



Statistical Generation of Ocean Forcing With Spatiotemporal Variability for Ice Sheet Models

Shivaprakash Muruganandham  and Alexander A. Robel , Georgia Institute of Technology, Atlanta, GA, 30332, USA

Matthew J. Hoffman  and Stephen F. Price, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA

Melting of ice at the base of floating ice shelves that fringe the Antarctic ice sheet has been identified as a significant source of uncertainty in sea level rise projections. Part of this uncertainty derives from chaotic internal variability of the coupled ocean-atmosphere system. For numerical ice sheet model projections, this uncertainty has not previously been quantified because of the prohibitive computational expense of running large climate model ensembles. Here, we develop and demonstrate a technique that generates independent realizations of internal climate variability from a single climate model simulation. Building on previous developments in model emulation, this technique uses empirical orthogonal function decomposition and Fourier-phase randomization to generate statistically consistent realizations of spatiotemporal variability fields for the target climate variable. The method facilitates efficient sampling of a wide range of climate trajectories, which can also be incorporated within ice sheet or other physical models to represent feedback processes.

The ocean melts the underside of ice shelves, the floating portions of large continental ice sheets. Such “basal melt” has been identified as a source of significant uncertainty in global sea level rise projections as it has the potential to drive rapid ice sheet retreat.¹ A part of this uncertainty derives from the unpredictable internal variability in ocean temperature and salinity around ice shelves. Recent satellite measurements have revealed large temporal variability in the basal melt from individual Antarctic ice shelves, each characterized by distinct regional and temporal signatures and driven by complex oceanic and atmospheric interactions.² This is understood to manifest commonly in the form of distinct “modes” of basal melt,³ driven by changing inflow of different ocean water masses on time scales spanning days to decades. The internal variability in the climate system that drives such changes is not yet well characterized in ice

sheet model simulations, adding to the uncertainty in projections of Antarctica’s future evolution.

“Large ensembles” of global climate model simulations have become invaluable tools to understand the role of internal climate variability in generating uncertainty in future projections. Similarly, large ensembles of ice sheet model simulations can be used to quantify ice sheet projection uncertainty, with each member influenced by an alternate “realization” of climate variability (or climate “forcing”). Previous attempts⁴ at quantifying ice sheet projection uncertainty due to internal climate variability have been hindered by the coarse grid resolution of existing large ensembles from climate models, a lack of climate models that can realistically simulate melting beneath ice shelves, and a lack of climate model simulation outputs beyond 2100 (when ice sheets are more likely to retreat rapidly). As running even a single climate model simulation that does not have these deficiencies can require tens of millions of core hours on high-performance machines, running a large number of such simulations is computationally prohibitive with currently available resources.

1521-9615 © 2023 IEEE
Digital Object Identifier 10.1109/MCSE.2023.3300908
Date of publication 1 August 2023; date of current version 25 October 2023.

Generators are computational tools that have long been used in the numerical weather prediction community⁵ to produce random realizations that statistically resemble observations or output from complex physical models at significantly reduced computational cost. Recently, there have been several proposed approaches that attempt to retain the spatiotemporal characteristics of internal variability in climate model simulations. Some approaches resample input model fields⁶ while others have relied on autoregressive models that account for temporal autocorrelations in the data.⁷ Many of these methods have been developed in the global climate modeling community, usually applied to coarse-resolution temperature and precipitation output fields. However, current approaches are not without challenges: some assume a Gaussian distribution for the model simulation data,⁷ while others,⁶ built to be used with global data, do not account for the nonrandom components of climate variability (e.g., the seasonal cycle) and spatial dependencies on evolving ice sheet geometry,⁸ which are particularly important for climate variability in polar regions.

In this article, we develop a general statistical technique that generates random realizations of climate variability based on output from a single long climate model simulation. We demonstrate how this method can be used efficiently to generate many realizations of variable basal melt projections for the Antarctic ice sheet that preserve the properties of the input model simulation.

METHOD

Here we describe our statistical method, which generates climate forcing for ice sheet models, including spatiotemporal variability. This technique ensures that generated variability is spatially and temporally consistent with the training data while also facilitating the sampling of a wide range of realizations in a computationally efficient manner. The method can, in principle, be applied to any input spatiotemporal climate dataset with certain properties. First, an empirical orthogonal function (EOF) decomposition is used to separate modes of spatiotemporal variability. The result is EOFs that are 2-D fields that capture the spatial fingerprint of each mode, and time series of projection coefficients [also known as *principal components (PCs)*] that capture the temporal variability of each mode. These PCs are then phase randomized, which retains the time scales of temporal variability while randomizing the particular phasing associated with the input dataset. Finally, a new realization of the climate forcing is generated by recomposing the projection coefficients

and EOF modes. The final product is a new statistically independent realization of the climate forcing field. A schematic of the method description is presented in Figure 1. The results are compared to the original data by quantifying the root-mean-square error between the power spectral density (PSD) of the generated data with respect to that of the input model simulation as well as the evaluation metrics outlined in Link et al.⁶ The specific metrics are not discussed in the “Results” section as they are tailored to the particular dataset being tested. However, adjustments to the method such as the use of normalization and a sparse principal component analysis (PCA), described later, are motivated in part by their ability to improve performance.

Data Preprocessing

The primary focus of this article is to generate realizations of the nonseasonal internal climate variability in a target dataset. Here, we start by describing the particular preprocessing needed when this method is used to generate forcing for ice sheet models. As such, we must preprocess by excluding other components of the spatial and temporal variability for which phase randomization is unnecessary because either 1) they are monotonic on the spatial and temporal scales of interest or 2) the phase of their periodicity can be predicted. At a given point in space, \mathbf{x}_0 , we can describe the target dataset as the sum of various components

$$F_{orig}(\mathbf{x}_0, t) = F_t(\mathbf{x}_0, t) + F_s(\mathbf{x}_0, t) + F_d(\mathbf{x}_0) + F_v(\mathbf{x}_0, t) \quad (1)$$

where $F_t(\mathbf{x}_0, t)$ is a monotonic function of time, $F_s(\mathbf{x}_0, t)$ is the seasonal cycle, $F_d(\mathbf{x}_0)$ is a monotonic function of the variable of interest to space or other variables (i.e., basal melt under ice sheets typically increases with the depth of the ice shelf beneath the ocean surface), and the residual $F_v(\mathbf{x}_0, t)$ is the internal variability component we are interested in. Thus, to focus on generating phase-randomized internal variability of the climate dataset, we start by removing the F_t , F_s , and F_d components.

