar to those of the quality functions, and are also defined separately for each component:

$$\overline{S} = \frac{\left|\overline{TF}_{\mathrm{hr,off}} - \overline{TF}_{\mathrm{hr,in}}\right|^{-1}}{1 + \left|\overline{TF}_{\mathrm{hr,off}} - \overline{TF}_{\mathrm{hr,in}}\right|^{-1}}, \tag{21}$$

$$S^S = \left[\frac{1 + r\left(TF^S_{\mathrm{hr,off}}, TF^S_{\mathrm{hr,in}}\right)}{2} \times \min\left(\frac{\sum_{i=1}^{12}|M_{i,\mathrm{hr,off}}|}{\sum_{i=1}^{12}|M_{i,\mathrm{hr,in}}|}, \frac{\sum_{i=1}^{12}|M_{i,\mathrm{hr,in}}|}{\sum_{i=1}^{12}|M_{i,\mathrm{hr,off}}|}\right)\right]^{1/2}, \tag{22}$$

$$S' = \left[\frac{1 + r\left(TF'_{\mathrm{hr,off}}, TF'_{\mathrm{hr,in}}\right)}{2} \times \min\left(\frac{\sigma\left(TF'_{\mathrm{hr,off}}\right)}{\sigma\left(TF'_{\mathrm{hr,in}}\right)}, \frac{\sigma\left(TF'_{\mathrm{hr,in}}\right)}{\sigma\left(TF'_{\mathrm{hr,off}}\right)}\right)\right]^{1/2}. \tag{23}$$

Strength functions thus attribute more weight to offshore gridpoints where ECCO estimates that a given TF component is similar to the inshore TF component close to the glacier front.

The last criterion is a simple localization function, L. It is formulated as an isotropic two-dimensional Gaussian function centered on the inshore location $(x_{\mathrm{in}}, y_{\mathrm{in}})$. The localization function attributes more weight to offshore gridpoints closer to the inshore location, and it is the same for the three TF components:

$$L = \exp\left[\frac{-\left((x_{\mathrm{off}} - x_{\mathrm{in}})^2 + (y_{\mathrm{off}} - y_{\mathrm{in}})^2\right)}{2\lambda_L^2}\right], \tag{24}$$

where λ_L is an isotropic decay lengthscale. We set $\lambda_L = 600$ km. This is chosen because the mean correlation between TF time-series within a given radius in ECCO falls below 0.7 for a radius above 600 km, when considering the 7 oceanic sectors around Greenland (Fig.1a). However, the decay lengthscale can be adjusted depending on the importance attributed to keeping offshore predictors close to the inshore location. Furthermore, Eq. (24) can easily be converted to an anisotropic function, with different decay scales in the zonal and meridional directions.

The functions Q, S, and L are combined into a single cost function J. In this process, one can use different weights ranging between 0 and 1 for the three functions, w_Q, w_S, and w_L,

$$J = \frac{1}{(1 - w_Q + w_Q Q)\,(1 - w_S + w_S S)\,(1 - w_L + w_L L)}, \tag{25}$$

and there is one cost function J per TF component. Here, we make the simplest choice of equally-weighted functions, with $(w_Q, w_S, w_L) = (1,1,1)$. While Oliver and Holbrook (2014) used a different predictor gridpoint per TF component, we find that the spatial differences between the cost functions of the different TF components are minor. Thus, for simplicity, we compute the total cost J_T as the average of the cost functions of the three components. The gridpoint minimizing J_T is used as the offshore gridpoint predictor. We provide examples of the searching method for the optimal offshore predictor gridpoint in Sect. 3.2. We use the optimal offshore predictor to construct a synthetic inshore TF time series following the method described in Sect. 2.2.1.

2.3 Fitting of statistical models

The QDM-corrected and inshore-extrapolated TF time series are entirely deterministic, but our goal is also to generate many TF time series that represent realizations of internal climate variability. Therefore, we calibrate statistical time series models to the deterministic time series, and use a stochastic term to represent residual variability in TF. Here, we consider output from two different AOGCMs, under two possible emission scenarios, and at 226 Greenland marine glaciers. Each combination of AOGCM and emission scenario has a given number of ensemble members. The ensemble members differ due to minor differences in the initial conditions, which can amplify during the AOGCM simulation owing to the sensitive dependence on initial conditions of the climate system (Hasselmann, 1976; Kay et al., 2015; Maher et al., 2019). For each individual glacier, the number of TF time series available from a given AOGCM under a specific emission scenario is limited by the number of ensemble members. Because AOGCMs are computationally expensive models, the number of members is in general of the order 1 to 10 for CMIP6 experiments until 2100. Statistical samples of such sizes of TF time series are thus too small to fully characterize the distribution of possible oceanic conditions affecting Greenland glaciers in the future. The statistical models that we develop here must (i) be representative of the deterministic inshore series, (ii) account for internal climate variability using a stochastic parameterization, and (iii) efficiently generate a large number of TF time series at low computational expense. Each statistical model is specific to a combination of AOGCM, emission scenario, and glacier. We detail their implementation in this section.

At a given inshore location, we process all the TF time series in the same manner (Eq. (8)). First, we decompose each series in a mean, a trend, a seasonality, and a residual component. In contrast to the procedure in the extrapolation process (Sect. 2.2.1), time series here are considered over the entire AOGCM run (1850-2100) and not only over their overlapping period with ECCO (1992-2009). For this reason, we estimate the sum of mean and trend components as a piecewise polynomial function with a breakpoint at date t_{brk}. We use a second-order polynomial, because it captures the non-linear warming trend by the end of the century in high-emission scenarios. Mathematically, the mean-plus-trend component that we estimate corresponds to:

$$\begin{cases} \overline{TF} + \dot{TF}(t) = a_0^{pre} + a_1^{pre}(t - t_0) + a_2^{pre}(t - t_0)^2 & \text{if } t \leq t_{brk} \\ \overline{TF} + \dot{TF}(t) = a_0^{post} + a_1^{post}(t - t_{brk}) + a_2^{post}(t - t_{brk})^2 & \text{if } t > t_{brk} \end{cases}, \tag{26}$$

where t_0 is the initial date of the time series. The breakpoint t_{brk} is not fixed but varies between different TF time series to optimize the fit of Eq. (26). The *pre* and *post* superscripts denote coefficients applying before and after the breakpoint t_{brk}, respectively. Ensemble members of a same AOGCM agree relatively well on general trends at the centennial time scale (Maher et al., 2019). For this reason, our method does not attribute internal variability to the mean-plus-trend component, which we take as entirely deterministic.

The seasonality is still evaluated as 12 monthly effects (Eq. (11)). However, we observe that seasonality in TF strongly increases by 2100 in high-emission scenarios, both

in raw AOGCM output and in our QDM-corrected inshore-extrapolated time series. For this reason, over the period 1850-2100, we represent each monthly effect M_i as a piecewise linear function with a breakpoint fixed at 2015:

$$\begin{cases} M_i(t) = b_0^{pre} + b_1^{pre}\,(t - t_0) & \text{if } t \leq 2015, \text{for } i = 1, ..., 12 \\ M_i(t) = b_0^{post} + b_1^{post}\,(t - 2015) & \text{if } t > 2015, \text{for } i = 1, ..., 12 \end{cases}, \tag{27}$$

and all the M_i values can be used to calculate the seasonal component (Eq. (11)). Assuming a single breakpoint avoids having different optimal breakpoints for different months, which would complicate the representation of seasonality. Here, we elect 2015 as the breakpoint because it is the transition between historical and prediction simulations of AOGCMs in the CMIP6 protocol. Finally, after removing the mean-plus-trend component (using Eq. (26)), and the seasonality component (using Eq. (27)) from a given *TF* series, we treat the remaining signal as the residual component, *TF'*.

The residual component is modeled as an autoregressive-moving-average (ARMA) process on an annual time scale. ARMA models have been principally popularized by Box and Jenkins (1976), and are extensively used in geophysical sciences to represent dynamical processes that exhibit memory (Storch & Zwiers, 1999; Mudelsee, 2010; Wilks, 2011). Their advantage is the ability to represent a large range of temporal autocorrelation features, while using only a small number of parameters. ARMA processes have been used previously to represent ocean temperatures and their residuals (e.g., Hasselmann, 1976; Hausfather et al., 2017; Cheng et al., 2022). It is important to remove the general trend, as well as the trend in monthly effects to ensure stationarity of the residual component time series, i.e., the residuals should not exhibit a trend or changing variability over time. By modeling the residual component as an ARMA process, our goal is to capture the different time scales of variability in oceanic conditions around Greenland, ranging from inter-annual to multi-decadal variability (Straneo & Heimbach, 2013). This temporal complexity is partly driven by the influence of the North Atlantic Oscillation and the Atlantic Multi-decadal Oscillation on variability in waters of both Polar and Atlantic origins (Dickson et al., 2000; Rignot et al., 2012). An ARMA model of autoregressive (AR) order p and moving-average (MA) order q, denoted ARMA(p,q), for a generic variable y is formulated as:

