
PHYSICS-CONSTRAINED GENERATIVE MACHINE LEARNING-BASED HIGH-RESOLUTION DOWNSCALING OF GREENLAND'S SURFACE MASS BALANCE AND SURFACE TEMPERATURE

PREPRINT

 **Nils Bochow**^{*1,2},  **Philipp Hess**^{†2,3}, and  **Alexander Robinson**^{‡4}

¹Department of Mathematics and Statistics, Faculty of Science and Technology, UiT The Arctic University of Norway, Tromsø, Norway

²Potsdam Institute for Climate Impact Research, Potsdam, Germany

³Technical University of Munich, Munich, Germany; School of Engineering & Design, Earth System Modelling

⁴Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

July 31, 2025

ABSTRACT

Accurate, high-resolution projections of the Greenland ice sheet's surface mass balance (SMB) and surface temperature are essential for understanding future sea-level rise, yet current approaches are either computationally demanding or limited to coarse spatial scales. Here, we introduce a novel physics-constrained generative modeling framework based on a consistency model (CM) to downscale low-resolution SMB and surface temperature fields by a factor of up to 32 (from 160 km to 5 km grid spacing) in a few sampling steps. The CM is trained on monthly outputs of the regional climate model MARv3.12 and conditioned on ice-sheet topography and insolation. By enforcing a hard conservation constraint during inference, we ensure approximate preservation of SMB and temperature sums on the coarse spatial scale as well as robust generalization to extreme climate states without retraining. On the test set, our constrained CM achieves a continued ranked probability score of 6.31 mmWE for the SMB and 0.1 K for the surface temperature, outperforming interpolation-based downscaling. Together with spatial power-spectral analysis, we demonstrate that the CM faithfully reproduces variability across spatial scales. We further apply bias-corrected outputs of the NorESM2 Earth System Model as inputs to our CM, to demonstrate the potential of our model to directly downscale ESM fields. Our approach delivers realistic, high-resolution climate forcing for ice-sheet simulations with fast inference and can be readily integrated into Earth-system and ice-sheet model workflows to improve projections of the future contribution to sea-level rise from Greenland and potentially other ice sheets and glaciers too.

Keywords Generative modeling · Greenland Ice Sheet · Surface Mass Balance · Downscaling

1 Introduction

The Greenland ice sheet (GrIS) is the second largest ice sheet with an ice volume of 7.4 m sea level equivalent [1]. Anthropogenic warming has driven a significant acceleration of Greenland mass loss over recent decades [2]. Currently, the mass loss of the GrIS is approximately evenly driven by dynamic discharge and a decrease in the surface mass balance (SMB), with ca. 60% of GrIS ice loss due to the changes at the surface [3, 4]. However, by the end of this

^{*}contact@nilsbochow.com

[†]philipp.hess@tum.de

[‡]alexander.robinson@awi.de

Summary of Comments on Physics-constrained generative machine learning-based high-resolution downscaling of Greenland's surface mass balance and surface temperature

This page contains no comments

century, the importance of surface processes is projected to increase as marine outlets retreat and dynamic processes will play a smaller role [4–6]. That means that the SMB will be the determining factor of the overall mass balance of the GrIS. Modeling the GrIS surface mass balance demands fine resolution to capture the steep orographic gradients along the ice-sheet margin that govern precipitation patterns [7]. High-resolution simulations also resolve the narrow ablation zone and peripheral outlet glaciers, where most of the melt and runoff occur [8]. These high-resolution SMB fields then feed into ice-sheet model intercomparisons (e.g., ISMIP6), improving projections of dynamic thinning and sea-level contribution [5]. Therefore, it is crucial to have realistic and high-resolution projections of the SMB in a changing climate.

Current methods to calculate the SMB have several disadvantages. They are either expensive to run, based on simple parameterisations or have low resolution. On short time scales, i.e., centennial scales, regional climate models such as MAR [9, 10] or RACMO [11] are used to simulate the surface mass balance. These RCMs are specifically designed to simulate the polar regions. However, they are computationally expensive to run and are usually not accounting for a changing ice-sheet topography. On longer time scales, the SMB is usually simulated via simple parameterisation schemes such as the positive degree days method (PDD) [12–14] or energy balance models [15–17]. Alternatively, the SMB can be directly derived from ESM runs as a sum of precipitation, melt and runoff in a first approximation [18, 19]. However, the ESM-derived SMB fields are limited to the, often coarse, resolution of the ESM and cannot resolve the steep gradients of the ice sheet, leading to biases in SMB estimates [20].

Given these computational constraints and resolution trade-offs across centennial and longer scales, computationally fast approaches are required to efficiently produce high-resolution SMB estimates. Previous approaches to downscaling the SMB of Greenland are often based on statistical downscaling [21, 22]. These approaches leverage the relationship between elevation, SMB and temperature, and have shown promising results in downscaling RCM results to higher resolutions. However, these statistical methods are inherently limited by their prescribed relationships between melt and elevation, which, for example, do not hold for the Antarctic ice sheet and can therefore not easily be extended to other regions [22]. Additionally, these statistical methods cannot easily integrate complementary inputs that could improve the downscaling task [22]. Therefore, novel machine-learning techniques, which can be trained to learn complex relationships, including topographic and climatological patterns, from high-resolution simulations and then run orders of magnitude faster are a promising approach. In contrast to previous statistical approaches, we aim to provide a framework in this work that is able to downscale SMB and T_s fields from other sources than RCMs, that is, deriving high-resolution fields directly from ESM output or from simpler physics-based models. Machine learning (ML) methods have gained tremendous attention in climate science in the last years, ranging from reconstruction of climate fields [23], ML-based global circulation and weather models [24, 25], downscaling [26] and weather forecasting [27–29]. In the glaciological community, ML has, for example, been used to improve ice-flow simulations [30, 31], uncover flow laws of Antarctic ice shelves [32], improve upon traditional ice-shelf melt parameterisations [33], forecast glacier-wide SMB [34–36] or as emulators for subglacial hydrology models [37].

Generative modeling approaches, and diffusion models in particular, have shown promising results in downscaling of ESM fields [26, 38–40]. The principle behind downscaling with generative diffusion models is based on so-called denoising, where a network is first trained to remove noise added to a high-resolution sample. Once trained, the network can then be applied for downscaling by iteratively generating realistic small-scale patterns from noise while following the large-scale patterns of the low-resolution ESM. However, most applications of generative downscaling so far are (i) limited to univariate fields (e.g. precipitation), (ii) struggle with training instabilities or mode collapse [41], (iii) require a large amount of sample steps to generate realistic fields [42] and/or (iv) lack physical mass conservation [26].

Here, we introduce a generative modeling-based methodology to derive high-resolution surface mass balance and temperature at the surface (T_s) fields for the Greenland ice sheet. The SMB and T_s are the most important, minimal surface forcing fields needed to drive an ice sheet model. We use a generative consistency model (CM) [26, 43] to demonstrate realistic and efficient downscaling of low-resolution SMB and temperature fields by at least a factor of 16 (Fig. 1, quadrupling the resolution gain of prior work, which was limited to a factor of 4 [26]). In contrast to diffusion models that require dozens or even hundreds of time-steps, a CM directly learns a consistency function that maps a noised version of the target field back to its clean counterpart, allowing for fast generation of samples. By adding hard constraints during the sampling process, that is, local mass conservation of the SMB and conservation of T_s of the low-resolution field, we are able to generalize to more extreme climate conditions without the need for retraining. In theory, our model allows the simultaneous downscaling of arbitrarily many variables given high-quality training data. Compared to previous SMB downscaling approaches, our method does not explicitly prescribe any elevation-SMB relationship.

We train our CM on high-resolution simulations of SMB and T_s from the regional climate model MARv3.12 [9, 10]. Once trained, the CM can be used to generate high-resolution SMB fields from ESM outputs using either directly the bias corrected SMB fields from the ESM or using a hybrid setup, combining the generative model with a positive

 Number: 1 Author: Subject: Highlight Date: 09.10.2025, 16:10:30
I don't understand this part because your model is trained on RCM data, so how can it downscale SMB from another source than an RCM?

 Number: 2 Author: Subject: Comment on Text Date: 09.10.2025, 16:04:28
introduce in the sentence above when it first comes up

 Number: 3 Author: Subject: Comment on Text Date: 09.10.2025, 16:12:12
Here there is also earlier work by van der Meer et al 2023: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022MS003593>

degrees days (PDD) approach (Fig. 2). We are able to produce realistic SMB and T_s fields with a 5 km resolution openly available output from ESMs and MAR (Fig. 1). Ultimately, our methodology can be used for any climate field and potentially be directly integrated into ESMs.

2 Results

We train and provide two versions of the CM (cf. Methods). In the first version, we only downscale the SMB, while in the second version we additionally downscale temperature fields T_s . In both versions, we include the ice sheet height and the monthly average of the insolation as additional conditioning input. Additionally, we implement a hard conservation constraint during inference to conserve the SMB and temperature on a regional level during the downscaling (c.f. Methods). If not explicitly stated differently, we always show the hard-constrained CM version. In the following, we concentrate on the SMB downscaling task, since downscaling the temperature is generally easier due to the relative smoothness of the temperature fields. The model workflow is depicted in Fig. A.1. We use monthly output from the RCM MARv3.12 forced by different ESMs over the time period 1950-2100 (1), in total 23 model runs, i.e., 21432 monthly fields (c.f. Methods).

Here, we evaluate how our model performs on two different downscaling tasks. First, we show the downscaling results of the CM using artificially coarsened MAR test set data. This has the advantage that we have paired low- and high-resolution fields for evaluations. Secondly, we show how the CM can be used to derive high-resolution SMB and T_s fields directly from unpaired low-resolution ESM output.