Climate datasets or model output frequently include monotonic trends like F_t due to human-caused emissions, which have produced monotonic trends in many climate variables over the industrial era and into the future. Low-order regression (typically linear) on the target variable at each spatial location or over subsets of the target dataset are sufficient to identify any trends, and then remove them. The annual seasonality, F_s , is the 12-month periodicity that occurs in most of the climate variables. Depending on the problem at hand, there may be other periodic components relevant at this stage, such as diurnal or tidal cycles. As

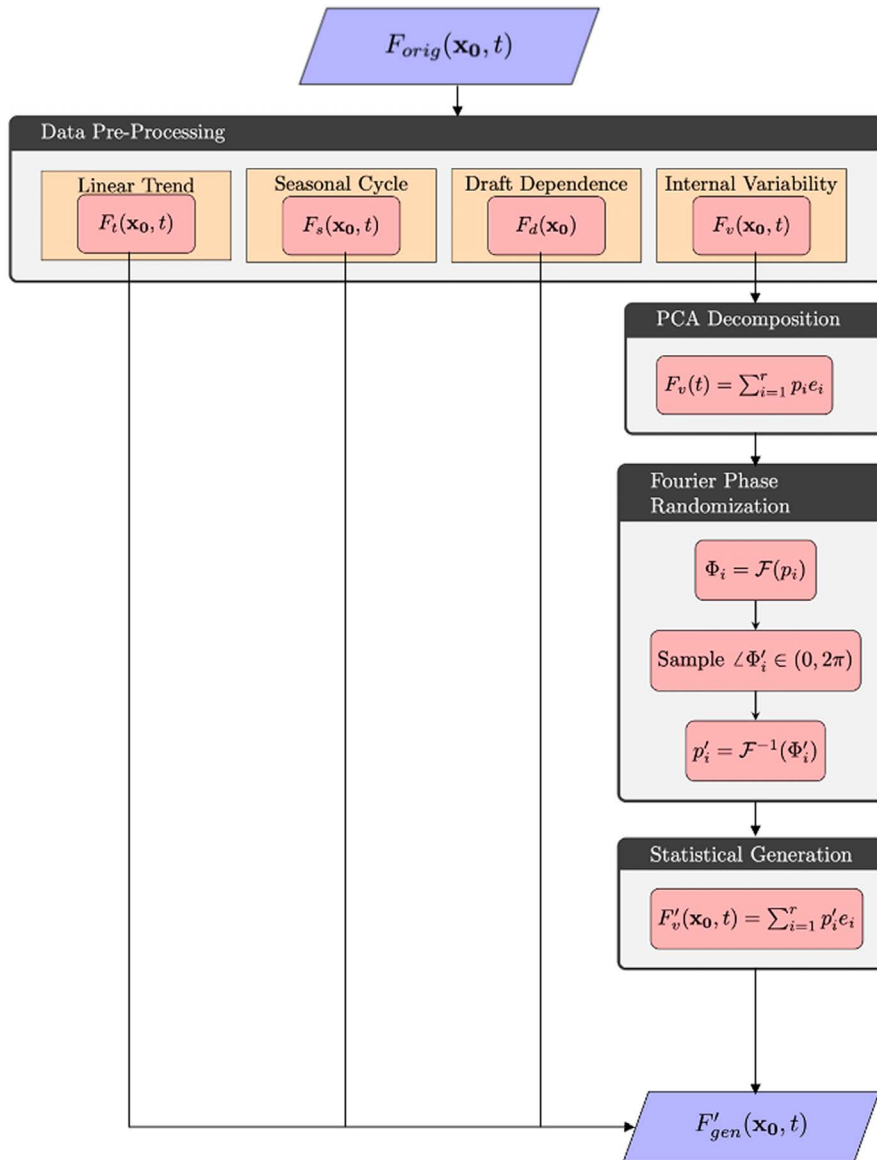


FIGURE 1. Generator workflow. PCA: principal component analysis.

the phase of seasonality is known, we can remove it at this stage and add it back the following generation. This step implicitly includes a centering of the temporal mean, which should be performed separately in the absence of such periodic components. The process of phase randomization is applied to only nonannual periodicities within the data. Many climate variables have purely spatial monotonic trends, F_d , due to other physical processes that are not variable in time. The variable we test our method on here, basal melting underneath ice shelves, increases with the draft (also known as *depth*) of ice shelves below the ocean surface due to pressure dependence of the melting

temperature of ice, which does not change over time. In a sense, this component is the spatial analog of F_t . We isolate this melt-draft dependence in the dataset using a separate least squares linear regression (although high-order regressions can be used as well) for each ice shelf. The result is a statistical estimate of the $F_d(t)$ component, which is removed to isolate the spatial components of internal variability. When the temporal trend, seasonality, and draft dependence are removed from the raw dataset, statistical tests for stationarity (e.g., augmented Dickey–Fuller) can be performed to ensure that the preprocessing steps yield an appropriate anomaly dataset for the method that follows.

Nonannual internal climate variability is denoted as F_v and treated as input data for the methods described hereafter. $F_{vn}(\mathbf{x}_0, t) = F_v(\mathbf{x}_0, t) / \sigma_{F_v}(\mathbf{x}_0)$ is further normalized with its standard deviation at each spatial grid point to diminish the influence of the magnitude of basal melt rate on the variance. F_{vn} is on some spatial grid (structured or unstructured) and sampled at regular time intervals. This data is organized in a $(t \times s)$ matrix

$$F_{vn} = [f_1, f_2, \dots, f_t]^T$$

which contains t observations of the variable f_t defined at s spatial grid points on an arbitrary number of spatial dimensions.

EOF Decomposition

To isolate the spatial patterns of variability in our target climate variable, we use EOF decomposition. This is a popular analysis technique commonly used in the earth sciences to find a set of orthogonal spatial patterns, where each EOF is associated with a time series of uncorrelated PCs. The decomposition returns an ordered set of EOF “modes” (i.e., patterns of variability) that describe uncorrelated spatial and temporal structures in the dataset, with the most dominant modes in climate datasets sometimes having identifiable physical origins (e.g., the El Niño–Southern Oscillation).