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \tag{28}$$

where the φ_1, ..., φ_p are the AR coefficients, and the θ_1, ..., θ_q are the MA coefficients. The ϵ_t term is a Gaussian noise term. In an ARMA model, the φ_i coefficients capture the memory of the process, and the θ_i coefficients represent the persistence of random noise effects in the system. In our specific case, ϵ_t allows us to prescribe covariance between different glaciers by being randomly sampled from a multivariate Gaussian of which the covariance matrix has dimensions equal to the total number of glacier front locations (i.e., 226). From this multivariate Gaussian, one can sample a random vector $\boldsymbol{\epsilon_t}$ of size equal to the number of glaciers, and which consists of the individual ϵ_t applied at each glacier front:

$$\boldsymbol{\epsilon_t} \sim N(0, \Sigma). \tag{29}$$

For each of the *TF'* time series (i.e., the residual variability), we calibrate all possible combinations of ARMA models of both AR orders (p) and MA orders (q) ranging from 0 to 4. For each possible ARMA(p,q) calibrated model, we evaluate its Bayesian Information Criterion (BIC) (Schwarz, 1978). In fitting a statistical model to data, the BIC measures the likelihood of the model, but penalizes the number of parameters used in the model. Its purpose is to find the adequate balance between model fit and model complexity, thus avoiding overfitting. It is calculated as:

$$\text{BIC} = m \ln(n) - 2 \ln\left(\hat{L}\right), \tag{30}$$

where m is the total number of parameters, n is the number of data points, and $\hat{L}$ is the value of the likelihood function evaluated with the calibrated model. In our case, $m = p+q$ and $n = 250$ (i.e., number of yearly residual values in a 1850-2100 time series). For a given glacier, emission scenario, and AOGCM, there is one combination of (p,q) minimizing the BIC per individual ensemble member. To decide which of all these potential best-fitting ARMA models we use, we take the most favored (p,q) combination among the selected (p,q) combinations for the different members, and denote this preferred combination as (p^*,q^*). In case of a tie among the selected models, we take the one of lower order (i.e., lower $p+q$). As such, for a given glacier, emission scenario, and AOGCM, we have a single ARMA(p^*,q^*) model representative of the residuals. To calibrate the φ_i coefficients, θ_j coefficients, and the marginal variance of ϵ_t (see Eq.(28)), we fit all the ensemble member time series with the (p^*,q^*) combination. We take the average of the φ_i, θ_j, and marginal variance values across the ARMA(p^*,q^*) models of all the ensemble members. This averaging procedure is equivalent to considering the parameter values from the different ensemble members as a sample of the true population of AOGCM parameter values. As per the central limit theorem, the average parameter values are thus asymptotically representative of a mean AOGCM simulation. This procedure is repeated for each combination of glacier, AOGCM, and emission scenario. The calibrated ARMA(p^*,q^*) model allows us to generate a large set of time series at a given glacier, representative of the *TF'* residuals modeled by the corresponding AOGCM under the forcing of the corresponding emission scenario.

The long-term mean and trend (Eq. (26)) and monthly effects (Eq. (27)) in *TF* are taken from the across-members AOGCM ensemble mean at each glacier and for each emission scenario. These ensemble mean deterministic components can be added to an ARMA-generated stochastic *TF'* residuals time series to produce a *TF* time series.

We can account for inter-glacier covariability when generating the ϵ_t, as highlighted by Eq. (29). In other words, the calibrated marginal variances are used to fill the diagonal entries of the covariance matrix, but the off-diagonal entries can be estimated from the empirical correlation of the *TF'* time series of a same member at all the glacier front locations. Because the number of values to be estimated in the covariance matrix is very large compared to the number of yearly samples of *TF'*, it is preferable to compute a sparse correlation matrix instead of relying on the empirical correlation matrix (e.g., Hu & Castruccio, 2021). We employ the commonly-used graphical lasso method to compute a sparse correlation matrix $\hat{C}_i$ for the ith ensemble member (Friedman et al., 2008). We average the $\hat{C}_i$ across all the ensemble members to yield an estimated correlation matrix $\hat{\overline{C}}$ for a combination AOGCM-emission scenario. From $\hat{\overline{C}}$ and the diagonal matrix of the individual calibrated marginal variances at all the glacier front locations, $\overline{K}$, we compute the sparse covariance matrix $\hat{\overline{\Sigma}}$:

$$\hat{\overline{\Sigma}} = \overline{K}^{\frac{1}{2}} \hat{\overline{C}} \overline{K}^{\frac{1}{2}}, \tag{31}$$

and $\hat{\overline{\Sigma}}$ is used as the covariance matrix for generating random *TF'* realizations across all the glaciers, following Eqs. (28, 29). As a reminder, $\hat{\overline{\Sigma}}$ is specific to a given AOGCM-emission scenario configuration, is an average across all the ensemble members, and captures covariance between *TF'* at all the glacier front locations. In the Results section, Figure 15 gives a visual intuition for the difference between an empirical and a sparse correlation matrix.

To summarize, the stochastic time series models described in this subsection provide a computationally efficient way to generate time series of variability in *TF*. The parameters of the models are calibrated to the temporal variability characteristics of the deterministic QDM-corrected and inshore-extrapolated AOGCM *TF* time series. In addition, the time series models can represent spatial covariance in *TF*. This latter aspect is critical when evaluating the response to oceanic forcing of the Greenland ice sheet as a whole, instead of on a glacier-by-glacier basis.

3 Results

To illustrate our method, we use outputs from two AOGCMs that took part in CMIP6. The AOGCMs chosen for this study are MIROC-ES2L (Hajima et al., 2020) and IPSL-CM6A (Boucher et al., 2020), as these models are updated versions of two of the AOGCMs used in the ISMIP6 Greenland intercomparison (Goelzer et al., 2020). However, the method outlined in Sect. 2 is applicable to any AOGCM. We consider both AOGCMs in a low- and a high-emission scenario: ssp126 and ssp585, respectively. The forcing scenarios cover the period 2015-2100, and are preceded by the AOGCM historical runs over 1850-2015. Note here that we only use the 1950-2015 period as calibration period, because it overlaps the historical AOGCM run and the period of high observational weight values in EN4 (Fig. 3). Under these emission scenarios, MIROC-ES2L and IPSL-CM6A have 10 and 5 ensemble members, respectively. Finally, we investigate *TF* conditions at the 226 outlet glaciers with marine termini (Wood et al., 2021) (see Fig. 1a). In total, our analysis involves 6780 *TF* time series.

3.1 Quantile Delta Mapping results

In this section, we illustrate and evaluate the QDM approach described in Section 2.1. Figure 5 shows examples of the QDM application at two different locations using member r1 of MIROC-ES2L. In Figure 5a, the raw AOGCM *TF* shows a positive bias of $\sim$ 3 K, and an underestimation of temporal variability with respect to EN4. After QDM, these differences are corrected, while the features of relative changes and future trends in the raw AOGCM output are preserved. In contrast, in Figure 5b, the AOGCM is in good agreement with EN4, both for the mean and the amplitude of variability. As such, the QDM-corrected *TF* time series remains similar to the raw time series.

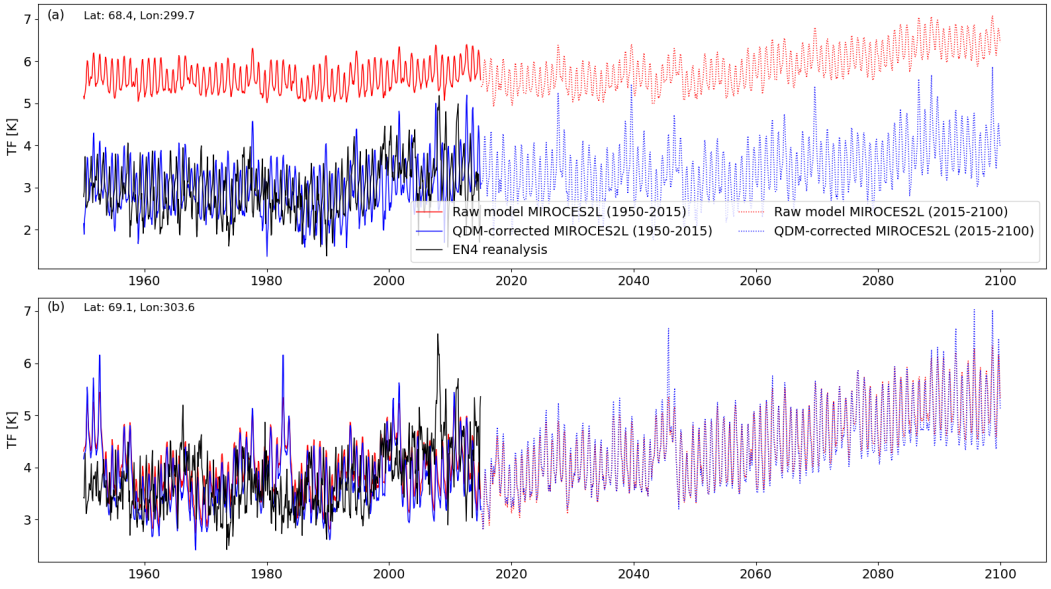

Figure 5. *TF* time series for Member r1 of ssp585 experiment from MIROC-ES2L at two different locations (a and b, at fa and fb in Fig. (1b)). QDM is applied on the raw MIROC-ES2L *TF* time series using the *TF* time series of the EN4 nearest neighbor gridpoint. The black line shows the EN4 nearest neighbor *TF* time series. Red lines show the *TF* time series from MIROC-ES2L. Blue lines show *TF* time series of the MIROC-ES2L after QDM correction. Dotted lines show the projection of the MIROC-ES2L *TF* time series beyond the calibration period (1950-2015). The two locations are chosen for illustrative purposes.