2.1 Downscaling low-resolution MAR

The held-out test set consists of 16 random years between the time period 1950 and 2100. In total, 2448 months have been held out from the train and validation set.

To test the ability of the CM to downscale from low to high resolution, we first coarsen the MAR fields by a factor of 16, that is, 16 times lower resolution than the high-resolution fields (80x80 km). This corresponds to a typical spatial resolution as provided by state-of-the-art ESMs. Additionally, we test our CM approach to downscale spatial fields from a factor 32 lower resolution (160x160 km). We downscale these pooled fields with our CM to recover the small-scale variability.

¹Qualitatively, both the downscaled SMB and T_s fields are visually indistinguishable from the ground truth, while the linearly interpolated fields show a reduced small-scale variability (Fig. 1&3). This is visible in the radially averaged power spectral density (PSD) of the SMB over the whole test set (Fig. 3l). The coarse field shows no small-scale variability (blue), while the linearly interpolated field shows a substantial underestimation of the small-scale variability (orange). In contrast, the downscaled fields using the CM show a highly accurate representation of small-scale variability compared to the ground truth. In particular, the PSD of the hard constrained field (CM HC) is visually indistinguishable from the ground truth.

The mean absolute error (MAE) of one realization of the unconstrained downscaled field compared to the ground truth MAR is 17 mmWE/month over the whole test set. The mean correlation of the unconstrained CM SMB field is $r = 0.84$. The hard constrained fields show a substantially lower MAE = 13.56 mmWE/month and higher correlation $r = 0.97$ with the ground truth compared to the unconstrained downscaled fields. Similarly, the hard constrained downscaled T_s field shows the lowest MAE = 0.16 K and correlation $r = 1.00$ compared to the ground truth (Fig. 3f). For comparison, the linearly interpolated field has a mean absolute error MAE = 0.53 K and correlation of $r = 0.98$ (Fig. 3j).

³Additionally, we coarsen the MAR fields by a factor 32 (160 km resolution) and show that the CM, in principle, still can downscale the fields (Fig. B.2). ⁴We show the same exemplary month as in Fig. 3 with coarsening factor 16. While the ⁵performance of the CM is still better than simply linearly interpolating, it is worse than for the downscaled field from the coarsened field by factor 16. The integrated monthly SMB is still mostly conserved with a residual of less than 25 Gt for almost all months (Fig. B.2b). However, due to the very coarse resolution of the input field, some information is lost during the coarsening process. ⁶Again, the error is largest at the margins, especially the southeastern margin (Fig. B.2c,d). It should be noted that due to the coarse resolution and the overall conservation of the per-cell SMB, the pixel edges are slightly visible in the MAE and the output fields. This is, for example, visible in southwestern Greenland with negative SMB in Fig. B.2e, where the relatively sharp edge from the coarse field (Fig. B.2f) propagates into the downscaled field. This is per construction due to the hard-constraining but could be improved by using the linearly interpolated field as input with the tradeoff of losing information due to the interpolation process. However, the small spatial scales are still substantially better resolved than in the linearly interpolated field and the error is generally lower (Fig. B.2g).

Page: 3

 Number: 1 Author: Subject: Cross-Out Date: 17.10.2025, 09:07:54
It's not qualitative if you're comparing by eye first here :)

 Number: 2 Author: Subject: Comment on Text Date: 17.10.2025, 09:04:18
Small comment but the order of your figures is a bit weird. I'd expect Fig 1 and 3 to follow each other and not have the method figure in between.

 Author: Subject: Comment on Text Date: 17.10.2025, 09:06:55
Also am I right in seeing that Fig 1 repeats things from Fig 3? I think having Fig 3 is good enough to not have any repetition.

 Number: 3 Author: Subject: Comment on Text Date: 17.10.2025, 09:09:33
This somehow repeats what you said above in the second paragraph. I would remove it from the second paragraph and keep it here only. Because the first time i read it I was a bit confused about why it did not show up in Fig 1 & 3 then.

 Number: 4 Author: Subject: Cross-Out Date: 17.10.2025, 09:10:26

 Number: 5 Author: Subject: Comment on Text Date: 17.10.2025, 09:11:36
Could you put the evaluation metrics in () here in the text so that we can compare the qualitative measures and see by how much it is better?

 Number: 6 Author: Subject: Comment on Text Date: 17.10.2025, 09:16:29
Again I think it's weird that Fig B2 comes in the text before Fig B1. You should probably change that ordering. Also for those two figures and Fig3, could you put the low resolution on the left and the higher resolutions together? If there is suddenly a low resolution image in the middle it's hard to compare the CM to the linear MAR and GT.

 Number: 7 Author: Subject: Comment on Text Date: 17.10.2025, 09:18:23

 Author: Subject: Comment on Text Date: 17.10.2025, 09:19:52
i might be wrong but not sure that came up before, except if it's in a figure legend? It should be in the text if there is an "again". Also do you have an idea on why that is? I i remember correctly, the same margin artifact problem was observed in van der Meer et al. 2023.

2.1.1 Generalizability to extreme climates

Our approach also permits downscaling of fields that are at the far end of the SMB distribution, that is, for high-emission scenarios at the end of this century with extremely negative SMB for most of Greenland. While the unconstrained model underestimates the melt or overestimates the SMB for most of Greenland (Fig. B.1b), the constrained version is able to realistically downscale the SMB field (Fig. B.1a). The MAE for one realization of this specific month is 45% lower for the constrained CM (MAE=50.2 mmWE) than for a simple linear interpolation (MAE=90.7 mmWE). Similarly, the correlation is higher for the constrained CM. However, the MAE of the unconstrained model (MAE=335.6 mmWE and $r = 0.93$) is substantially higher than even for the coarsened field (MAE=104.9 mmWE and $r = 0.93$). For the temperature fields, the linearly interpolated field shows the best metrics (Fig. B.1i). The MAE of the constrained CM is twice as large (MAE= 0.36 K) as for the linearly interpolated field (MAE= 0.18 K). This is due to artifacts in the downscaled fields outside the ice sheet interior (Fig. B.1f). However, it should be noted that the MAE is still very small compared to the absolute values of the temperature. The unconstrained version fails to generate realistic temperature fields (Fig. B.1g). By hard-constraining the CM, the downscaled fields approximately conserve the SMB even for very negative SMB (Fig. 5). The maximum absolute error in the monthly integrated SMB of the downscaled field over the whole ice sheet compared to the ground truth is 6.8 Gt. In contrast, linearly interpolating the coarsened fields leads to a clear overestimation of the SMB for very negative SMB fields (Fig. 5b). For comparison, the maximum absolute error of the integrated SMB of the linearly interpolated field is 86.6 Gt. For small noising scales ($t < 1$) the CM reproduces the hard boundaries of the pooled fields, which can be avoided by linearly interpolating the fields before feeding into the CM. However, since linear interpolation is not a conserving operation, this leads to a minor loss of information, which ultimately leads to worse quality of the downscaled fields compared to directly downscaling the pooled fields.

2.1.2 Determining optimal noising scale

To determine the optimal noising scale, which reasonably recovers the small-scale variability while also providing the smallest errors, we evaluate the power spectral density (PSD) (Fig. 4) and the continuous ranked probability score (CRPS) for different noising scales (Fig. 6). While the PSD is a measure of the variability on varying spatial scales, the CRPS is a metric that compares the ground truth to an ensemble of predictions. It can be seen as a probabilistic generalization of the mean absolute error. For this, we generate 50 different realizations for each month of the test set for different noising intensities and calculate the CRPS. We find that a noising time of $t = 5$ results in the lowest CRPS of 6.31 mmWE/month (Fig. 6b), while also ensuring agreement with the high resolution PSD (Fig. 4). For noising scales larger than $t = 10$, the CRPS converges to 6.34 mmWE/month (Fig. 6e-g). The CRPS is generally largest at the margins, where the SMB variability is the largest and lowest in the interior with little variability. Similarly, the CRPS for the surface temperature is lowest for $t = 5$ with CRPS= 0.1 K (Fig. 6g).

2.2 Downscaling NorESM

Generally, most ESMs do not directly provide a SMB field. However, the SMB can be approximated as the sum of precipitation, evaporation and runoff [18, 19]. Naturally, the CM is able to directly downscale from this first approximation of the SMB. However, the quality of the SMB field strongly depends on the ESM's ability to correctly simulate these variables. Most CMIP6 ESM's strongly overestimate the SMB or underestimate the melt in the higher emission scenarios. Hence, the SMB fields provided by the ESM's are not in accordance with the RCM simulations, even after bias correction (Fig. C.1a-d). The (bias-corrected) temperature, on the other hand, reproduces the overall warming trend shown by MAR reasonably well (Fig. C.1e-h). We compare the downscaled fields to the MAR fields for the same scenario. However, it has to be noted that the fields are not paired in the strict sense since MAR is driven by other variables and hence the spatial patterns do not match. For this downscaling exercise, we use the Norwegian ESM (NorESM2) under the SSP-5.85 forcing scenario, in order to show the general applicability of the approach. First, we bias correct the surface temperature as well as the SMB fields using Quantile Data Mapping (QDM) with the MAR fields driven by the reanalysis ERA5, using the historical period 1950-2015 as the reference period. These fields are then used as input into the constrained CM. While the CM is able to realistically downscale the SMB, the fields for the melting season at the end of the century underestimate the melt (Fig. C.1c). This follows from the QDM-corrected SMB fields which show the same behavior and are used as the input (Fig. C.1b). Per construction, the CM downscaling follows the large-scale patterns and hence cannot correct the low SMB if the input field show a low SMB bias. The MAE of the field downscaled by the CM is only slightly lower than that of the QDM-corrected field alone, but it effectively recovers the variability at small spatial scales (Fig. C.1&C.2). If the bias-corrected fields show very sharp and large transitions between individual grid cells or simply show very unrealistic spatial SMB fields, the conservation of the SMB can lead to wrong values, where they are not expected since the CM is trying to conserve the overall SMB of the locally defined region (superblock). In other words, the constrained CM cannot simply correct fields that are spatially incoherent. An example of this behavior is shown in Fig. C.3. The QDM-corrected field produces an unrealistically narrow accumulation area with positive SMB at the southeastern margin and a strong and abrupt gradient between

 Number: 1 Author: Subject: Comment on Text Date: 17.10.2025, 09:20:37

Can you start a new paragraph for temperature ? This paragraph is very very heavy with all the metrics.