EOF decomposition identifies the eigenvalues of the input data covariance. At its core, EOF decomposition is accomplished through the generalized matrix factorization technique known as *singular value decomposition* (SVD), which allows one to rewrite any matrix as a product of three matrices

$$F_{(t \times s)} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$$

where $\mathbf{\Lambda}_{(t \times s)}$ is a diagonal matrix, and $\mathbf{U}_{(t \times t)}$ and $\mathbf{V}_{(s \times s)}$ are orthonormal square matrices. The elements along the leading diagonal of $\mathbf{\Lambda}$ are called *singular values*, denoted by $\lambda_{i,i}$ for $i \in (1, r)$, where r is the rank of the input matrix F , and $r \leq \min(t, s)$. The column vectors of \mathbf{U} and \mathbf{V} are the left and right singular vectors. If we were to perform an SVD on F and then form the covariance matrix C , we would have

$$C = (\mathbf{U} \mathbf{\Lambda} \mathbf{V}^T)^T (\mathbf{U} \mathbf{\Lambda} \mathbf{V}^T) = \mathbf{V} (\mathbf{\Lambda}^T \mathbf{\Lambda}) \mathbf{V}^T$$

where \mathbf{V} has the EOFs as columns, and the projection coefficients are contained in $\mathbf{P} = \mathbf{F} \mathbf{V}$. EOF decomposition of the input data matrix $F_{(t \times s)}$ can thus be rewritten via SVD as a series expansion of basis functions [EOFs, i.e., $e_i(s)$] with associated expansion functions [PC time series, i.e., $p_i(t)$]

$$F_{vn}(t, s) = \sum_{i=1}^r p_i(t) e_i(s)^T. \quad (2)$$

This decomposition can be understood as a series of spatial patterns ($e_i(s)$) and their associated temporal patterns ($p_i(t)$). The EOFs thus obtained are orthonormal, i.e.,

$$e_i^T e_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

and the PCs are uncorrelated temporal structures in the dataset, i.e.,

$$\text{cov}(p_i, p_j) = 0 \text{ if } i \neq j \quad (4)$$

where cov is the covariance operator.

Sparse PCA Decomposition

Although EOF analysis is a nomenclature specific to the climate and meteorology communities, the technique is quite common in statistics and machine learning and generally referred to in the literature as PCA. So far, we have described the canonical PCA method, however, this approach can be inconsistent in its estimation of PC weighting⁹ in the case of high-dimensional datasets, where the number of features (the spatial grid points) is far greater than the number of observations (the time-series length). This gives rise to spurious representations of spatial EOF structure, and in our method can show up as spurious high-frequency noise in the generator output. In such cases, sparse PCA algorithms can improve performance, and are preferred. Such methods modify canonical PCA by optimizing for sparsity in EOFs with a user-defined “penalization factor” to control the degree of sparsity.⁹ The result is EOFs with localized nonzero values, and zero values over most of the spatial domain. In forcing such sparsity, however, the method introduces correlations among PCs, for which this must be tested and accounted.

Fourier-Phase Randomization

The core of the generation method described here is Fourier-based phase randomization, which is commonly used in time-series analysis to generate surrogate data benchmarks for hypothesis testing with Gaussian inputs.¹⁰ Each of the PC time series from the EOF decomposition is passed through a Fourier transform, which yields a set of complex values, each corresponding to a frequency. The imaginary part of these complex values is the phase of each frequency component of the Fourier transform. To generate a single new realization of a PC, this imaginary phase component is sampled from a uniform distribution between (0) and (2π) . Then, the inverse Fourier transform is applied to the phase-randomized Fourier series to generate a new PC in the time domain.

Fourier-phase randomization takes advantage of the Wiener–Khinchin theorem, which equates the PSD of a wide-sense stationary random process with the Fourier transform of the time autocorrelation function (ACF) associated with that process. In this case, the projection coefficients $p_i(t)$ are treated as stationary random processes (which is ensured through preprocessing tasks and subsequent statistical tests for stationarity). The time correlation can then be captured via the Wiener–Khinchin theorem, which states

$$\mathcal{F}(C(p)) = |\mathcal{P}(f)|^2 \quad (5)$$

where \mathcal{F} is the Fourier transform operator, $C(p)$ is the time ACF of $p(t)$, and $\mathcal{P}(f) = \mathcal{F}(p(t))$, i.e., the Fourier transform of the PCs. As only the magnitudes of $\mathcal{P}(f)$ are used in the equation, randomizing the phases of $\mathcal{P}(f)$ will have no impact on equality. This allows us to generate an alternate random function, p' , such that $|\mathcal{P}'| = |\mathcal{P}|$, and taking its inverse Fourier transform, as (5) guarantees that both will have the same mean and time ACF, and therefore the same power spectrum. The generated $p'_i(t)$ are the phase-randomized surrogate realizations.

FOR OUR DEMONSTRATION, WE USE A 150-YEAR CONTROL RUN FROM A COUPLED E3SM SIMULATION UNDER PREINDUSTRIAL CONDITIONS, WITH NO INFLUENCE FROM HUMAN-CAUSED GREENHOUSE GAS EMISSIONS.

Statistical Generation

The randomized PCs $p'_i(t)$ can now be incorporated into (2) with the associated EOFs to generate an independent randomized realization of the original input data

$$F'_{vn}(t, s) = \sum_{i=1}^r p'_i(t) e_i(s)^T \quad (6)$$

By virtue of the Wiener–Khinchin theorem and the orthonormal nature of EOF decomposition, this method ensures that the main spatial and temporal patterns of variability are retained.

Phase randomization can be repeated as many times as necessary to build an ensemble of independent realizations of the variability fields. We multiply the generated variability field by the point-by-point standard deviations $F'_v(\mathbf{x}_0, t) = \sigma_{F_v}(\mathbf{x}_0) F'_{vn}(\mathbf{x}_0, t)$ to

renormalize the variability before it is added back into (1) to construct independent realizations of the climate model output

$$F'_{gen}(\mathbf{x}_0, t) = F'_i(\mathbf{x}_0, t) + F'_s(\mathbf{x}_0, t) + F'_d(\mathbf{x}_0, t) + F'_v(\mathbf{x}_0, t). \quad (7)$$

The work presented in this article makes heavy use of Python packages for climate data analysis, particularly the Model for Prediction Across Scales (MPAS)-analysis^a diagnostics package and scikit-learn for machine learning. The former is built to work with the unstructured meshes of the ocean and cryosphere components in the Energy Exascale Earth System Model (E3SM). The use of openly available Python packages for the entire generation workflow ensures that this method will be easily usable and adaptable by the broader climate community.