In order to validate the QDM method, we evaluate the direct match between modeled *TF* pre- and post-QDM correction with *TF* from in-situ CTD profile data. We use raw CTD profile data available from the EN4 dataset (Good et al., 2013). Because CTD profile data are used as a constraint in the EN4 objective analyses, they cannot serve as independent data to evaluate the improvement in model fidelity with respect to observations after QDM correction. For this reason, we perform a second QDM calibration with only a subset of the calibration period: 1950-2000 instead of 1950-2015. The CTD data from the period 2000-2015 is subsequently used to evaluate the models after the QDM procedure was applied over this reduced 1950-2000 calibration period. In this way, the EN4 objective analyses used for the QDM-correction are independent of the raw CTD data used for the model evaluation. We compute summary statistics for the agreement between *TF* from the CTD profile observations with the raw AOGCM *TF* and with the QDM-corrected AOGCM *TF*. Specifically, from the EN4 dataset, we take all the CTD profile data within the 7 oceanic sectors outlined in Figure 1a, extending at least until 500 m depth, and falling into the period 2000-2015. For each individual CTD profile, we find the nearest AOGCM gridpoint, and compute its *TF* at the time step closest to the CTD collection date. MIROC-ES2L and IPSL-CM6A have 10 and 5 members, respectively. Because there are 17 390 CTD profiles fulfilling our conditions, this results in 173 900 and 86 950 pairwise comparisons of modeled versus observed *TF* for MIROC-ES2L and IPSL-CM6A, respectively. The pairwise comparisons and summary statistics are shown in Figure 6. For MIROC-ES2L, the QDM correction reduces the bias magnitude by 42%. The Root Mean Squared Error (RMSE) is reduced by 29% and the proportion of variance explained (R^2) improves from 0.62 to 0.78. For IPSL-CM6A, the negative bias is amplified after the QDM correction from -0.11 to -0.29 K. However, the RMSE decreases by 24% and R^2 increases from 0.63 to 0.83. These results show that, even in a case where calibration is performed only using the EN4 objective analyses over 1950-2000, QDM improves the agreement of AOGCMs *TF* with respect to 2000-2015 CTD profiles.

3.2 Extrapolation results

For each inshore location, our method selects one offshore predictor location by minimizing the average cost function (Eq. (25)) for the mean ($\overline{TF}$), seasonality (TF^S), and residual variability (*TF'*) components (see Section 2.2.2). As an example of offshore predictors selection, we show some total cost functions, J_T, for member r1 of IPSL-CM6A. We show J_T for three of the largest Greenland outlet glaciers, and situated in geographically distinct areas: Helheim glacier in the South-East (Fig. 7a), Sermeq Kujalleq (also called Jakobshavn Isbræ) in the Central-West (Fig. 7b), and Petermann glacier in the North (Fig. 7c) (see also Fig. 1b for the glacier locations). For Helheim glacier, the predictor location is further North, upstream along the East Greenland Current (Fig. 1b). This current, originating at about 80° North and flowing southward, provides a physical connection to the offshore predictor (Strass et al., 1993). In this case, this location is preferred to closer gridpoints because of stronger discrepancies of the AOGCM with respect to ECCO close to the coast. The predictor location of Sermeq Kujalleq shows the shortcoming of the bathymetry used by coarse-resolution AOGCMs: the extent of the white areas in the map shows where ocean depth is less than 100 m in the AOGCM, whereas ECCO can simulate ocean dynamics up to the fjord mouth (see Fig. 1b). Finally, the predictor for Petermann glacier shows that predictability for *TF* in the Northern most latitudes is very low for the rest of the ocean around Greenland. The limit between low- and high-predictability areas is likely attributable to the location of the southern edge of sea-ice in winter and to the respective influences of Atlantic and Arctic waters (Straneo et al., 2022).

For the same member r1 of IPSL-CM6A, we show the time series resulting from the inshore extrapolation process at Helheim glacier in Figure 8. In this particular case, the extrapolation results in a cooler inshore mean, and enhanced seasonality and residual variability. Figure 9 shows the effect of the extrapolation for all the members of the

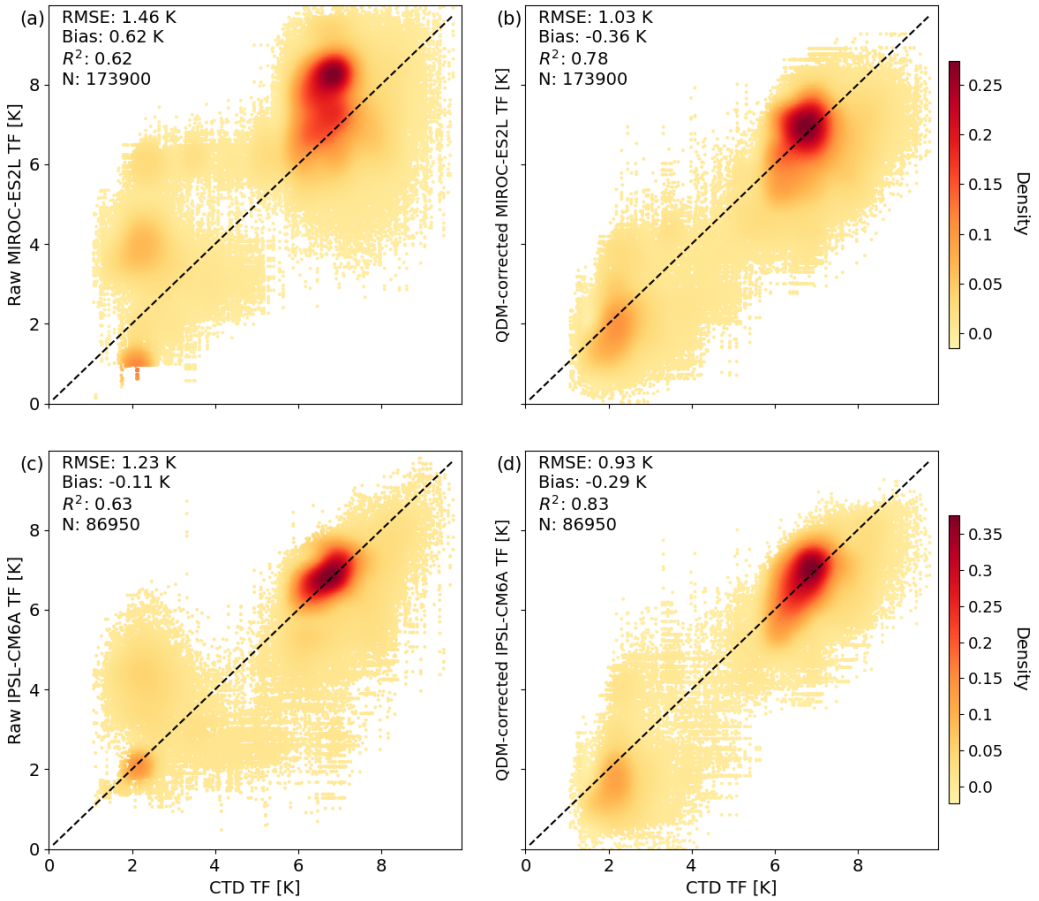

Figure 6. Comparison of raw AOGCM *TF* (a, c) and QDM-corrected AOGCM *TF* (b, d) with in-situ CTD profile data. For the purpose of evaluation, the calibration period is limited to 1950-2000, and the fit statistics are evaluated with respect to CTD data from 2000-2015 only. Evaluation is performed for MIROC-ES2L (a, b) and IPSL-CM6A (c, d). RMSE, R^2, and N denote root mean squared error, coefficient of determination, and total number of pairwise comparisons, respectively.

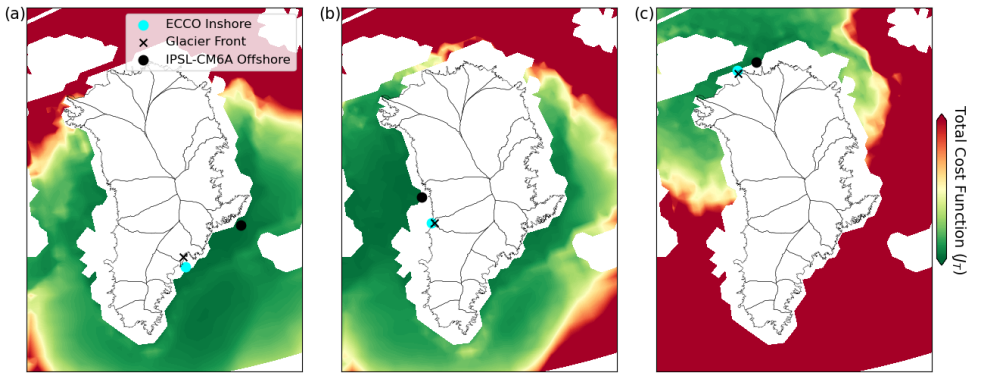

Figure 7. Optimal offshore predictors for r1 member of the IPSL-CM6A AOGCM. Shown at Helheim glacier (a), Sermeq Kujalleq (b), and Petermann glacier (c).

two AOGCMs in both emission scenarios and for all glaciers. Figure 9a shows boxplots of three parameters: (i) α, the inshore-to-offshore ratio in mean TF ($\overline{TF}$) (see Eq. (10)), (ii) γ, the inshore-to-offshore ratio in the total amplitude of the monthly effects in TF (see Eq. (13)), and (iii) β the inshore-to-offshore ratio in the standard deviation of the residual variability in TF (TF') (see Eq. (15)).

The parameter α has a mean of 0.84, and an inter-quartile range of [0.75;0.99] (Fig. 9a). $\overline{TF}$ is thus, on average, slightly reduced from the offshore to the inshore. Potential physical causes are (i) shallower bathymetry blocking access to warm Atlantic Waters, (ii) Arctic Waters being carried along the coast via the East Greenland Coastal Current and, subsequently, via the West Greenland Current (Fig. 1b), (iii) vertical mixing during the transit from the shelf towards the glacier, and (iv) the cooling effect of freshwater discharge from the ice sheet (Buch, 2002).

The ratio in amplitude of monthly effects between inshore and offshore derived from ECCO (γ) spans a larger range of values (Fig. 9a). The median (1.10) and mean (1.11) show that the distribution of γ is approximately centered around 1. However, the distribution is characterized by a high number of outliers in both the lower- and upper-end, with minimum and maximum values of 0.17 and 2.94, respectively. This large range of values suggests that processes influencing differences between offshore and inshore seasonal amplitude vary between geographical regions. This is illustrated by the map of seasonal amplitude in ECCO, which we approximate as half the difference between maximum and minimum monthly effects (Fig. 10a). In the South-East, Central-East, and South-West basins, seasonal amplitude inshore is substantially larger. On the other hand, the difference between offshore and inshore amplitude is small in the Central-West, North-West, North, and North-East basins. These between-regions differences explain the large range of γ.