 Author: Subject: Comment on Text Date: 17.10.2025, 09:23:46

It might also be a good idea to finish this section with a summary of its main point, as it's so heavy :)

negative and positive SMB (Fig. C.3b). While the CM downscales to a realistic SMB at the southeastern margin, the CM then is forced to compensate with a negative SMB west of the accumulation area due to the overall conservation of the SMB (Fig. C.3c). To avoid this behavior, the input field can be linearly interpolated beforehand or the superblock size has to be tuned accordingly.

The original temperature fields are substantially too cold (Fig. C.1e&C.2e), while the bias corrected fields show a similar temperature range compared to the MAR fields (Fig. C.1f&C.2f). This shows the importance of bias-correcting ESM fields for any further analysis. The CM fields show the smallest MAE with 0.68 K for a random warm month and 0.72 K for a random cold month. Over the whole time period, the MAE of the downscaled SMB fields is 38.9 mmWE and for the temperature 1.17 K, compared to the MAR fields forced by NorESM2. The largest differences are at the southeastern and western margins (Fig. C.4).

Since the SMB simulated by the NorESM underestimates the melt at the end of this century, we show an additional approach using a simple SMB parameterisation. Specifically, we show how our approach can be used in a *hybrid* manner by (i) using a physically-motivated model to get a low-resolution estimate of the SMB and (ii) run the CM to get a high-resolution field from this estimate. We run a simple offline PDD model with the bias corrected NorESM2 (SSP-5.85) temperature field and the precipitation field as input to get a very simple estimate of the SMB. For simplicity, we do not bias correct the precipitation field before calculating the SMB, which might improve the SMB further.

We find a MAE of 33.3 mmWE over the whole time period and the MAE of the temperature remains 1.17 K (Fig. C.5a,b). For comparison, the MAE of the PDD-derived SMB field is slightly higher with 34.6 mmWE and 1.2 K for the temperature (Fig. C.5c,d). Again, the margins show the largest error and the CM is able to recover the small-scale variability. While the PDD approach does not substantially improve the performance compared to the QDM approach, it shows that the CM can be used in a hybrid manner, i.e., using it in combination with a physics-based model.

3 Discussion

Our approach is built as a conditional generative model, not a classical regression, that learns to sample from a distribution of realistic high-resolution SMB fields given coarse inputs rather than to deterministically predict the SMB from external input such as the temperature or precipitation. However, similar to a regression, our recovered SMB and T_s fields strongly depend on the quality of the low-resolution input fields, that is, the ability of the ESM or physics-based model to accurately simulate the SMB and T_s at the coarse scale. We tried including the temperature as additional input to the CM during training (i.e., predictor) for the SMB (not shown). While we were able to generate realistic samples, the warming trend and interannual variations were underestimated in future emission scenarios.

A more promising alternative approach would be to separate the regression, i.e. predicting SMB from ESM variables, and downscaling tasks [44]. For example, a U-net could be trained to learn a regression from several input variables to give a first low-resolution estimate of the SMB with subsequent downscaling using the CM. Depending on the chosen regressor variables such an approach would allow a coupling with an external ice sheet model and therefore a first prototype of a hybrid SMB-ice sheet model. The CM model we present here is therefore a first step in this direction.

While our approach to constraining the model via mass conservation generally assumes some confidence in the overall low-resolution SMB fields, it allows generalization to more extreme or even out-of-sample SMB fields without the need of retraining the CM. This could be used to downscale regional climate models when running on a high-resolution grid is too expensive or time consuming. The subsequent downscaling step with the CM ensures conservation of the overall SMB fields and trend while providing realistic high-resolution fields at the same time. However, we also find that the CM reaches its limit beyond a downscaling factor of 32. Alternatively, our model can be used to generate unconditional samples that represent constant present-day forcing with the learned natural variability. This can be used for, e.g., spinup simulations, when a constant climate with variability is desirable.

Due to the nature of data-driven approaches, our results are strongly dependent on the quality of the training data. Here, we only use the regional climate model MAR, however, additional regional climate models such as RACMO [11] are available. It has been shown, that these show variations by a factor of two in future projections of the SMB [45]. Naturally, our downscaled fields are therefore biased towards the MAR projections. Nonetheless, our constraining approach overcomes this problem to some extent by enforcing approximate conservation of the regional SMB. The hard constraining is especially useful in the context of embedding downscaling approaches within existing process-based models such as ESMs. In ESMs, conservation of, for example, water, mass and energy is important to enforce physical consistency between the individual components. Our constraining approach ensures conservation of these variables, while allowing higher resolution. It should be noted, that our constraining approach is relatively simple and can be improved by, for example, incorporating the constraint layer directly into the network architecture, using a different

This page contains no comments

constraint approach (e.g., multiplicative constraints) [46] or by simply improving our region-growing algorithm, which might fail for some special cases.

Unlike precipitation, which is defined (even if zero) everywhere and can be treated as a spatially continuous field, surface mass balance is strictly confined to the ice-sheet mask. Downscaling SMB, therefore, must handle sharp domain boundaries at the ice edge, where values abruptly drop to zero, rather than the smooth transitions typical of precipitation fields. This leads to several technical challenges on how to treat the outside of the ice sheet in spatially-aware generative models. We decided to simply treat the SMB outside the ice sheet as zero. In the case of a non-evolving topography, as given by the training data, this is the most natural choice. However, this also means that our model is currently restricted to ice sheet geometries that do not extend beyond present-day extent due to the nature of the training data. With increasing availability of high-resolution RCM simulations with evolving topography this problem might be solved [47]. Similarly, the resolution of the output field is mostly limited by the training data. Training data with even higher resolution would also allow to generate even higher resolution downscaled fields. However, for very high resolutions, the training might be limited by the available GPU memory.

4 Conclusion

This study introduces a physics-constrained generative machine learning approach, based on consistency models to downscale the spatial SMB and T_s fields of the GrIS. We train our model on the regional climate model MARv3.12 and show that our approach is able to downscale the SMB and T_s fields by a factor of up to 32, from 160 km resolution to 5 km resolution in an efficient and realistic manner. The resulting fields are visually indistinguishable from the ground truth and the CM is able to successfully recover the small-scale variability of the SMB and T_s fields. Furthermore, we show how our model can be used to directly derive high-resolution SMB and T_s fields from the ESM by combining the CM with either (i) a bias-correction method (QDM) or (ii) a physics-based method (PDD).

Our approach is a first step towards a more general machine-learning driven SMB model. An application of our approach to Antarctica is straightforward but likely needs retraining and corresponding high-quality training data. Alternatively, our approach could be integrated into ESMs either to downscale SMB fields provided by the land component or to directly downscale precipitation fields. Usually, the resolution of the precipitation fields generated by the atmospheric component is several times lower than the native resolution of the ice sheet model. Hence, the precipitation does not resolve the topography of the ice sheets at all, which in turn leads to unrealistic spatial precipitation distribution and ultimately SMB. Using the CM with hard-constraints to downscale the precipitation fields during runtime could solve this problem. Overall, generative modeling is a promising approach to downscale arbitrary climate fields in computationally fast manner.

5 Methods

5.1 Consistency model

We use a consistency model architecture [43], recently introduced for downscaling [26] to derive high-resolution SMB and temperature fields. CMs are inspired by iterative diffusion models [48, 49], but offer several advantages. The idea behind CM is to learn a direct mapping from noisy inputs to clean data such that any point along the same noise trajectory (for the same underlying data) maps to the same clean state [43]. In contrast to previously introduced diffusion models, which need many evaluation steps (usually in the range of 10-2000) to generate realistic samples, CM can therefore generate realistic samples in only one step. Consistency models learn a consistency function $f(x(t), t, y) = x(t_{\min})$, which is self consistent:

$$f(\mathbf{x}(t), t, \mathbf{y}) = f(\mathbf{x}(t'), t', \mathbf{y}) \forall t, t' \in [t_{\min}, t_{\max}], \quad (1)$$

with the time-independent conditioning fields $\mathbf{y}$, the input fields $\mathbf{x}(t)$, i.e., the T_s and SMB and the time t . That is, the model's output is invariant to the noise level. Following Song et al. (2023) and Hess et al. [26, 43], we set $t_{\min} = 0.002$ and $t_{\max} = 80$. We impose the boundary condition $f(x(t_{\min}), t_{\min}) = x(t_{\min})$ for $t = t_{\min}$. This is parameterised in the model via skip connections

$$f(\mathbf{x}, t, \mathbf{y}; \theta) = c_{\text{skip}}(t)\mathbf{x} + c_{\text{out}}(t)F(\mathbf{x}, t, \mathbf{y}; \theta), \quad (2)$$

where $F(\cdot)$ is a U-Net with parameter θ .