RESULTS

Data Description

Here, we demonstrate that the method described in the previous section can efficiently generate many realizations of climate forcing for ice sheet models that are statistically similar to the single long climate model simulation from which they are derived. For our demonstration, we use a 150-year control run from a coupled E3SM simulation under preindustrial conditions, with no influence from human-caused greenhouse gas emissions. This simulation uses the E3SM V1 Cryosphere configuration, which allows for ocean circulation in Antarctic ice shelf cavities and heat and mass exchange between the ocean and the base of the ice shelves. This configuration was designed to simulate realistic Antarctic sub-ice shelf melting within a coupled, global climate model, allowing for the expression of relevant modes of coupled climate variability (e.g., the El Niño–Southern Oscillation). This is in contrast to typical model configurations used in large ensembles, which are relatively coarser (100-km resolution) and do not allow for ocean circulation beneath floating ice shelves. The E3SM V1 Cryosphere configuration and relevant simulations are described in detail by Comeau et al.¹¹

We use relevant output from the MPAS-Ocean component of E3SM, which includes land-ice freshwater flux, i.e., the flux of freshwater across the ice-ocean model boundary due to melting or freezing. This is related to the basal melt rate of the ice shelves, which we plot in Figure 2 in meters of water equivalent per year (m/yr), assuming a constant freshwater density of 1000 kg/m³. Here we use monthly averaged output

^a<https://github.com/MPAS-Dev/MPAS-Analysis>

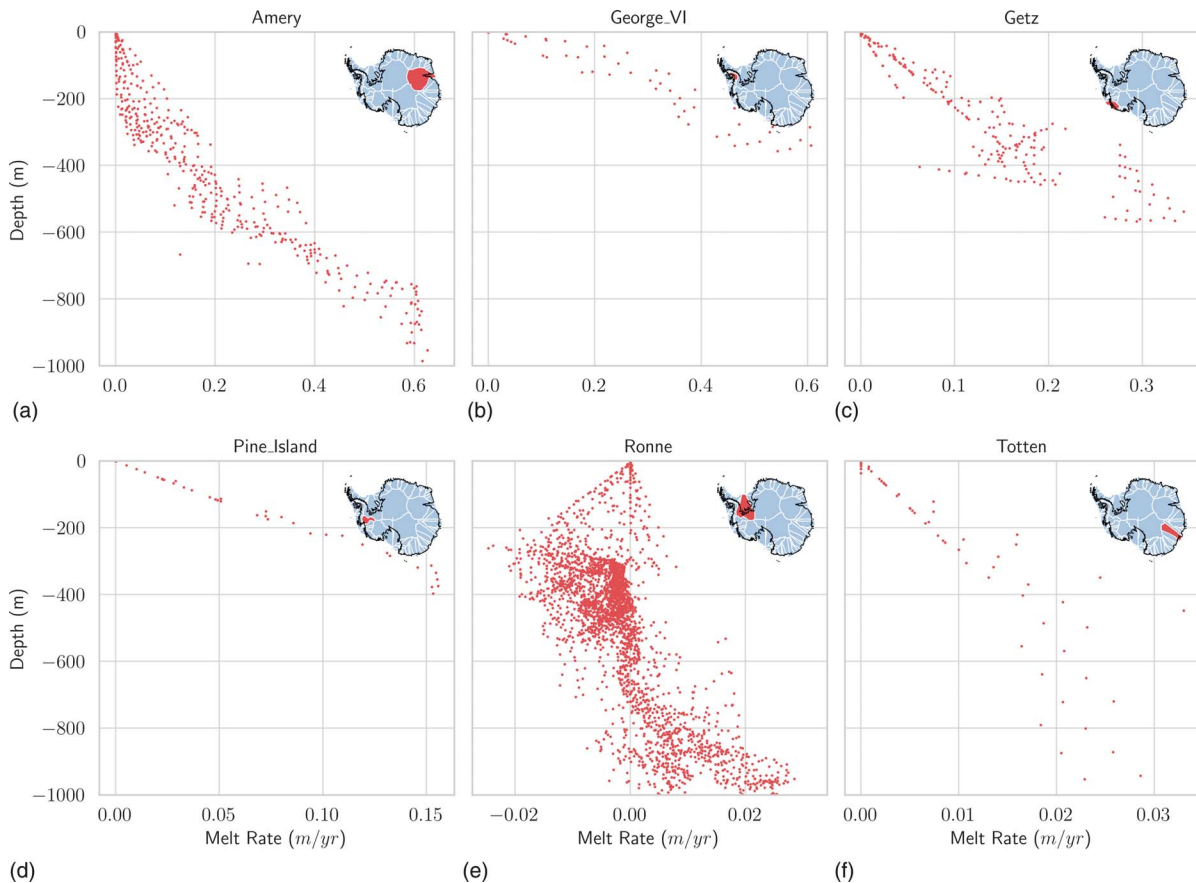


FIGURE 2. Dependence of melt rate on depth (ice draft), i.e., $F_d(x_0)$ from (1), shown for the temporal mean melt rates across six sample ice shelf catchments: (a) Amery, (b) George VI, (c) Getz, (d) Pine Island, (e) Ronne, and (f) Totten. The spatial boundary of each ice shelf is noted in red in the inset maps. Note the difference in scale of the melt-rate magnitude across different ice shelves. m/yr: meters of water equivalent per year.

from the E3SM simulation. For the purposes of this article, the model's output was initially regridded to a rectangular grid with a spatial resolution of 10 km, although the results shown here are not particularly sensitive to the choice of regridded resolution. It was also trimmed to remove the influence of transient model spin-up time, resulting in a 125-year-long simulation at monthly resolution. This length of model time and resolution are sufficient to capture any multidecadal, low-frequency variability of the ocean forcing present in the model simulation. We stress that this particular model's output is used as a demonstration of the statistical generation method described in the previous section; in the future, we aim to apply this same analysis to improved E3SM simulations that include much higher spatial resolution in the ocean (~ 10 km) around Antarctica, and a longer duration.

Boundaries of the various Antarctic ice shelf basins considered in this article are taken from E3SM

definitions and characterize the ice sheet domain as 100 distinct regions used to calculate individual melt-draft relationships and subsequent results. We find no significant linear trend in the model's output due to the absence of any prescribed greenhouse gas emissions forcing on global climate during the preindustrial era. Consequently, basal melt is stationary in the trimmed model's output. As described in the previous section, the seasonality and draft dependence of basal melt rates is first removed prior to EOF decomposition. Figure 2 presents the draft dependence of six sample catchments from different regions of Antarctica, selected to highlight the degree of variability in the spatial distribution of basal melt across the ice sheet. At some smaller ice shelves, such as Pine Island [Figure 2(d)], nearly the entire spatial variation in melt can be explain by this linear dependence on depth. At other large ice shelves, like Ronne [Figure 2(e)], there is considerable spatial complexity in melt rates