Finally, the β parameter is closely centered around 1 (Fig. 9a, mean=0.99, median=1.01), with no strong outlier. Figure 10b shows that ECCO residual variability (TF') is larger closer to the coast, which should favor β values larger than 1. The reason for β remaining close to 1 is that offshore predictors are generally located close to the inshore location, as the examples in Figure 7 show. As such, inshore locations and offshore predictor locations have, on average, equal standard deviation in residual variability in ECCO.

3.3 Statistical models of the residual component

Our procedure for isolating the residual variability by removing the trend and seasonality components (Eqs. (26),(27)) effectively renders the TF' time series stationary. This is validated as the null hypothesis of non-stationarity in the Augmented Dickey Fuller test (Dickey & Fuller, 1981) is rejected with significance for all the 6780 TF' time series (p-values$<$0.05). For each TF' time series, we calibrate ARMA models (Eq. (28)) and select the best-fitting ARMA model by minimizing the BIC (Eq. (30)) among the models tested (see Sect. 2.3). Histograms in Figure 11 show the optimal (p,q) orders selected. Our procedure finds that, in general, low order ARMA models fit the TF' time series best. An autoregressive order of 1 (i.e., $p = 1$) is selected for 51 % of the cases, and more specifically an ARMA(1,0) (i.e., AR(1)) for 42 % of the cases. None of the TF' time series have a best-fitting ARMA model that is pure white noise (i.e., $p = 0$, $q = 0$), and only 3 ($<$0.1%) have the most complex ARMA model tested (i.e., $p = 4$, $q = 4$) as the best-fitting model.

Our goal is to find an ARMA model representative of TF' at a given glacier for an AOGCM under a specific emission scenario. In other words, we derive a single ARMA model from the multiple realizations provided by the different ensemble members of an AOGCM. As explained in Sect. 2.3, we find the across-members optimal ARMA(p^*,q^*) combination, calibrate it to all ensemble members, and average the noise variance, the φ_i, and the θ_j coefficients across the different members (see Eq. (28)). To generate a TF'

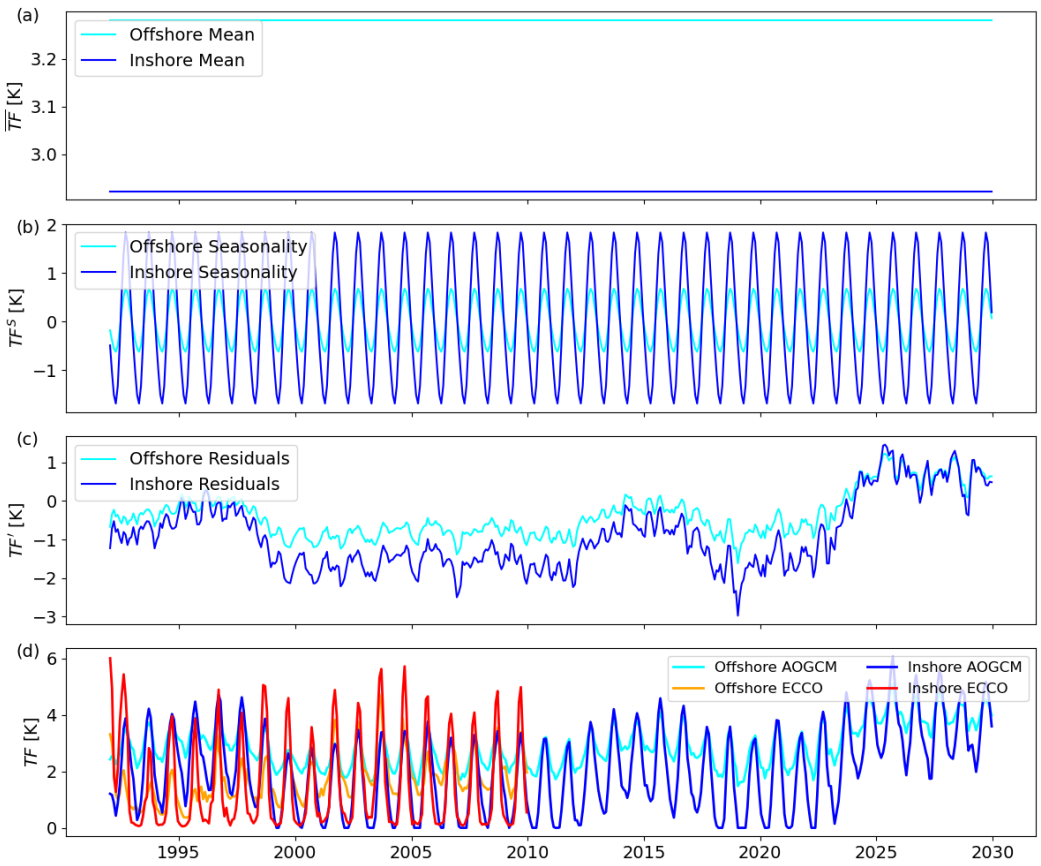

Figure 8. Offshore and extrapolated inshore TF components for r1 member of the IPSL-CM6A AOGCM in emission scenario ssp585, shown at Helheim glacier (see Fig. 1b for location). Time series of (a) offshore and extrapolated inshore mean component ($\overline{TF}$), (b) offshore and extrapolated inshore seasonality component (TF^S), and (c) offshore and extrapolated inshore residual variability component (TF'). Time series (d) of the total offshore and extrapolated inshore TF. Time series are only shown over the period 1992-2030 for the sake of figure clarity.

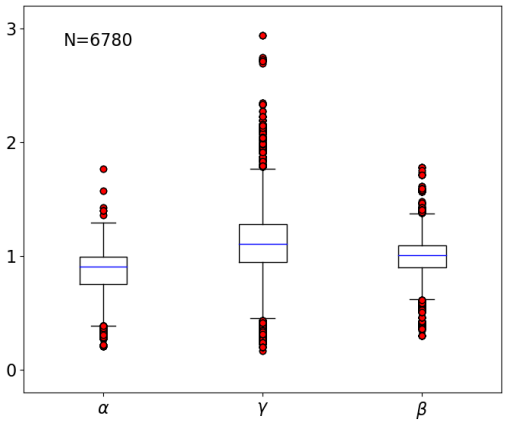

Figure 9. Statistics of the offshore-inshore relations at all glaciers for all the ensemble members of all the AOGCMs in all the emission scenarios considered in this study (total of 6780 offshore-inshore relations). Boxplots of the ratio in $\overline{TF}$ (α, see Eq. (10)), ratio in the sum of absolute monthly effects (γ, see Eq. (13)), and ratio in standard deviation of the residual variability (β, see Eq. (15)).

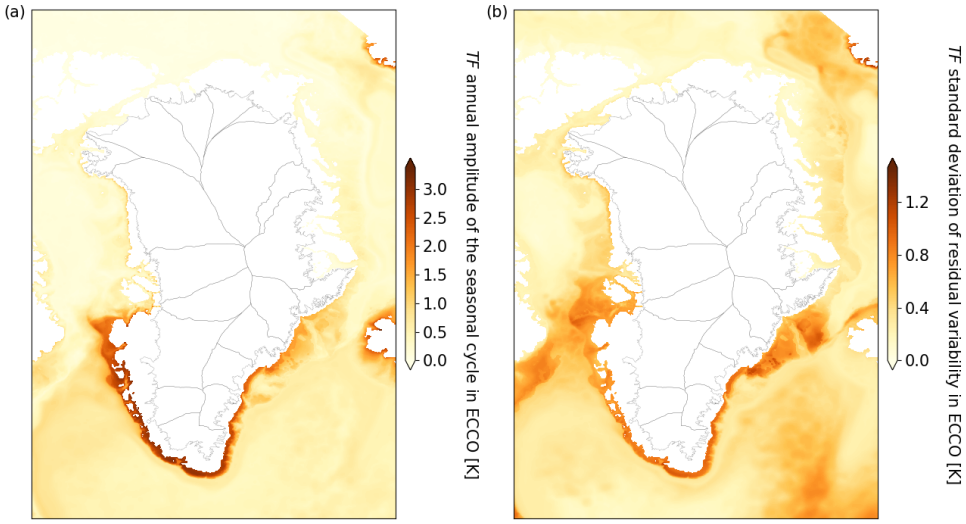

Figure 10. Maps of the high-resolution (4 km × 4 km) ECCO: (a) amplitude of the seasonal cycle, and (b) standard deviation in the residual variability. The amplitude of the seasonal cycle is half the difference between the maximum and minimum monthly effects (see Eq. (11)).

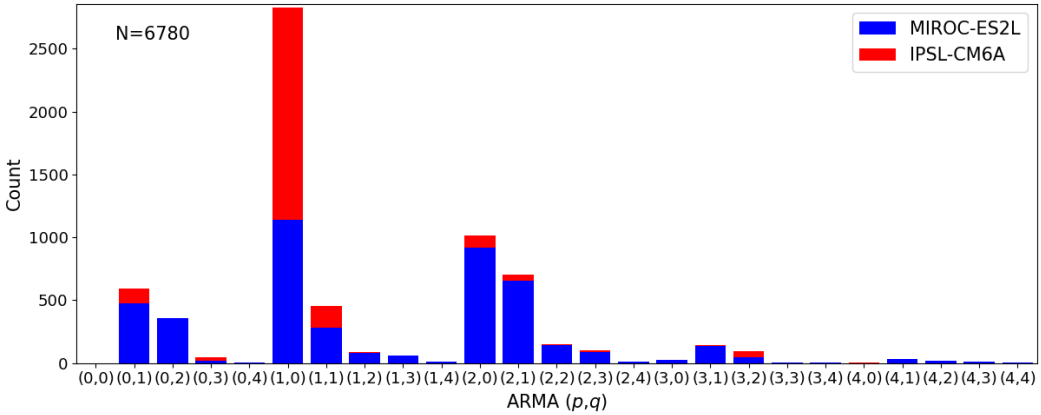

Figure 11. Histograms of best fitting ARMA models for the 6780 QDM-corrected, inshore-extrapolated time series of residual variability in *TF*. Selection is based on the BIC. The autoregressive order is p, and the moving-average order is q in Eq. (28).

time series at a glacier for a given AOGCM-emission scenario combination, we can then use the ARMA model corresponding to that glacier. Adding back the mean, trend, and seasonality component yields a statistically-generated *TF* time series for the glacier. The mean-plus-trend (see Eq. (26)) and seasonality components (see Eq. (27)) are taken from the across-member ensemble mean.