The parameters c_{skip} and c_{out} are defined as

$$c_{\text{skip}} = \frac{\sigma_{\text{data}}^2}{((t - t_{\min})^2 + \sigma_{\text{data}}^2)}, \quad c_{\text{out}}(t) = \frac{\sigma_{\text{data}} t}{\sqrt{t^2 + \sigma_{\text{data}}^2}}. \quad (3)$$

This page contains no comments

The consistency training objective is given by [43]

$$\mathcal{L}(\theta, \bar{\theta}) = \mathbb{E}_{\mathbf{x}, n, t_n} [d(f(\mathbf{x} + t_{n+1}\mathbf{z}, t_{n+1}, \mathbf{y}; \theta), f(\mathbf{x} + t_n\mathbf{z}, t_n, \mathbf{y}; \bar{\theta}))]. \quad (4)$$

The distance measure $d(\cdot, \cdot)$ is given by $d(\mathbf{x}, \mathbf{y}) = \text{LPIPS}(\mathbf{x}, \mathbf{y}) + \|\mathbf{x} - \mathbf{y}\|_1$ [26, 43], where LPIPS is the Learned Perceptual Image Patch Similarity [50].

The training objective is minimized by stochastic gradient descent (Adam optimizer) on the model parameters θ , while updating $\bar{\theta}$ with an exponential moving average (EMA):

$$\bar{\theta} = \text{stopgrad}[w(k)\bar{\theta} + (1 - w(k)\theta)]. \quad (5)$$

The decay schedule $w(k)$ is given by

$$w(k) = \exp\left(\frac{s_0 \log(w_0)}{N(k)}\right), \quad (6)$$

with $w_0 = 0.9$. The EMA has been found to greatly stabilize training and improve final performance of the CM [43].

While, CMs can be distilled from diffusion models, we train the model from scratch. We use the same U-Net backbone as in Hess et al. (2025) [26]. We train for 200 epochs with a batch size of 4. Training on a single NVIDIA H100 Tensor Core GPU takes approximately 5 days.

While, CMs can generate samples in only one step, we use multi-step sampling (max. 10 steps) to enhance sample quality and avoid artifacts [43]. Starting from an initial noisy input, we progressively add scaled Gaussian noise at discrete time steps and denoise the field using the CM. One evaluation, i.e. downscaling one monthly field, for $t = 5$ with hard constraining takes ca. 1.6 s on a NVIDIA H100 Tensor Core GPU.

5.2 Bias Correction

We use Quantile Data Mapping (QDM) for bias correcting the ESM fields when used as input for the PDD or CM using the Python library xclim [51], which follows the algorithm introduced by Cannon et al. (2015) [52]. The idea of QDM is to correct systematic biases in modeled time series compared to a reference period of observed data or ground truth (here the RCM MAR) while conserving model-projected relative changes in quantiles [52]. It is applied for each spatial grid cell and variable individually. The QDM algorithm uses two steps. First, the future model output is detrended by quantile and bias corrected to the reference data by quantile mapping. Second, the absolute or relative change in quantiles is superimposed on the bias-corrected model output. The time-dependent cumulative density function (CDF) $\tau_{m,p}(t) = F_t(x_{m,p}(t))$ of the modeled projected time series $x_{m,p}$ over a time window is calculated (e.g., the future period). The relative change in quantiles between the modeled reference period ref and a future time t is given by

$$\Delta_m(t) = \frac{F_t^{-1}(\tau_{m,p}(t))}{F^{(-1)}(\tau_{m,\text{ref}}(t))} = \frac{x_{m,p}(t)}{F^{(-1)}(\tau_{m,\text{ref}}(t))} \quad (7)$$

with the inverse CDF F^{-1} . The modeled $\tau_{m,p}(t)$ quantile is then bias corrected by applying the inverse CDF derived from the observed values over the reference period

$$\hat{x}_{o:m,\text{ref}:p}(t) = F_{o,\text{ref}}^{(-1)}(\tau_{m,p}(t)). \quad (8)$$

Finally, the bias-corrected modeled projected time series at time $\hat{x}_{m,p}(t)$ is given by

$$\hat{x}_{m,p} = \hat{x}_{o:m,\text{ref}:p}(t) + \Delta_m(t). \quad (9)$$

We use 100 quantiles and do a monthly-based bias correction.

5.3 Conservation Constraints

To allow the CM to downscale fields that are far out-of-sample, we implement the option to enforce a hard conservation constraint during the inference [46]. This does not require a retraining of the CM.

The initial conditions are pooled using adaptive pooling with a factor that should correspond to the resolution ratio between the low and high resolution fields. This gives a field with so-called superblocks, i.e., pooled pixels. The first denoising step is not changed and gives a first guess of the downscaled field. After the first denoising step, we add the residual between each superblock and the corresponding high-resolution pixels. That is, to each intermediate output

 Number: 1 Author: Subject: Comment on Text Date: 17.10.2025, 08:48:36

 Author: Subject: Comment on Text Date: 17.10.2025, 08:49:51

I believe this was not introduced. Maybe you confused the target θ and the approximation by the model $\tilde{\theta}$? Just explain both and make sure the functions are consistent :)

 Number: 2 Author: Subject: Comment on Text Date: 17.10.2025, 08:51:40

Here it would be nice to have a reminder of how many samples are in the training set to have an idea what this means for the training time on GPU :)

 Number: 3 Author: Subject: Comment on Text Date: 17.10.2025, 08:50:51

Is there a reason why you took such a small batch size ?

pixel during the multistep sampling $\tilde{y}_i$ the residual between the superblock x and the mean of the corresponding pixels $\frac{1}{n} \sum_{i=1}^n \tilde{y}_i$ is added:

$$y_j = \tilde{y}_j + x - \frac{1}{n} \sum_{i=1}^n \tilde{y}_i. \quad (10)$$

The residuals are corrected to account for the number of land pixels in each superblocks. That is, the residuals are only distributed onto the valid GrIS pixels and not into the ocean. In the last denoising step, we do not enforce this constraint. To avoid artifacts in the downscaled field due to the pooling step, we use the *area* interpolation and Gaussian filter on the residual field before adding the residual field in each step.

To avoid cases where all residuals of a superblock are distributed to only a few pixels of the GrIS, which would lead to unrealistic SMB values at the GrIS margins, we implement a simple greedy region-growing algorithm. Each superblock cell that has less than a chosen number of corresponding land pixels in the high-resolution field (we choose 16^2 pixels) annexes its largest neighboring blocks until it exceeds the chosen threshold. This continues until all superblocks are assigned a region with at least 16^2 pixels. We find that 16^2 - 32^2 pixels is a reasonable choice ensuring realistic fields and avoiding artifacts in the downscaled fields.

Since the conservation of the SMB should happen in physical units, we transform the correction field and intermediate fields back to physical units before we apply the constraint. Subsequently, we transform the fields back to the transformed units before running the next step of the CM.

5.4 PDD Model

We use a simple positive degree day (PDD) model based on the PyPDD implementation [53]. The PDD is a semi-physical method that takes the precipitation and the near-surface air temperature as input and estimates the surface mass balance. Specifically, the PDD are calculated according to [54, 55]

$$\text{PDD} = \int_0^A \left[\frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{T_{\text{ac}}(t)^2}{2\sigma^2}\right) + \frac{T_{\text{ac}}(t)}{2} \text{erfc}\left(-\frac{T_{\text{ac}}(t)}{\sqrt{2}\sigma}\right) \right] dt, \quad (11)$$

with the time interval A , the time-dependent seasonal near-surface temperature $T_{\text{ac}}(t)$, the standard deviation of the temperature $\sigma = 5$ K and the error function erfc . Following the standard conventions, we assume that all precipitation contributes to snow accumulation whenever air temperature $T \leq 0^\circ\text{C}$, and that the snow-fraction (hence accumulation) declines linearly to zero as T rises from 0 to 2°C . Melt is then calculated via positive degree-day factors of $3 \text{ mm}^\circ\text{C}^{-1}\text{d}^{-1}$ for snow and $8 \text{ mm}^\circ\text{C}^{-1}\text{d}^{-1}$ for ice. For more details we refer to Seguinot et al. (2013) [54].

5.5 Training

We use monthly output from the regional climate model Modèle Atmosphérique Régional (MAR) version 3.12 forced by different Earth system models (Tab. 1) over the time period 1950-2100 [9, 10]. In total we use 23 different model runs, totaling to 21432 monthly fields. We hold out 4 independent runs from MAR forced by the ESM NorESM2 for additional testing. We regrid the MAR fields to 5 km resolution, corresponding to the IMSIP6 grid [5]. Additionally, we mask the fields by the current ice sheet mask given by MAR, that is, we set the SMB to 0 outside of the current ice sheet geometry.

We condition our CM on the ice sheet height and the monthly mean insolation as additional input besides the SMB. This ensures that generated unconditional samples, i.e., samples without initial conditions, adhere to the topography of the GrIS. The insolation allows the CM to learn the seasonal melt cycle and generate unconditional samples that adhere to it. The input fields are simply concatenated before being fed to the U-Net.