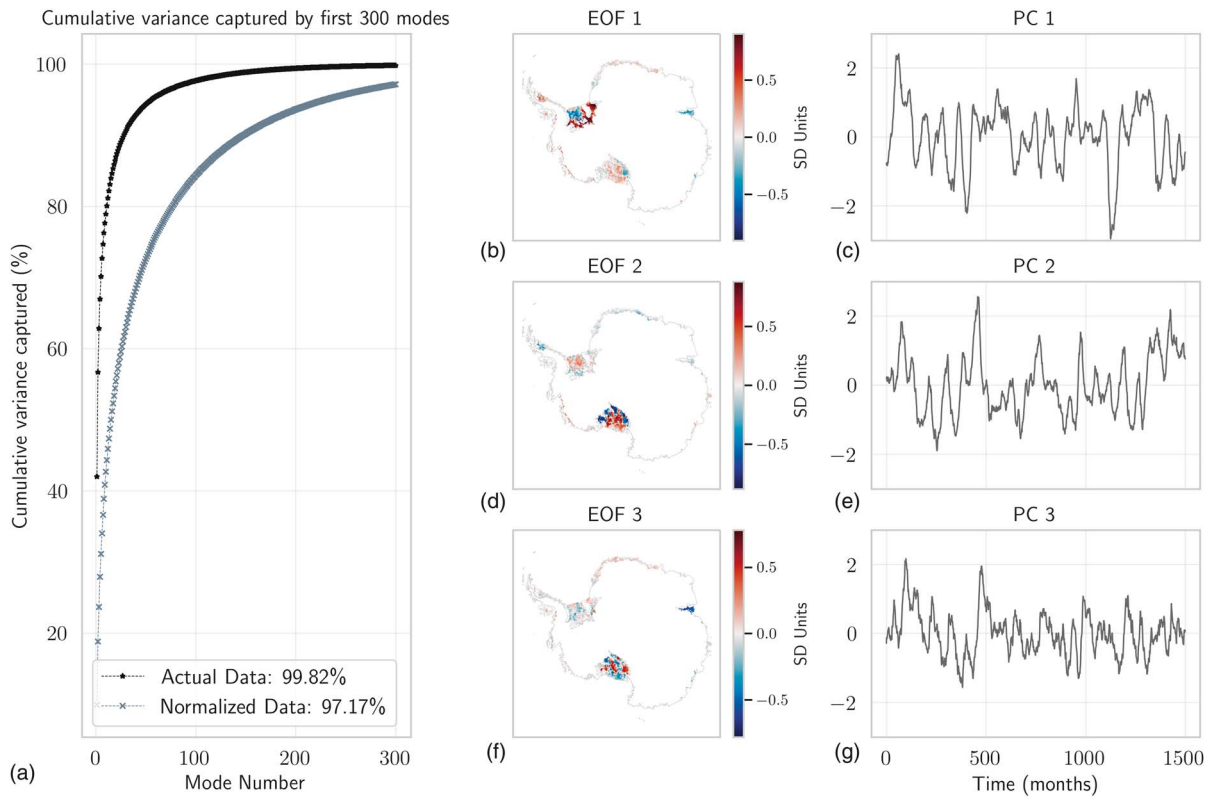


FIGURE 3. (a) Cumulative variance in the dataset captured by different modes of variability from the EOF decomposition. Only a sample of the total number of modes is shown here for the EOF decomposition, applied to actual and normalized model output. The percentage of explained variance for the dominant modes reduces upon normalization. (b), (d), and (f) The three most dominant modes of spatial variability from the EOF decomposition of the normalized output, with their associated PC time series in (c), (e), and (g). The values in (b)–(g) are scaled by the square root of the eigenvalues for consistency. SD: standard deviation.

caused by more complex ocean circulation patterns and the Coriolis effect.

Generating Realizations of Basal Melt Variability

Following preprocessing of the basal melt output from E3SM, we are left with a dataset of monthly basal melt anomalies that include only the nonannual and nondepth-dependent components that we seek to use for the generation method. We start by performing a traditional EOF decomposition, as described in the previous section [Figure 3(a)]. The rank of the input dataset provides an upper bound on the number of distinct modes of variability obtained by the decomposition (in this case, 1500 modes). In this particular dataset, most of the variance is captured by the first 10 modes, with very little difference in explanatory power in higher-order modes.

Figure 3(b)–(g) presents a sampling of the dominant spatial fields of variability alongside their respective projection coefficients. When each grid point from

the model is first normalized by the local temporal variance, dominance of the first mode in the nonnormalized decomposition is suppressed, and spatial patterns of other modes of variability emerge strongly in the Filchner-Ronne [Figure 3(b)] and Ross ice shelves [Figure 3(d)]. We use the normalization procedure in our method because it ultimately improves the performance of the generation method.

As explained in the previous section, following EOF decomposition, we Fourier transform each PC time series. To generate a new spatiotemporal realization of basal melt over Antarctic ice shelves, we randomize all Fourier phases of all PCs, perform an inverse Fourier transform, and recombine the PC-weighted EOFs. To assess the performance of the generation method in matching the statistics of the original dataset, we compare power spectral densities of the generated output to the original basal melt fields, averaged over all of Antarctica as well as individual ice shelf catchments. The generated output, reconstructed using the