The statistically-generated *TF'* time series are qualitatively similar to the deterministic *TF'* QDM-corrected inshore-extrapolated time series to which they are calibrated. This is illustrated in Figure 12a, showing an example of a statistically-generated *TF'* for the ARMA model corresponding to IPSL-CM6A in the scenario ssp585 at Helheim glacier. Subsequently, *TF* is reconstructed (Fig. 12b) from the ARMA-generated *TF'*. Both time series are compared to the corresponding *TF* and *TF'* from one of the original IPSL-CM6A QDM-corrected and inshore-extrapolated ensemble members, which we refer to as the deterministic time series. These comparisons provide a visual intuition for the ability of ARMA processes to reproduce the residual variability, and our general fitting procedure to reproduce the characteristics of *TF* time series. We also find a good agreement between the deterministic and statistically-generated *TF'* in terms of standard deviation and 1-year autocorrelation (Figure 12a). For illustrative purposes, we show statistically-generated *TF* time series for a single randomly-selected member of IPSL-CM6A under ssp585 at six of the largest Greenland outlet glaciers from different regions in Figure 13. Again, these examples demonstrate that QDM-corrected inshore-extrapolated *TF* time series with either deterministic or stochastically-generated residuals are qualitatively similar. In contrast, *TF* of the raw AOGCM from the glaciers' nearest-neighbor locations show some systematic differences in terms of mean and variability, and can fail to reproduce *TF* values of nearby CTDs (Fig. 13).

Quantitatively, we find good agreement in the timescale and amplitude of variability between the stochastically-generated *TF'* and their deterministic counterparts. In Figure 14a , we compare the 1-year autocorrelation of all the 6780 deterministic *TF'* time series of our dataset to a stochastic *TF'* generated for the corresponding AOGCM, emission scenario, and glacier. We find that 59% of the variance in the 1-year autocorrelation is explained by the ARMA models, with an RMSE<0.1 and a bias magnitude<0.02. We perform the same comparison for the standard deviation of the time series in Figure 14b, showing that 89% of the variance is explained, the RMSE is <0.1 K and the bias is negligibly small.

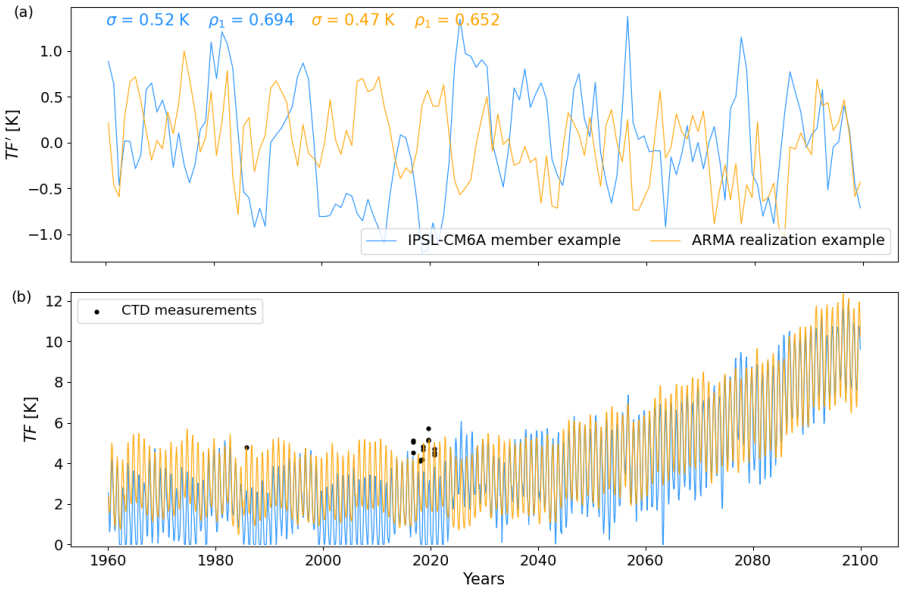

Figure 12. QDM-corrected and inshore-extrapolated time series of (a) *TF'* and (b) *TF* using a a deterministic ensemble member of IPSL-CM6A under ssp585 at Helheim glacier (blue curves), and a statistically-generated time series using the optimally-calibrated ARMA model (orange curves). The blue *TF'* curve shows the residual variability as given by the deterministic QDM-corrected and inshore extrapolated AOGCM member, and the orange *TF'* curve shows the residual variability simulated as an ARMA process. The blue and orange *TF* curves in panel (b) add the residual variability shown in panel (a) to the mean, trend, and seasonality components. The standard deviation (σ) and 1-year autocorrelation (ρ_1) of the *TF'* time series are shown in (a). Black dots in (b) show *TF* measurements from CTD data located at 100 km or less to the glacier front. Time series are only shown over the period 1960-2100 and with a single randomly-selected member for the sake of figure clarity. See Figure 1b for the location of Helheim glacier.

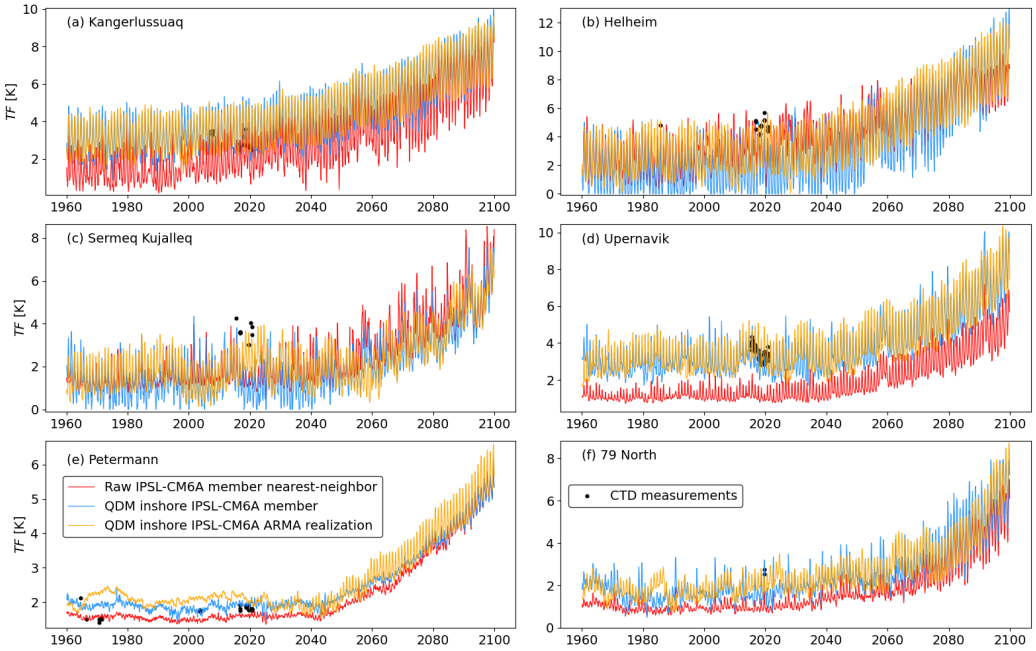

Figure 13. Raw nearest-neighbor AOGCM (red), deterministic QDM-corrected and inshore-extrapolated (light-blue), and ARMA realization of the QDM-corrected and inshore-extrapolated (orange) time series of *TF* at six large Greenland glaciers in different oceanic sectors. Time series are for the IPSL-CM6A AOGCM under the ssp585 emission scenario. The orange curves use a statistically-generated time series of *TF'* using the optimally-calibrated ARMA models for each glacier. The blue curves use the residual variability as given by the deterministic QDM-corrected and inshore-extrapolated AOGCM member as *TF'*. Both the orange and blue curves add their respective *TF'* to the mean, trend, and seasonality components, resulting in their respective *TF* time series. Black dots show *TF* measurements from CTD data located at 100 km or less to the glacier front. Time series are only shown over the period 1960-2100 and with a single randomly-selected member for the sake of figure clarity. See Figure 1b for locations of the glaciers.

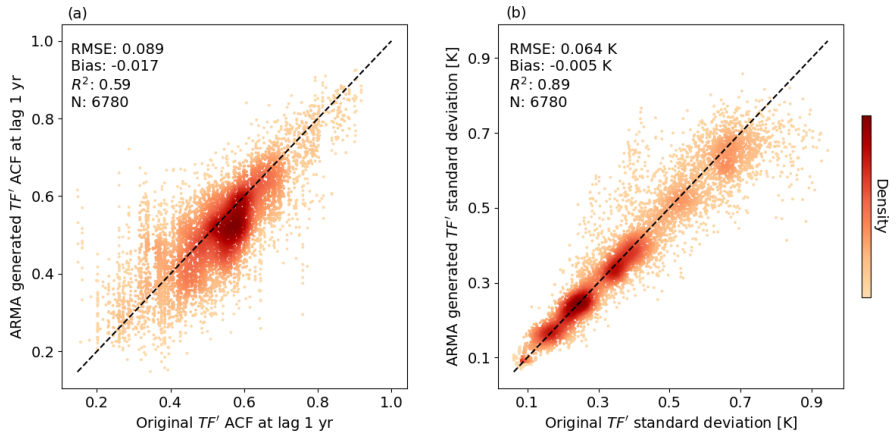

Figure 14. Comparison of (a) autocorrelation function at lag 1-year, and (b) standard deviation. Each pairwise comparison consists of a deterministic *TF'* time series and a *TF'* time series generated from the corresponding calibrated ARMA model. RMSE, R^2, and N denote root mean squared error, coefficient of determination, and total number of pairwise comparisons, respectively.