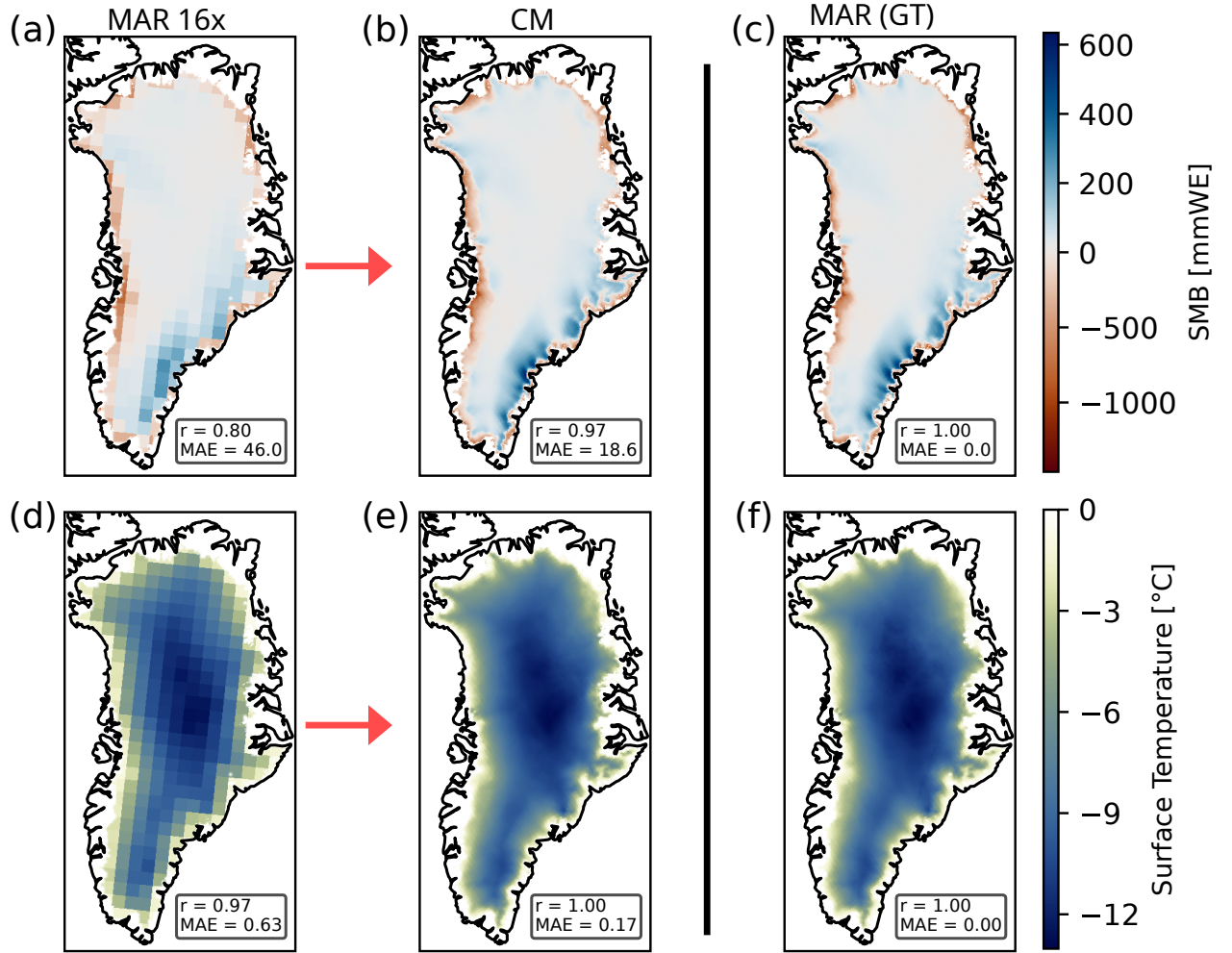


Figure 1: **Exemplary downscaling of surface mass balance and temperature at surface for random month from test set.** (a) Random month from held-out test set pooled with a factor of 16 (80 km) resolution. This serves as input to our model. (b) Downscaled SMB field with the CM model using hard constraints to 5 km resolution. The small scale variability is recovered, especially at the ice-sheet margins. (c) High-resolution ground truth (GT) for this individual month. The GT and the downscaled field are visually indistinguishable. (d, e, f) Same as (a, b, c) but for the temperature at surface.

Table 1: **Selected MARv3.12 runs for the training forced by CMIP5 and CMIP6 models.** All models used for training and the respective scenario are denoted. The MAR runs forced by NorESM are held out for validation.

Forcing (Model)	CMIP Scenarios (runs)
ACCESS1.3	hist, RCP-8.5
CESM2	hist, SSP-1.26, SSP-2.45, SSP-5.85
CNRM-CM6	SSP-5.85
CNRM-ESM2	hist, SSP-5.85
ERA-5	hist
IPSL-CM6A-LR	hist, SSP-5.85
MPI-ESM1-2-HR	hist, SSP-1.26, SSP-2.45, SSP-5.85
UKESM1-0-LL	hist, SSP-2.45, SSP-5.85
NorESM2	hist, SSP-1.26, SSP-2.45, SSP-5.85

This page contains no comments

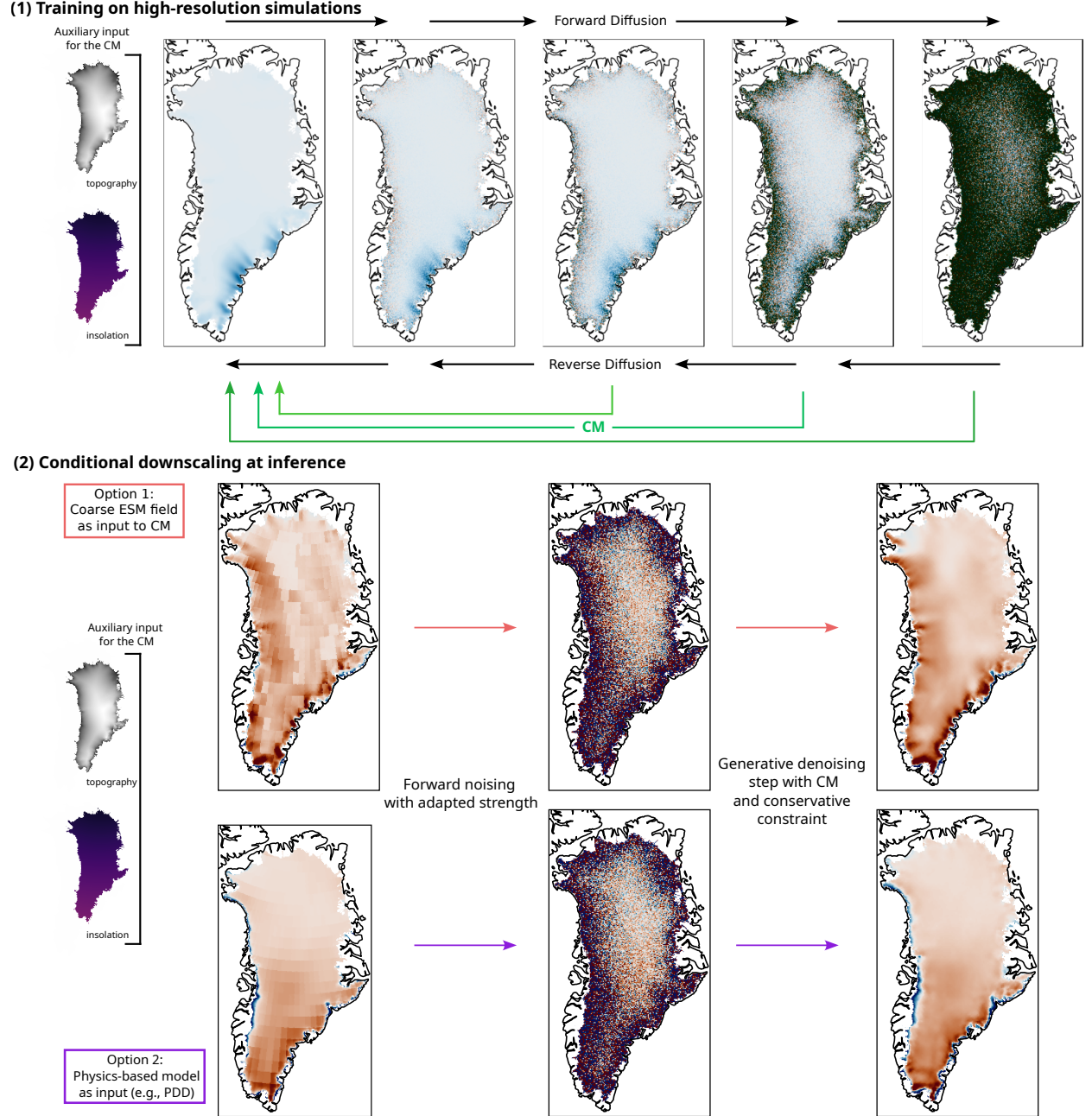


Figure 2: **Schematic depiction of the consistency model (CM) workflow.** We train the CM on the high-resolution RCM MARv3.12 with topography and insolation as auxiliary input. During inference, the low resolution input fields are noised with an adaptive noising strength and subsequently denoised by the trained CM with conservative constraints. The small-scale variability is recovered by the CM.