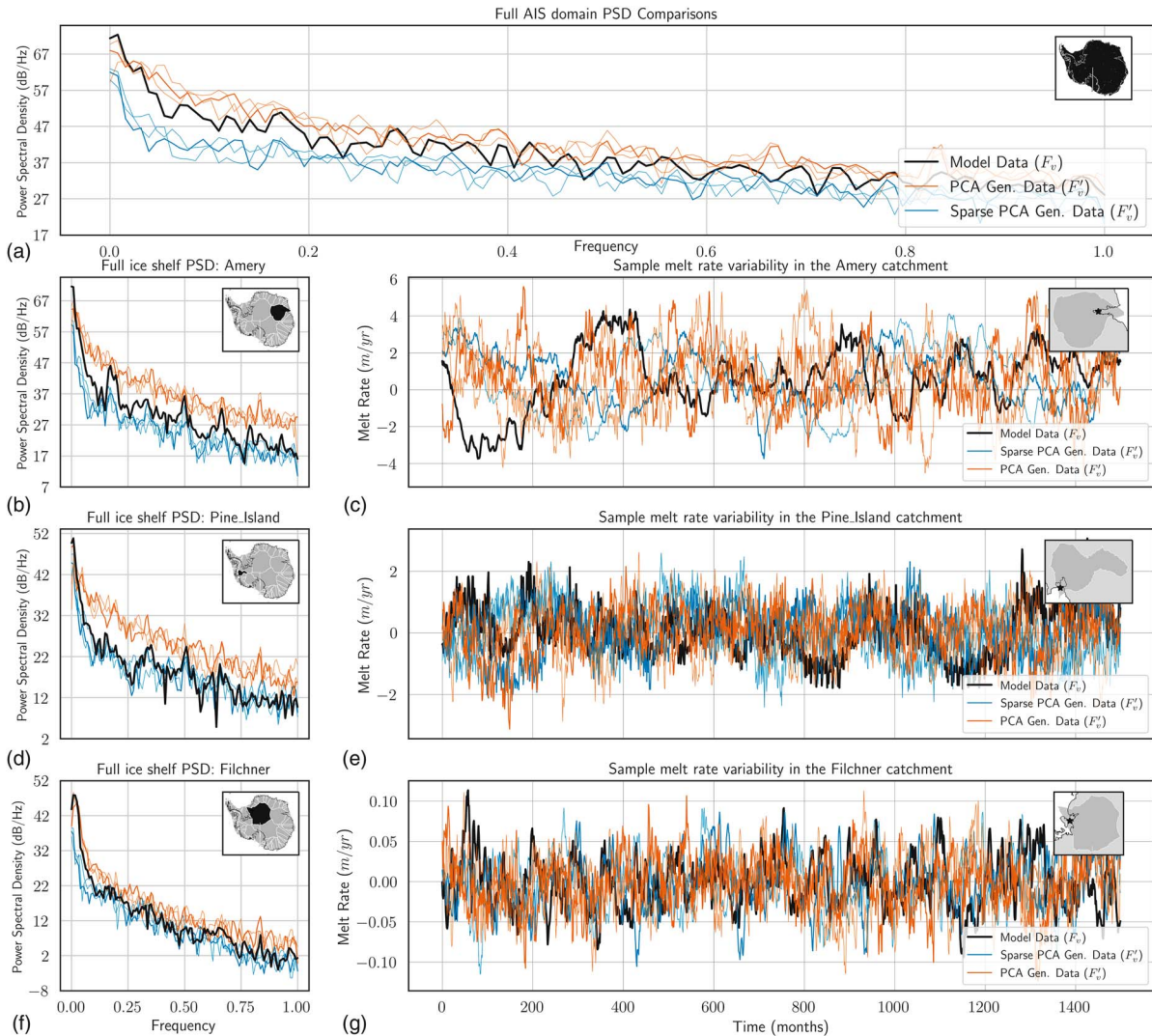


FIGURE 4. Comparison of the model simulation data (black) with randomized realizations generated using a standard EOF/PCA decomposition (orange) and a sparse PCA algorithm (blue). PSD of the cumulative melt rate across (a) the full ice sheet, and individual catchments of (b) Amery, (d) Pine Island, and (f) Filchner. (c), (e), and (g) Time evolution of a sample grid location (marked in the inset map) from each basin. AIS: Antarctic ice sheet; gen.: generated; m/yr: meters of water equivalent per year.

randomized PCs with the EOFs of the input dataset, reproduces well the full spectrum of frequencies in the entire ice sheet domain [the orange lines in Figure 4(a)]. This is because the decomposition is performed over the entire Antarctic ice sheet domain, and all spatial grid points are treated equally. However, the time series for individual catchments are obtained by integration over spatial subdomains of which traditional PCA is “unaware.” Depending on the nature of the EOF decomposition, high-frequency variability may be introduced to some ice shelves, resulting in the departure of power spectral densities of the generated data

from the original model data at subannual time scales [Figure 4(b), (d), and (f)].

To remedy this issue of spurious high-frequency variability, we make use of a sparse decomposition algorithm that optimizes an L1 regularization penalty on the EOF modes. To retain spatiotemporal characteristics of internal variability across individual basins, we want to obtain power spectral densities that are as close as possible to the original dataset. Figure 4 shows the sparse generator results (blue lines) for select catchments, and evolution of a sample grid location in each catchment. Here, we found that sparse

generators with a higher penalization factor introduced higher correlations among PCs, and that lower values were sufficient to result in considerable improvement in the ice shelf power spectral densities [Figure 4(b), (d), and (f)], without much loss in its performance over the full ice sheet domain. Of the nonzero correlations introduced by sparse PCA, the only ones that were significant were between higher-order modes that captured relatively lower variances. Thus, we are confident that the added correlations (which are not preserved during phase randomization of the generation) represent a negligible contribution to the total variance in the dataset.

DISCUSSION

The method described in this article builds on the method described in Link et al.,⁶ where a phase-randomized EOF-based method was used to emulate variability fields of surface atmospheric temperature at the global scale. However, the field generator in that method assumed stationarity and did not account for periodic components or spatial dependencies of the variable on other fields, which are significant for variability modes in polar ocean systems. The generated output “inherits” the spatial complexity at all scales of variability, as evidenced by the fields visible in different basins. Another difference with Link et al.⁶ is the use of a sparse PCA algorithm, which is more flexible and improves generator performance for high-dimensional datasets like the basal melt rates investigated here. We are also able to isolate the dependence of basal melt rates on ice shelf draft through a regression polynomial fit. This removes the effect of ice sheet geometry, and the generator is passed only what is deemed to be internal climate variability in the model. Adding this draft dependence back to the basal melt rate following the generation of randomized variability ensures that the melt rates predicted through this method include feedbacks on ice sheet state. Such draft dependence can be straightforwardly incorporated within ice sheet models directly. The time evolution of four grid locations with this spatial draft dependence and seasonality added back in is presented in Figure 5. These points are chosen across separate catchments, where different aspects of the variability are important. For example, the impact of seasonality is much higher in a point sampled from the George VI ice shelf, which is strongly influenced by the seasonal cycle of sea ice [Figure 5(c)], while draft dependence takes precedence for the Filchner ice shelf [Figure 5(d)]. As expected, decadal variability dominates at a deeper point on the Amery ice shelf in Figure 5(a).

Previous work has identified long time scales of variability (i.e., interdecadal to multidecadal) in ocean

forcing to be dominant drivers of uncertainty in future ice sheet projections, particularly when sectors of the ice sheet lose stability.¹² By using a sparse version of the EOF decomposition, we are able to ensure an accurate representation of subannual variability when compared to a traditional EOF decomposition. This approach also prioritizes the accurate generation of low-frequency variability across the ice sheet domain, as shown by the multidecadal signals in Figure 4. Another benefit of the EOF-based approach taken from this article is that the influence of different temporal and spatial modes of variability can be explicitly modified to study the contribution of different modes to uncertainty in ice sheet projections. Spectral filters between the Fourier transform and the randomization steps can be used to explicitly modify the importance of different time scales of variability within different modes, which may be relevant to their particular use case.

Other recent emulator approaches have been limited to reduced-order problems, or assume normally distributed input data, which is not a requirement of the method described in this article. One study applied to Antarctica-sampled, uniform-idealized melt-rate trajectories (absent temporal variability) across large sectors of the continent to generate ensemble members.⁸ However, this approach neglected the draft dependence of melt rates and used a sigmoid function to smooth over the complex spatial patterns that have been identified to be present within and across ice shelf basins.² Such a method is useful for sampling uncertainty in the mean melt rate (due to ocean model process uncertainty), but not due to the inherent variability of oceanic and atmospheric processes.