Finally, the correlation matrices ($\hat{\overline{C}}$, see Eq. (31)) show that the graphical lasso method (see Sect. 2.3, Friedman et al., 2008) effectively reduces the empirical correlation structure and prevents overfitting caused by the small sample size. This is illustrated by comparing the empirical and sparse correlation matrices in Fig. 15 for the case of the IPSL-CM6A AOGCM under the ssp585 emission scenario. Furthermore, the correlation matrices provide further evidence that our method effectively isolates the residual component of *TF* variability. As shown in the four sparse correlation matrices (Fig. 16), the correlation structure is similar for a single AOGCM under different emission scenarios, but differs between AOGCMs. The independence of $\hat{\overline{C}}$ on the emission scenario can only be obtained if the non-stationary deterministic changes associated to the emission scenario are removed. This is achieved here by removing polynomials with glacier-specific parameters in Eqs. (26) and (27), which adequately capture the non-stationary patterns. In contrast, the correlation patterns differ between AOGCMs, as expected due to their different internal dynamics. For example, Fig. 16 shows that correlation in IPSL-CM6A extends over a smaller neighborhood than in MIROC-ES2L.

4 Discussion

The method developed in this study is complementary to, and extends previous work for parameterizing ocean thermal forcing in Greenland ice sheet model simulations (Rignot et al., 2016; Slater et al., 2019, 2020). The QDM correction adjusts the distribution of AOGCM *TF* with respect to observational datasets, the extrapolation method corrects for too-coarse model resolution, and stochastic realizations of residual variability in *TF* sample internal climatic variability in model projections. These steps are independent of each other, and can be performed individually. While the melt parameterization of Xu et al. (2012) and Rignot et al. (2016) assumes that *TF* is given at the fjord mouth, the retreat parameterization of Slater et al. (2019) depends on sector-averaged values. The latter approach could therefore skip the extrapolation step, or alternatively be re-calibrated while accounting for extrapolation. Furthermore, each step can be applied with any choice of model and reanalysis products, making the method highly flexible. The accuracy of the *TF* time series generated depends directly on the quality of

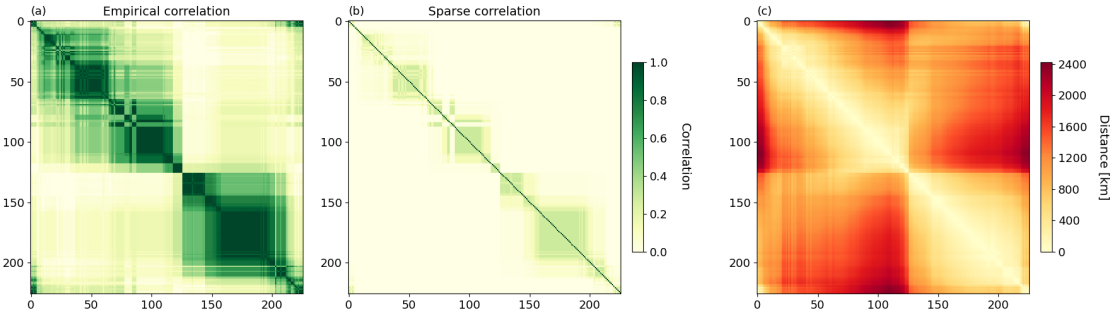

Figure 15. Across-members average of the (a) empirical and (b) sparse correlation matrices for the residual variability between the 226 outlet glacier locations of the IPSL-CM6A AOGCM under the emission scenario ssp585. Numbers on the x- and y-axis denote glacier number, with 0 corresponding to the Northern-most glacier, and glacier numbers increase in the clock-wise direction. Distances between glaciers are shown in (c). The separation around number 125 corresponds to the transition from East to West Greenland. See Fig. 1a for the glacier locations.

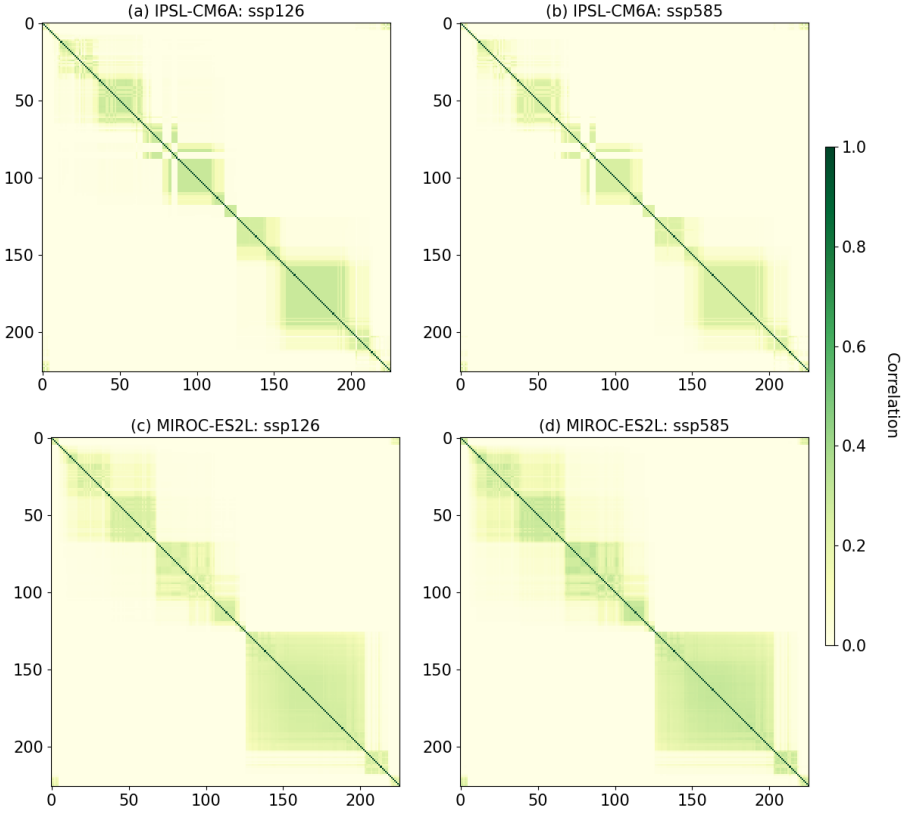

Figure 16. Across-members average of the sparse correlation matrices for the residual variability ($\hat{\overline{C}}$) between the 226 outlet glacier locations of the results from (a) IPSL-CM6A under scenario ssp126, (b) IPSL-CM6A under scenario ssp585, (c) MIROC-ES2L under scenario ssp126, and (d) MIROC-ES2L under scenario ssp585. Numbers on the x- and y-axis denote glacier number, with 0 corresponding to the Northern-most glacier, and glacier numbers increase in the clock-wise direction. The separation around number 125 corresponds to the transition from East to West Greenland. See Fig. 15c for distances between glaciers and Fig. 1a for the glacier locations.

the products used. Disagreement between different reanalysis products on ocean heat content can be particularly high around Greenland (Palmer et al., 2017), and one could explore the sensitivity of the final *TF* generated to the reanalysis product used. High-resolution ocean model outputs are scarce, but are likely to increase in coming years such that the extrapolation method could use multi-model averages of offshore-inshore relationships. The low computational expense of the ARMA statistical models also facilitates the generation of large numbers of *TF* time series, efficiently sampling the irreducible climatic uncertainty associated with natural variability. This is important, as there is an increased awareness of ice sheet sensitivity to variability in ocean conditions (e.g., Hoffman et al., 2019; Robel et al., 2019; Christian et al., 2020), and as numerical tools become available to investigate this sensitivity (Verjans et al., 2022).

The QDM correction method is well-established in the climate- and hydrology-modeling communities (e.g., Gudmundsson et al., 2012; Themeßl et al., 2012; Cannon et al., 2015). It corrects the distributional properties of AOGCM output, and permits the projection of future changes without invoking a stationarity assumption. Our independent evaluation of the method against CTD data shows that it increases explained variance from the AOGCM by $\sim 25\%$ and decreases the RMSE by ~ 0.3 K ($\sim 25\%$). However, one potential limitation is that it linearly adds AOGCM-projected changes in the future period (Eq. (5)), regardless of the model biases in the observational period. This assumption therefore does not correct a potential bias in the long-term climate sensitivity of the AOGCM. This though, is an intrinsic issue of all projections, and argues for the use of multiple different AOGCMs to force ice sheet models.

The extrapolation method (see Sect. 2.2) offers large flexibility in the formulations of both the regression parameters in the linear relations (Eqs. (9-15)), as well as of the criteria used in the predictor selection (Eqs. (18-25)). For example, additional refinements can be applied to the predictor selection by including connectivity between gridpoints based on passive tracer advection (Oliver & Holbrook, 2014), or by using multiple offshore predictors in a multiple regression method. We perform the calibration of the ARMA models on the residual variability in *TF* of the AOGCM time series, here spanning 250 years. Alternatively, residuals could be calibrated from pre-industrial control runs of AOGCMs, which can be >1000 years long. However, this approach faces the limitation that pre-industrial control AOGCM simulations cannot be QDM-corrected prior to the ARMA fitting, as reanalysis products are based on data from the industrial era.