This page contains no comments

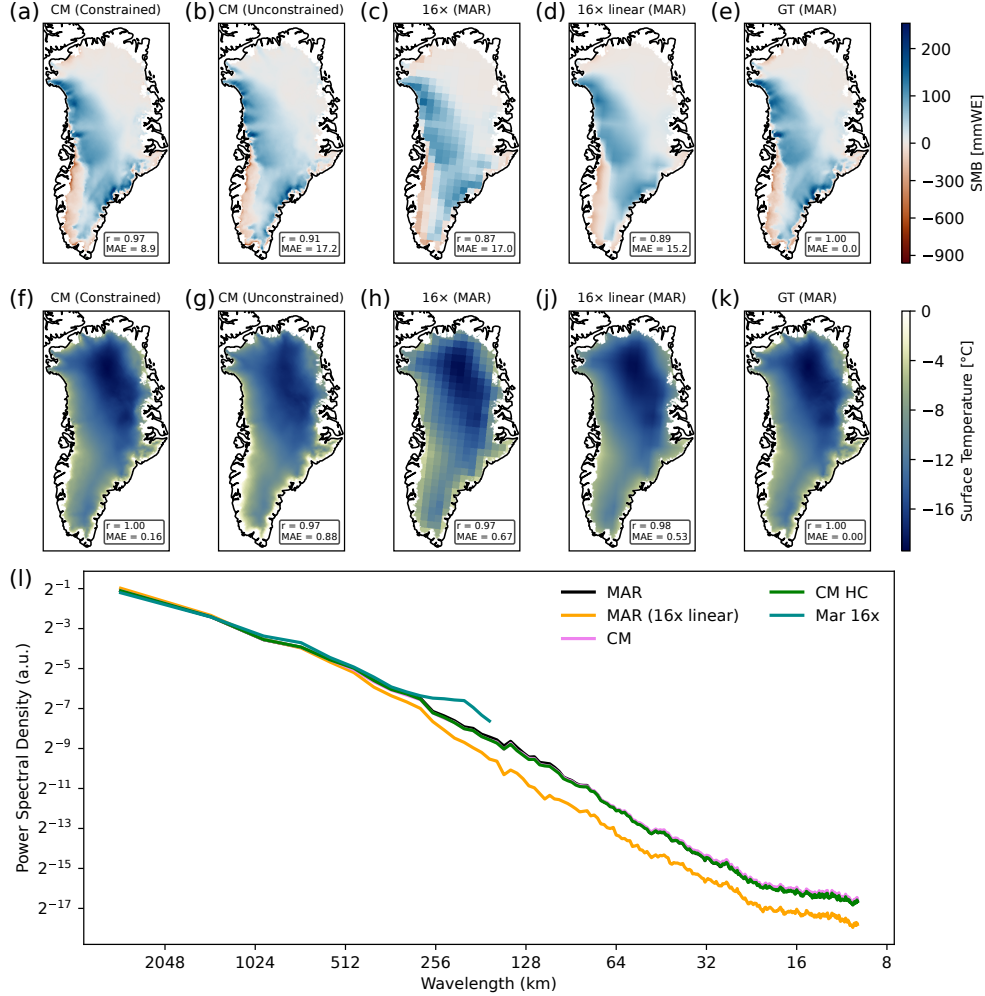


Figure 3: **Qualitative comparison of downscaled SMB and T_s fields.** (a) Downscaled SMB field of test set with hard constraints during inference for warm month with negative SMB at the margins (blue) with a 5 km resolution. The mean absolute error (in mmWE) and Pearson correlation with the GT field are denoted. (b) Same as (a) but unconstrained CM. (c) Coarsened MAR field by a factor of 16 (80 km resolution), which is the input for the CM. The small scale variability is lost. (d) Same as b but linearly interpolated. (e) Ground truth MAR field with 5 km resolution. The downscaled fields and GT are visually indistinguishable. (f-j) Same as a-e but for the surface temperature. (i) Radially averaged power spectral density of the whole test set for all depicted fields. The coarsened field has no small scale variability (blue). The linear interpolated field shows substantially lower small scale variability compared to the ground truth. The ground truth MAR field and CM fields almost show the same PSD.

This page contains no comments

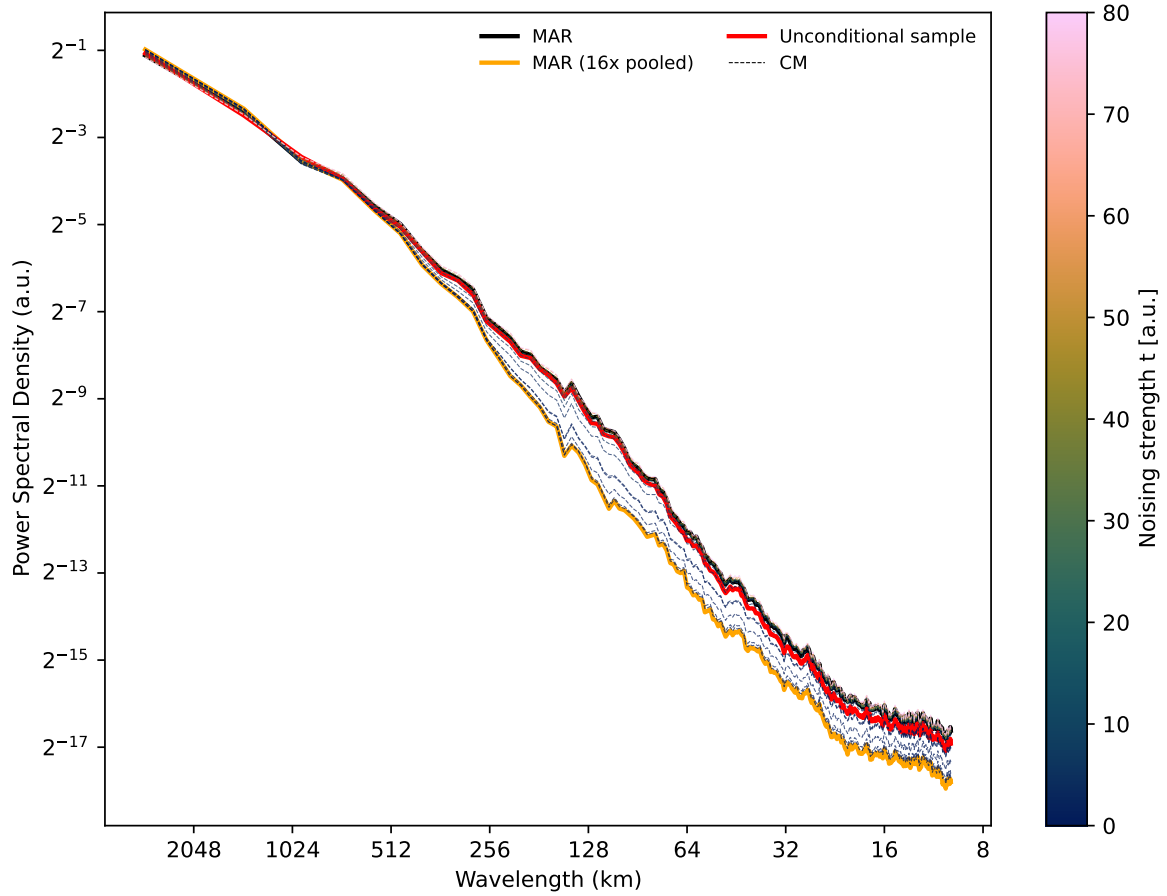


Figure 4: **Normalized power spectral density (PSD) for different noise scales.** The mean radially averaged normalized PSD for different noising scales is shown. We downscale the test set that was coarsened by a factor of 16 and linearly interpolated using the CM (unconstrained). Additionally, the PSD of an unconditional ensemble is shown (red). The coarsened field (orange) generally underestimates the small scale variability, i.e., the variability below 512 km. The higher the noising scale, the more the PSD of the downscaled fields resembles the PSD of the original field (black). Larger noise scales correspond to less pairing with the initial fields.

This page contains no comments

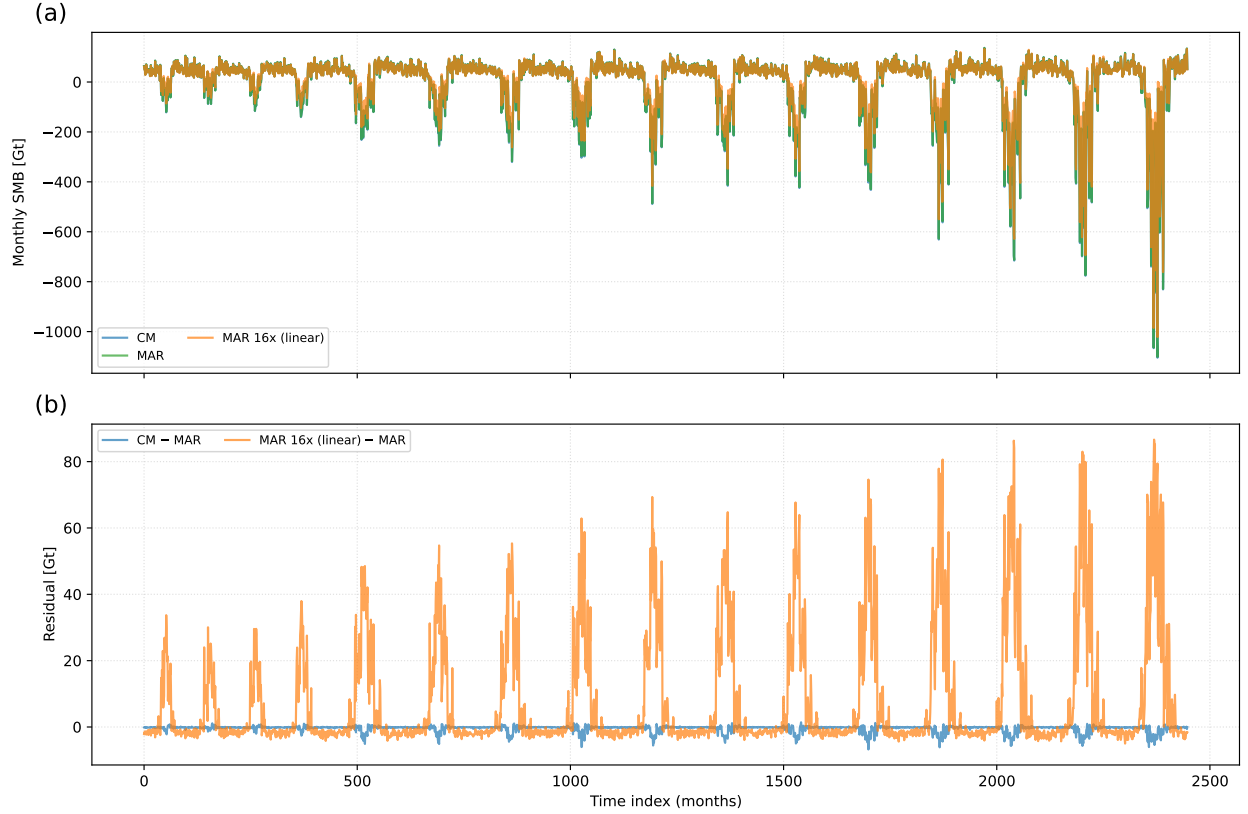


Figure 5: **Monthly integrated SMB over test set for linearly interpolated SMB and downscaled field.** (a) Monthly integrated SMB over whole test set for MAR, CM and linearly interpolated MAR field from the coarsened MAR field by a factor of 16. (b) Residuals between downscaled field and ground truth MAR and linearly interpolated SMB field and MAR. The linearly interpolated fields overestimates the SMB compared to MAR during the melting seasons. The downscaled fields approximately conserve the SMB even for strong melt.