The approach presented in this article is also not without its limitations. We see that although the PSD of the original dataset is retained, the generator output attenuates the magnitude of variability at a few specific grid points within the larger ice shelves (notably for Amery). Additional research will be required to identify the optimal range of the penalization factor to best reconstruct the input dataset. We expect that a weaker penalization of sparsity in the EOF decomposition may potentially improve generator performance, although with the potential downside of introducing spurious subannual variability, as in traditional PCA. Such an approach may also be used in conjunction with a low-pass filter to reduce added spurious subannual variability. Other sparse PCA formulations, such as methods that modify and adapt the least absolute shrinkage and selection operator approach for PCA decomposition, or the exactly uncorrelated sparse PCA method, are promising and may provide better performance

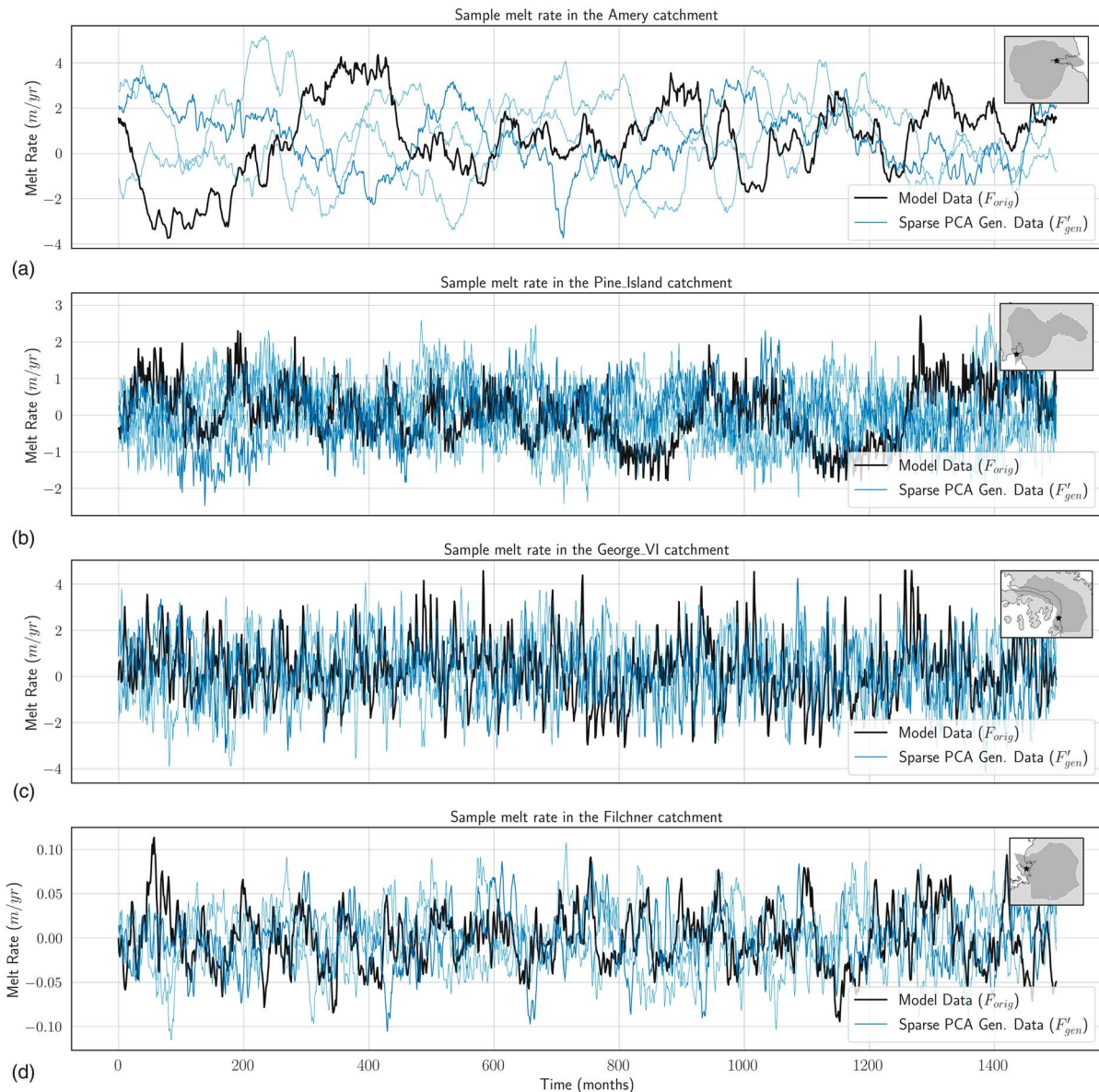


FIGURE 5. Time evolution of select grid locations (marked in the inset) in four ice shelves: (a) Amery, (b) Pine Island, (c) George VI, and (d) Filchner, with seasonality and spatial draft dependence added back in for the overall melt rate at that location. gen.: generated; m/yr: meters of water equivalent per year.

than the implementation used here. Alternative methods that make use of dynamic-mode decomposition, which have been applied to the stochastic parameterization of atmospheric flow problems, may also be of further interest here.

We emphasize again that the primary motivation for this article is to support uncertainty quantification workflows in constraining the Antarctic ice sheet response to variability in ocean forcing. The method we described, tested, and validated accomplishes the

goal that we have set out for it for the target model's output. However, our larger goal is to provide generation capabilities for the widest possible range of climatic fields, and further improvements could further broaden the applicability of this method.

CONCLUSION

The evolution of mass loss from Antarctica has been identified as a major source of uncertainty in future

sea level projections. Ice shelves that fringe the continent are particularly susceptible to warming oceans, with some already having begun to thin due to increased melting at the base. Model-based probabilistic projections of ice sheet change require robust uncertainty quantification frameworks to better understand the influence of intrinsic climate variability on projection uncertainties, with serious implications for policy decisions and climate impact studies. However, full-physics-based numerical climate models are prohibitively expensive to run and are therefore limited in their ability to sample the large parameter spaces necessary for quantifying this uncertainty. To this end, we devised a computationally efficient method to generate many realizations of climate forcing on ice sheets, which reproduces the statistics of simulations from physical climate models.

We stress that the E3SM V1 Cryosphere Campaign dataset was used as a demonstration of the generator method and anticipate using more recent and better calibrated simulation runs to force an ice sheet model ensemble in future work. The method was also built with flexibility in mind so that it can be applied to the widest possible range of climate fields beyond those considered here. This method would also be applicable to other studies where the output from climate models is necessary to force a physical model of, e.g., oceans, ecosystems, the solid earth, or other components of the Earth system. Additionally, the implementation of this method in Python and use of widely available open source packages makes it easy to use in conjunction with many existing modeling workflows.

In the future, we plan to incorporate this method directly within ice sheet models such that ice sheet model ensembles can directly take ocean simulation output and generate forcing for ensemble simulations of ice sheet evolution under variable ocean forcing, including feedbacks with ice shelf shape on basal melt rates. In this way, this generation method can reduce the computational cost of generating large ensemble simulations of ice sheets under historical and future forcing scenarios.