Our procedure does not capture two-way interactions between the ocean and the ice sheet. Input of freshwater due to outlet glacier melt and subglacial discharge are important components of fjord circulation. This could not only modify *TF* properties in shallow waters, but also alter the relationship between *TF* properties on the continental shelf and glacier frontal melt (Slater et al., 2018; Wagner et al., 2019; Jackson et al., 2020). Capturing heat transfer through fjords and up to the ice-ocean boundary layer will require detailed, high-resolution fjord models in the future (e.g., Cowton et al., 2015; Zhao et al., 2021). The work presented here is relevant to such future model developments, as it can provide boundary forcing to fjord models by applying our method to both ocean temperature and salinity, and thus density and stratification, in a similar manner. Nevertheless, in the current context of ocean modeling, our approach remains consistent with the assumption of the Xu et al. (2012) and Rignot et al. (2016) parameterization that the *TF* variable is prescribed from the fjord mouth. The interplay of atmospheric and oceanic forcing may also be an important factor in governing outlet glacier frontal melt, because of the dependence of melt rates to both ocean temperatures and to suglacial discharge, which is sourced from surface melt (Slater & Straneo, 2022). By exploiting the statistical nature of our method, covariance of *TF* with surface melt can be estimated using output from regional climate models and enforced in the *TF* generation. Thus, using ocean and atmospheric model output, correlation can capture inter-

actions of *TF* with variables influencing marine-terminating glacier melt, as well as with variables influencing inland ice sheet mass balance.

We provide the entire workflow of QDM-correction, inshore extrapolation, and statistical generation of residuals as open-source code (see *Open Research*). These resources allow potential users to compute *TF* time series as detailed in this study. Furthermore, we provide ensembles of *TF* time series at the 226 outlet glaciers for the four combinations of AOGCMs (IPSL-CM6A and MIROC-ES2L) and CMIP6 emission scenarios (ssp126 and ssp585) used in this study as an open-access dataset (see *Data availability statement*). These ensembles of 1000 members each, spanning 1850-2100, can be used by the glaciology modeling community to force ice sheet model simulations at the scale of the Greenland ice sheet, or sub-regions of the ice sheet.

5 Conclusions

We propose a statistical method to estimate ocean thermal forcing for the Greenland ice sheet on a range of different timescales. Starting from AOGCM output, and based on output from ocean reanalysis products and high-resolution models, this method bias-corrects, adjusts distributional properties, extrapolates, captures spatial correlation, and samples variability in ocean thermal forcing. The correction of bias and variability amplitude via a quantile delta mapping approach reproduces the distributional properties of reanalysis products, while preserving the climate sensitivity of the AOGCM in the form of trends and future changes. The extrapolation method derives simple and independent linear relationships between offshore and inshore thermal forcing at different timescales from a high-resolution ocean model. The relationships are subsequently applied to AOGCM output to estimate ocean conditions at fjord mouth locations. Finally, we use autoregressive moving-average models to represent the residual variability observed in thermal forcing. These statistical models not only reproduce temporal characteristics in modeled residual variability, but also capture spatial covariance in ocean thermal forcing.

The workflow developed here offers a complementary approach to glacier melt parameterizations that are applied in ice sheet model simulations. Generation of thermal forcing is computationally straightforward owing to the purely statistical nature of each step in the method, and ensembles of time series accompany this study as a data product. Given the current state of climate and ocean modeling, ice sheet model predictions will likely continue to face a dearth of long-term high-resolution ocean model output availability. This highlights the need to use statistical techniques in order to bridge the gap between existing climate model output and ice sheet model requirements for boundary forcing. Our method is a first step in this direction. It offers the advantages of relative ease of computation, of addressing several limitations of AOGCMs at once, and of being applicable to any combination of AOGCM, reanalysis product, and high-resolution ocean model.

6 Open Research

All code (python scripts) to reproduce the method described in this study are available as a Zenodo repository (Verjans, 2023):
https://doi.org/10.5281/zenodo.7808874
The repository includes all intermediary and final output files for member r1 of MIROC-ES2L under scenario ssp585 as an example. The repository also includes samples of 1000 *TF* 1850-2100 time series generated following the method presented at the 226 marine glacier fronts for the four combinations of AOGCMs and CMIP6 emission scenarios used in this study. The raw climate model output from the CMIP6 experiments can be downloaded from:
https://esgf-node.llnl.gov/search/cmip6/

The raw output of the Hadley Centre EN4.2.1 monthly objective analyses can be downloaded from:
https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html
Output from the high-resolution ECCO Arctic forward run can be downloaded from:
https://ecco-group.org/

Acknowledgments
This work was funded by a grant from the Heising-Simons Foundation. HS was also funded by the NSF Navigating the New Arctic Program. Computing resources were provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta. We thank three anonymous reviewers for providing comments that helped to improve the quality of this study. VV thanks Lizz Ultee for guidance about the graphical lasso method, Hong Zhang for helping with ECCO output processing, and John Christian for insights about climate variability.

References

Balmaseda, M. A., Alves, O., Arribas, A., Awaji, T., Behringer, D., Ferry, N., ... Stammer, D. (2009). Ocean initialization for seasonal forecasts. *Oceanography*, *22*, 154-159.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... Vuichard, N. (2020). Presentation and evaluation of the ipsl-cm6a-lr climate model. *Journal of Advances in Modeling Earth Systems*, *12*.

Box, G., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day.

Buch, E. (2002). Present oceanographic conditions in greenland waters. *Scientific Report 02-02*.

Camus, P., Menéndez, M., Méndez, F. J., Izaguirre, C., Espejo, A., Cánovas, V., ... Medina, R. (2014). A weather-type statistical downscaling framework for ocean wave climate. *Journal of Geophysical Research*, *119*, 7389-7405.

Cannon, A., Sobie, S. R., & Murdock, T. (2015). Bias correction of gcm precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, *28*, 6938-6959.

Castruccio, S., & Stein, M. L. (2013). Global space–time models for climate ensembles. *The Annals of Applied Statistics*, *7*, 1593-1611.

Cheng, L., Foster, G., Hausfather, Z., Trenberth, K. E., & Abraham, J. P. (2022). Improved quantification of the rate of ocean warming. *Journal of Climate*.

Christian, J. E., Robel, A. A., Proistosescu, C., Roe, G. H., Koutnik, M. R., & Christianson, K. (2020). The contrasting response of outlet glaciers to interior and ocean forcing. *The Cryosphere*.

Cottier, F. R., Nilsen, F., Skogseth, R., Tverberg, V., Skardhamar, J., & Svendsen, H. (2010). Arctic fjords: a review of the oceanographic environment and dominant physical processes..

Cowton, T. R., Slater, D. A., Sole, A. J., Goldberg, D. N., & Nienow, P. (2015). Modeling the impact of glacial runoff on fjord circulation and submarine melt rate using a new subgrid-scale parameterization for glacial plumes. *Journal of Geophysical Research*, *120*, 796-812.

Cowton, T. R., Sole, A. J., Nienow, P., Slater, D. A., & Christoffersen, P. (2018). Linear response of east greenland's tidewater glaciers to ocean/atmosphere warming. *Proceedings of the National Academy of Sciences*, *115*, 7907 - 7912.

Cui, B., Toth, Z., Zhu, Y., & Hou, D. (2012). Bias correction for global ensemble forecast. *Weather and Forecasting*, *27*, 396-410.

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *49*(4), 1057–1072. Retrieved 2022-

10-28, from http://www.jstor.org/stable/1912517

Dickson, R. R., Osborn, T. J., Hurrell, J. W., Meincke, J., Blindheim, J., Adlandsvik, B., ... Maslowski, W. (2000). The arctic ocean response to the north atlantic oscillation. *Journal of Climate*, *13*(15), 2671 - 2696. Retrieved from https://journals.ametsoc.org/view/journals/clim/13/15/1520-0442_2000_013_2671_taortt_2.0.co_2.xml doi: 10.1175/1520-0442(2000)013⟨2671:TAORTT⟩2.0.CO;2

Fagundes, M., Litvin, S. Y., Micheli, F., Leo, G. A. D., Boch, C., Barry, J. P., ... Woodson, C. B. (2020). Downscaling global ocean climate models improves estimates of exposure regimes in coastal environments. *Scientific Reports*, *10*.

Felikson, D., Bartholomaus, T. C., Catania, G. A., Korsgaard, N. J., Kjær, K. H., Morlighem, M., ... Nash, J. D. (2017). Inland thinning on the greenland ice sheet controlled by outlet glacier geometry. *Nature Geoscience*, *10*, 366-369.

Fettweis, X., Hofer, S., Krebs-Kanzow, U., Amory, C., Aoki, T., Berends, C. J., ... Zolles, T. (2020). Grsmbmip: intercomparison of the modelled 1980–2012 surface mass balance over the greenland ice sheet. *The Cryosphere*.

Friedman, J. H., Hastie, T. J., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9 3*, 432-41.

Gillard, L. C., Hu, X., Myers, P. G., & Bamber, J. L. (2016). Meltwater pathways from marine terminating glaciers of the greenland ice sheet. *Geophysical Research Letters*, *43*(20), 10,873-10,882. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL070969 doi: https://doi.org/10.1002/2016GL070969

Goelzer, H., Nowicki, S., Payne, A., Larour, E., Seroussi, H., Lipscomb, W. H., ... van den Broeke, M. (2020). The future sea-level contribution of the greenland ice sheet: a multi-model ensemble study of ismip6. *The Cryosphere*, *14*(9), 3071–3096. Retrieved from https://tc.copernicus.org/articles/14/3071/2020/ doi: 10.5194/tc-14-3071-2020

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). En4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*(12), 6704-6716. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JC009067 doi: https://doi.org/10.1002/2013JC009067

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., & Engen-Skaugen, T. (2012). Technical note: Downscaling rcm precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrology and Earth System Sciences*, *16*, 3383-3390.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., ... Kawamiya, M. (2020). Development of the miroc-es2l earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5), 2197–2244. Retrieved from https://gmd.copernicus.org/articles/13/2197/2020/ doi: 10.5194/gmd-13-2197-2020

Hasselmann, K. (1976). Stochastic climate models part i. theory. *Tellus A*, *28*, 473-485.

Hausfather, Z., Cowtan, K. D., Clarke, D. C., Jacobs, P., Richardson, M. I., & Rohde, R. A. (2017). Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Science Advances*, *3*.

Hoffman, M. J., Asay-Davis, X. S., Price, S. F. D., Fyke, J. G., & Perego, M. (2019). Effect of subshelf melt variability on sea level rise contribution from thwaites glacier, antarctica. *Journal of Geophysical Research: Earth Surface*, *124*, 2798 - 2822.