This page contains no comments

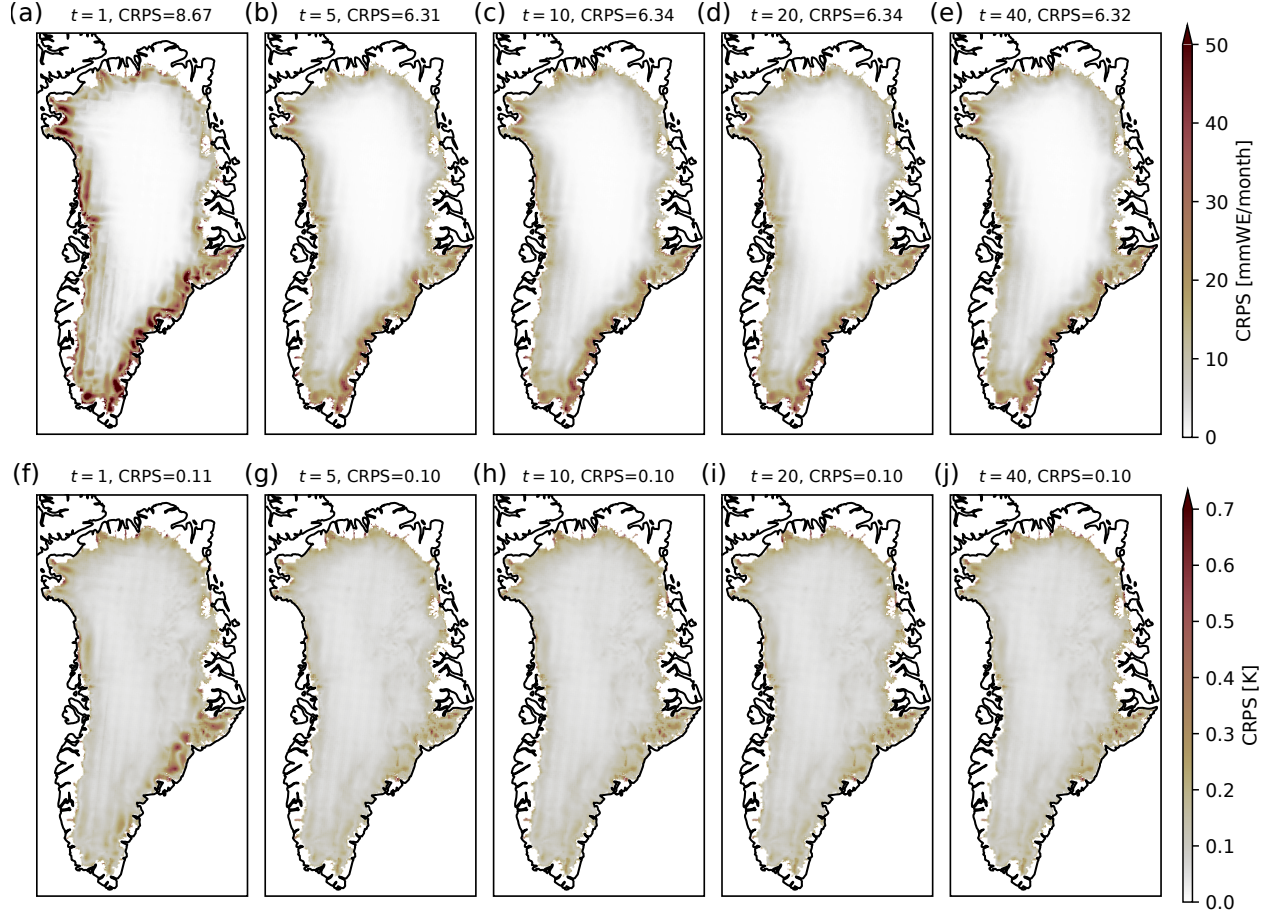


Figure 6: **Continuous ranked probability scores (CRPS) for different noise scales over the test set for SMB and T_s .** (a) Spatial mean CRPS for the noising time $t = 1$. For each month of the test set, an ensemble of 50 realizations is generated and compared to the corresponding ground truth. The spatiotemporal mean CRPS is 8.67 mmWE/month. The CRPS is largest at the margins of the ice sheet, where the variability is largest. (b,c,d,e) Same as a but for different noising times. The noising time $t = 5$ leads to the lowest CRPS. (f-j) Same as a-e but for the surface temperature. The CRPS converges to 0.1 K for $t > 5$.

This page contains no comments

Author Contributions: NB conceived and designed the study with input from PH and AR. NB and PH wrote the code. NB analyzed all data. All authors interpreted the results. NB wrote the manuscript with input from AR and PH.

Code and Data Availability: The original CM code is available at <https://github.com/p-hss/consistency-climate-downscaling>. Our modified version for Greenland as well as the training and test data will be made available upon publication.

Competing Interests: The authors declare no competing interests.

Acknowledgments: This work was supported by the UiT Aurora Centre Program, UiT - The Arctic University of Norway (2020), and the Research Council of Norway (project number 314570). This is ClimTip contribution #80; the ClimTip project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101137601. Alexander Robinson received funding from the European Union (ERC, FORCLIMA; grant no. 101044247). The authors gratefully acknowledge the Ministry of Research, Science and Culture (MWFK) of Land Brandenburg for supporting this project by providing resources on the high performance computer system at the Potsdam Institute for Climate Impact Research. The colormaps for the plots are taken from [56].

This page contains no comments

Appendix

A Unconditional samples

To show that our models correctly learn the distribution of the underlying training data, we generate 100 unconditional SMB samples for each month and compare them to the training set (Fig. A.1). While, the SMB modeled by MAR is not a ground truth in the classical sense, we treat it as such in lack of observational SMB fields and refer to it as such in the following. Here, we refer to unconditional samples as samples that are generated from noise, i.e., without any initial guesses of the SMB. However, these samples are conditioned on the auxiliary inputs, i.e., the ice sheet height and insolation.

The CM is able to generate realistic unconditional SMB field samples, both for the winter and summer season (Fig. A.1a&b). The mean and standard deviation of the generated set closely follow the mean and standard deviation of the training set, except for the summer season (Fig. A.1h-j). In all summer months (JJA), the average SMB is greater than the corresponding training set mean SMB. Similarly, the standard deviation of the generated ensemble is smaller. This is almost exclusively due to the greater SMB at the margins in the generated ensemble. In other words, the generated ensemble underestimates the melt and variability of the SMB at the ice-sheet margins in the summer months. Interestingly, the generated ensemble mean SMB in May is smaller and the variability is greater than that of the training set (Fig. A.1g).

Generally, we are not interested in generating unconditional samples. Therefore, the overestimation of the SMB at the margins in the summer months is not a problem for the downscaling task per-se. We overcome this problem by introducing a hard constraint that conserves the SMB from the low-resolution fields on a regional scale. We find, that the hard-constraining generally improves the results. Alternatively, oversampling of the summer months during training could improve the unconditional samples for these months.