ACKNOWLEDGMENT

This work was supported by the U.S. National Science Foundation under Award 1947882 as a part of the Antarctic Ice Sheet Large Ensemble project. It was also supported by the Earth and Environmental System Modeling program, funded by the U.S. Department of Energy's Office of Science, Biological and Environmental Research program. The data and analysis code for the results presented in this article are available as an

archived Zenodo object at <https://doi.org/10.5281/zenodo.7633997>. Full source, documentation, and future updates may be found in the GitHub repository at https://github.com/mshiv/aislens_emulation.

REFERENCES

1. H. Seroussi et al., "ISMIP6 Antarctica: A multi-model ensemble of the Antarctic ice sheet evolution over the 21st century," *Cryosphere*, vol. 14, no. 9, pp. 3033–3070, Sep. 2020, doi: [10.5194/tc-14-3033-2020](https://doi.org/10.5194/tc-14-3033-2020).
2. S. Adusumilli et al., "Interannual variations in meltwater input to the Southern Ocean from Antarctic ice shelves," *Nature Geosci.*, vol. 13, no. 9, pp. 616–620, Aug. 2020, doi: [10.1038/s41561-020-0616-z](https://doi.org/10.1038/s41561-020-0616-z).
3. S. Jacobs, H. Helmer, C. Doake, A. Jenkins, and R. Frolich, "Melting of ice shelves and the mass balance of Antarctica," *J. Glaciol.*, vol. 38, no. 130, pp. 375–387, 1992, doi: [10.3189/S0022143000002252](https://doi.org/10.3189/S0022143000002252).
4. C. Y. Tsai, C. E. Forest, and D. Pollard, "The role of internal climate variability in projecting Antarctica's contribution to future sea-level rise," *Climate Dyn.*, vol. 55, nos. 7–8, pp. 1875–1892, Jul. 2020, doi: [10.1007/s00382-020-05354-8](https://doi.org/10.1007/s00382-020-05354-8).
5. D. S. Wilks and R. L. Wilby, "The weather generation game: A review of stochastic weather models," *Prog. Phys. Geography*, vol. 23, no. 3, pp. 329–357, Sep. 1999, doi: [10.1177/030913339902300302](https://doi.org/10.1177/030913339902300302).
6. R. Link, A. Snyder, C. Lynch, C. Hartin, B. Kravitz, and B. Bond-Lamberty, "Fldgen v1.0: An emulator with internal variability and space–time correlation for Earth system models," *Geosci. Model Develop.*, vol. 12, no. 4, pp. 1477–1489, Apr. 2019, doi: [10.5194/gmd-12-1477-2019](https://doi.org/10.5194/gmd-12-1477-2019).
7. V. Verjans, A. Robel, H. Seroussi, L. Ultee, and A. Thompson, 2022, "The stochastic ice-sheet and sea-level system model v1.0 (StISSM v1.0)," EGU sphere [Preprint], <https://doi.org/10.5194/egusphere-2022-699>.
8. M. Berdahl, G. Leguy, W. H. Lipscomb, and N. M. Urban, "Statistical emulation of a perturbed basal melt ensemble of an ice sheet model to better quantify Antarctic sea level rise uncertainties," *Cryosphere*, vol. 15, no. 6, pp. 2683–2699, 2021, doi: [10.5194/tc-2020-178](https://doi.org/10.5194/tc-2020-178).
9. I. M. Johnstone and A. Y. Lu, "On consistency and sparsity for principal components analysis in high dimensions," *J. Amer. Statistical Assoc.*, vol. 104, no. 486, pp. 682–693, Jun. 2009, doi: [10.1198/jasa.2009.0121](https://doi.org/10.1198/jasa.2009.0121).
10. T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D, Nonlinear Phenomena*, vol. 142, nos. 3–4, pp. 346–382, Aug. 2000, doi: [10.1016/S0167-2789\(00\)00043-9](https://doi.org/10.1016/S0167-2789(00)00043-9).
11. D. Comeau et al., "The DOE E3SM v1.2 cryosphere configuration: Description and simulated Antarctic

ice-shelf basal melting," *J. Adv. Model. Earth Syst.*, vol. 14, no. 2, Feb. 2022, Art. no. e2021MS002468, doi: [10.1029/2021MS002468](https://doi.org/10.1029/2021MS002468).

12. A. A. Robel, H. Seroussi, and G. H. Roe, "Marine ice sheet instability amplifies and skews uncertainty in projections of future sea-level rise," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 30, pp. 14,887–14,892, Jul. 2019, doi: [10.1073/pnas.1904822116](https://doi.org/10.1073/pnas.1904822116).

SHIVAPRAKASH MURUGANANDHAM is a graduate student at the Georgia Institute of Technology, Atlanta, GA, 30332, USA. His research focuses on ice sheet modeling. Muruganandham received his master's degree in cybernetic systems/control from Czech Technical University in Prague. He is a member of the American Geophysical Union and International Glaciological Society. Contact him at mshiv@gatech.edu.

ALEXANDER A. ROBEL is an assistant professor at the Georgia Institute of Technology, Atlanta, GA, 30332, USA. His research focuses on ice sheet modeling and sea level projections. Robel received his Ph.D. degree in earth and planetary sciences from Harvard University. He is a member of the American Geophysical Union and International Glaciological Society. Contact him at robel@eas.gatech.edu.

MATTHEW J. HOFFMAN is part of the Fluid Dynamics and Solid Mechanics group at Los Alamos National Laboratory, Los Alamos, NM, 87545, USA, where he leads the ice sheet modeling team. His research interests include glacier basal physics, ice sheet/ocean interactions, and regional sea level change. Hoffman received his Ph.D. degree in environmental sciences and resources from Portland State University. He is a member of the American Geophysical Union and International Glaciological Society. Contact him at mhoffman@lanl.gov.

STEPHEN F. PRICE is part of the Fluid Dynamics and Solid Mechanics group at Los Alamos National Laboratory, Los Alamos, NM, 87545, USA. He currently leads the polar processes and sea level rise science focus area for the U.S. Department of Energy's Energy Exascale Earth System Model. His research focuses on improving our understanding of glacier, ice sheet, and sea level change through the use of Earth system models. Price received his Ph.D. degree in geophysics from the University of Washington. He is a member of the American Geophysical Union and International Glaciological Society. Contact him at sprice@lanl.gov.

IEEE COMPUTER SOCIETY
Call for Papers

Write for the IEEE Computer Society's authoritative computing publications and conferences.

GET PUBLISHED
www.computer.org/cfp

IEEE COMPUTER SOCIETY IEEE