Holland, D. M., Thomas, R. H., de Young, B., Ribergaard, M. H., & Lyberth, B. (2008). Acceleration of jakobshavn isbræ triggered by warm subsurface ocean waters. *Nature Geoscience*, *1*, 659-664.

Hu, W., & Castruccio, S. (2021). Approximating the internal variability of bias-corrected global temperature projections with spatial stochastic generators. *Journal of Climate*.

IMBIE. (2020). Mass balance of the greenland ice sheet from 1992 to 2018. *Nature*, *579*, 233-239.

Jackson, R. H., Lentz, S. J., & Straneo, F. (2018). The dynamics of shelf forcing in greenlandic fjords. *Journal of Physical Oceanography*.

Jackson, R. H., Nash, J. D., Kienholz, C., Sutherland, D. A., Amundson, J. M., Motyka, R. J., ... Pettit, E. C. (2020). Meltwater intrusions reveal mechanisms for rapid submarine melt at a tidewater glacier. *Geophysical Research Letters*, *47*.

Jackson, R. H., Straneo, F., & Sutherland, D. A. (2014). Externally forced fluctuations in ocean temperature at greenland glaciers in non-summer months. *Nature Geoscience*, *7*, 503-508.

Jakobsson, M., Macnab, R. F., Mayer, L. A., Anderson, R. M., Edwards, M. H., Hatzky, J., ... Johnson, P. (2008). An improved bathymetric portrayal of the arctic ocean: Implications for ocean modeling and geological, geophysical and oceanographic analyses. *Geophysical Research Letters*, *35*.

Kay, J. E., Deser, C., Phillips, A. S., Mai, A., Hannay, C., Strand, G., ... Vertenstein, M. (2015). The community earth system model (cesm) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*, 1333-1349.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Dobrynin, M., ... Marotzke, J. (2019). The max planck institute grand ensemble: Enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems*, *11*, 2050 - 2069.

Marks, K. M., & Smith, W. H. F. (2006). An evaluation of publicly available global bathymetry grids. *Marine Geophysical Researches*, *27*, 19-34.

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., ... Zinglersen, K. B. (2017). Bedmachine v3: Complete bed topography and ocean bathymetry mapping of greenland from multibeam echo sounding combined with mass conservation. *Geophysical Research Letters*, *44*, 11051 - 11061.

Mouginot, J., Rignot, E., Bjørk, A. A., van den Broeke, M. R., Millan, R., Morlighem, M., ... Wood, M. (2019). Forty-six years of greenland ice sheet mass balance from 1972 to 2018. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 9239 - 9244.

Mudelsee, M. (2010). Climate time series analysis: Classical statistical and bootstrap methods. *Climate Time Series Analysis*.

Nguyen, A. T., Kwok, R., & Menemenlis, D. (2012). Source and pathway of the western arctic upper halocline in a data-constrained coupled ocean and sea ice model. *Journal of Physical Oceanography*, *42*, 802-823.

Nick, F. M., Vieli, A., Howat, I. M., & Joughin, I. R. (2009). Large-scale changes in greenland outlet glacier dynamics triggered at the terminus. *Nature Geoscience*, *2*, 110-114.

Noël, B. P. Y., van de Berg, W. J., Machguth, H., Lhermitte, S., Howat, I. M., Fettweis, X., & van den Broeke, M. (2016). A daily, 1 km resolution data set of downscaled greenland ice sheet surface mass balance (1958–2015). *The Cryosphere*, *10*, 2361-2377.

Oliver, E. C. J., & Holbrook, N. J. (2014). A statistical method for improving continental shelf and nearshore marine climate predictions. *Journal of Atmospheric and Oceanic Technology*, *31*, 216-232.

Palmer, M. D., Roberts, C. D., Balmaseda, M. A., Chang, Y.-S., Chepurin, G. A., Ferry, N., ... Xue, Y. (2017). Ocean heat content variability and change in an ensemble of ocean reanalyses. *Climate Dynamics*, *49*, 909-930.

Porter, D. F., Tinto, K. J., Boghosian, A. L., Csathó, B. M., Bell, R. E., & Cochran, J. R. (2018). Identifying spatial variability in greenland's outlet glacier response to ocean heat. *Frontiers in Earth Science*.

Rignot, E., Fenty, I. G., Menemenlis, D., & Xu, Y. (2012). Spreading of warm ocean waters around greenland as a possible cause for glacier acceleration. *Annals of Glaciology*, *53*, 257 - 266.

Rignot, E., Xu, Y., Menemenlis, D., Mouginot, J., Scheuchl, B., Li, X., ... de Fleurian, B. (2016). Modeling of ocean-induced ice melt rates of five west greenland glaciers over the past two decades. *Geophysical Research Letters*, *43*, 6374 - 6382.

Robel, A. A., Seroussi, H., & Roe, G. H. (2019). Marine ice sheet instability amplifies and skews uncertainty in projections of future sea-level rise. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 14887 - 14892.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. Retrieved 2022-10-28, from `http://www.jstor.org/stable/2958889`

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591-611.

Slater, D. A., Felikson, D., Straneo, F., Goelzer, H., Little, C. M., Morlighem, M., ... Nowicki, S. M. J. (2020). Twenty-first century ocean forcing of the greenland ice sheet for modelling of sea level contribution. *The Cryosphere*.

Slater, D. A., & Straneo, F. (2022). Submarine melting of glaciers in greenland amplified by atmospheric warming. *Nature Geoscience*, *15*, 794 - 799.

Slater, D. A., Straneo, F., Das, S. B., Richards, C. G., Wagner, T., & Nienow, P. (2018). Localized plumes drive front-wide ocean melting of a greenlandic tidewater glacier. *Geophysical Research Letters*, *45*, 12,350 - 12,358.

Slater, D. A., Straneo, F., Felikson, D., Little, C. M., Goelzer, H., Fettweis, X., & Holte, J. (2019). Estimating greenland tidewater glacier retreat driven by submarine melting. *The Cryosphere*.

Storch, H. v., & Zwiers, F. W. (1999). *Statistical analysis in climate research*. Cambridge University Press. doi: 10.1017/CBO9780511612336

Straneo, F., & Cenedese, C. (2015). The dynamics of greenland's glacial fjords and their role in climate. *Annual review of marine science*, *7*, 89-112.

Straneo, F., & Heimbach, P. (2013). North atlantic warming and the retreat of greenland's outlet glaciers. *Nature*, *504*, 36-43.

Straneo, F., Slater, D. A., Bouchard, C., Cape, M. R., Carey, M., Ciannelli, L., ... Vernet, M. (2022). An interdisciplinary perspective on greenland's changing coastal margins. *Oceanography*.

Straneo, F., Sutherland, D. A., Holland, D. M., Gladish, C. V., Hamilton, G., Johnson, H. L., ... Koppes, M. N. (2012). Characteristics of ocean waters reaching greenland's glaciers. *Annals of Glaciology*, *53*, 202 - 210.

Strass, V. H., Fahrbach, E., Schauer, U., & Sellmann, L. (1993). Formation of denmark strait overflow water by mixing in the east greenland current. *Journal of Geophysical Research*, *98*, 6907-6919.

Sutherland, D. A., Roth, G., Hamilton, G., Mernild, S. H., Stearns, L. A., & Straneo, F. (2014). Quantifying flow regimes in a greenland glacial fjord using iceberg drifters. *Geophysical Research Letters*, *41*, 8411 - 8420.

Sutherland, D. A., Straneo, F., Stenson, G. B., Davidson, F. J. M., Hammill, M. O., & Rosing-Asvid, A. (2013). Atlantic water variability on the se greenland continental shelf and its relationship to sst and bathymetry. *Journal of Geophysical Research*, *118*, 847-855.

Themeßl, M. J., Gobiet, A., & Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change*, *112*, 449-468.

Tsai, C.-Y., Forest, C. E., & Pollard, D. D. (2017). Assessing the contribution of internal climate variability to anthropogenic changes in ice sheet volume. *Geophysical Research Letters*, *44*, 6261-6268.

Verjans, V. (2023). Code for Bias correction and statistical modeling of variable oceanic forcing of greenland outlet glaciers: April 07, 2023 Release (Version 1.0) [Software]. Zenodo.
doi: https://doi.org/10.5281/zenodo.7808874

Verjans, V., Robel, A. A., Seroussi, H., Ultee, L., & Thompson, A. F. (2022). The stochastic ice-sheet and sea-level system model v1.0 (stissm v1.0). *Geoscientific Model Development*, *15*(22), 8269–8293. doi: 10.5194/gmd-15-8269-2022

Wagner, T. J. W., Straneo, F., Richards, C. G., Slater, D. A., Stevens, L. A., Das, S. B., & Singh, H. (2019). Large spatial variations in the flux balance along the front of a greenland tidewater glacier. *The Cryosphere*.

Walsh, K. M., Howat, I. M., Ahn, Y., & Enderlin, E. M. (2012). Changes in the marine-terminating glaciers of central east greenland, 2000–2010. *The Cryosphere*, *6*, 211-220.

Wilks, D. S. (2011). *Statistical methods in atmospheric sciences*. Academic Press.

Wood, M., Rignot, E., Fenty, I. G., An, L., Bjørk, A. A., van den Broeke, M. R., ... Zhang, H. (2021). Ocean forcing drives glacier retreat in greenland. *Science Advances*, *7*.

Xu, Y., Rignot, E., Menemenlis, D., & Koppes, M. N. (2012). Numerical experiments on subaqueous melting of greenland tidewater glaciers in response to ocean warming and enhanced subglacial discharge. *Annals of Glaciology*, *53*, 229 - 234.

Zhao, K., Stewart, A. L., & McWilliams, J. C. (2021). Linking overturning, recirculation, and melt in glacial fjords. *Geophysical Research Letters*, *49*.