This page contains no comments

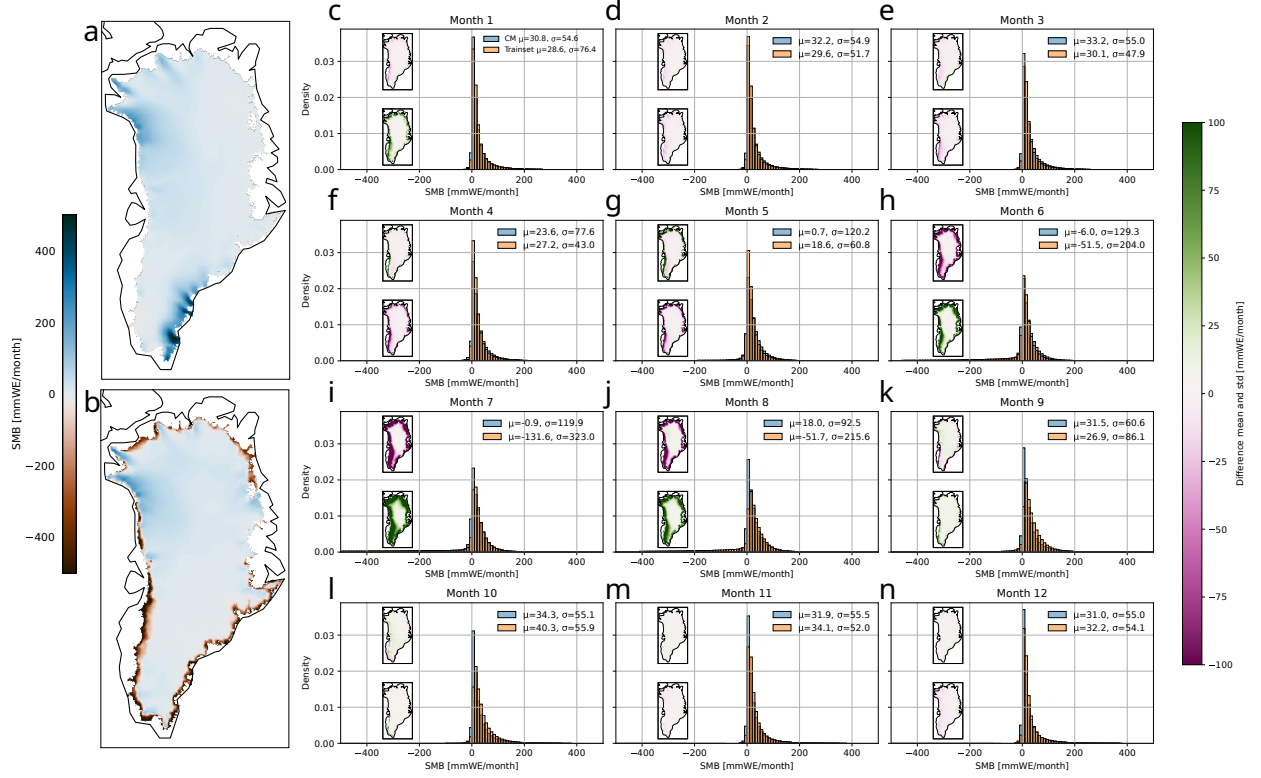


Figure A.1: **Unconditional sampling of SMB from trained CM.**(a) Exemplary unconditional sample of cold month with positive SMB. (b) Exemplary unconditional sample of warm month with negative SMB at the margins. (c) Histogram of generated SMB samples and training set for January. The mean and standard deviation of each set are denoted. The inset shows the spatial difference of the mean (top) and standard deviation (bottom) between generated samples and the training set. Green corresponds to areas where the CM samples have smaller mean/standard deviation, while pink areas denote regions where the CM has larger mean/standard deviation than the training set. **d-n** Same as **a** but for all other months of the year. The CM is mostly reproducing the mean and standard deviation well but overestimates the SMB in the summer months with strongly negative SMB.

This page contains no comments

B Examples of downscaled fields

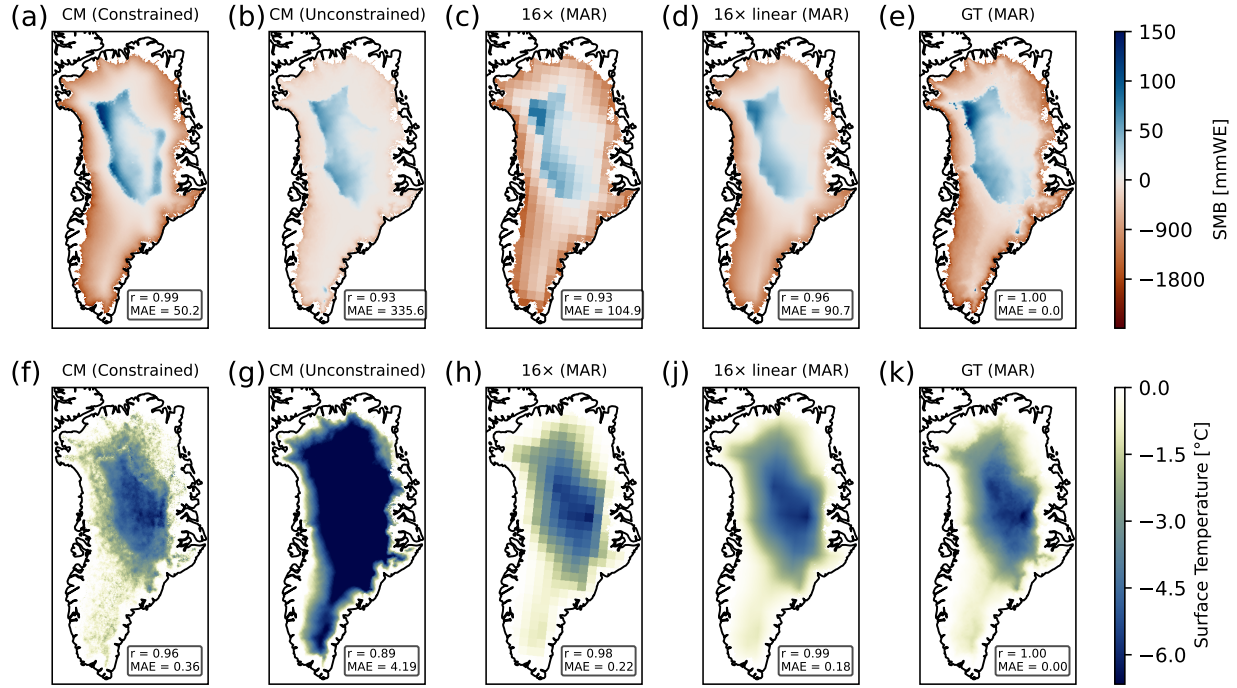


Figure B.1: **Exemplary downscaling of surface mass balance and temperature at surface for random month from test set.** (a) Random month from held-out test set pooled with a factor of 16 (80 km) resolution. This serves as input to our model. (b) Downscaled SMB field with the CM model using hard constraints to 5 km resolution. The small scale variability is recovered, especially at the ice-sheet margins. (c) High-resolution ground truth (GT) for this individual month. The GT and the downscaled field are visually indistinguishable. (d, e, f) Same as (a, b, c) but for the temperature at surface.

This page contains no comments

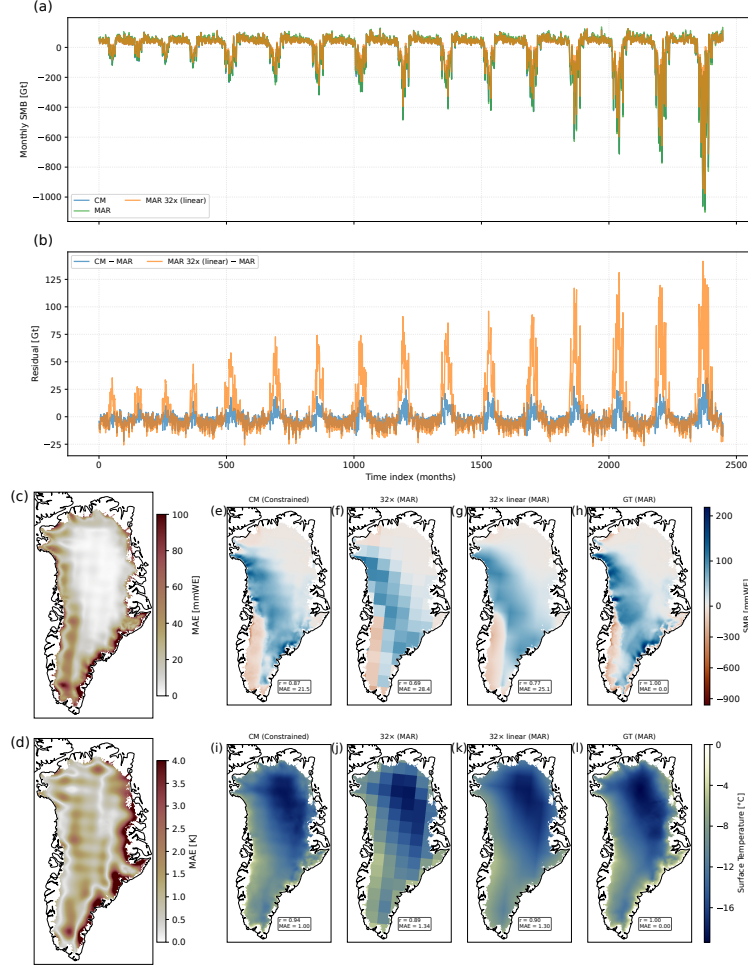


Figure B.2: **Downscaling of coarsened MAR test set by factor 32.** (a) Monthly integrated SMB over whole test set for MAR, CM and linearly interpolated MAR field from the coarsened MAR field by a factor of 32. (b) Residuals between downscaled field and ground truth MAR and linearly interpolated SMB field and MAR. The linearly interpolated fields overestimates the SMB compared to MAR during the melting seasons. While the downscaled fields also over- and underestimate the SMB they show substantially smaller residuals compared to the linearly interpolated fields. (c) MAE of downscaled SMB fields over whole test set. The coarse structure of the input field is still partially visible in the MAE field. The average MAE over the whole test set is 24.6 mmWE. (d) Same as c but for surface temperature. The MAE over the whole test set is 1.35 K. (e) Downscaled SMB field of test set with hard constraints during inference for warm month with negative SMB at the margins (blue) with a 5 km resolution. The mean absolute error (in mmWE) and Pearson correlation with the GT field are denoted. (f) Coarsened MAR field by a factor of 32 (160 km resolution), which is the input for the CM. The small scale variability is lost. (g) Same as f but linearly interpolated. (h) Ground truth MAR field with 5 km resolution. (i-j) Same as e-h but for the surface temperature.

This page contains no comments

C NorESM2 downscaling

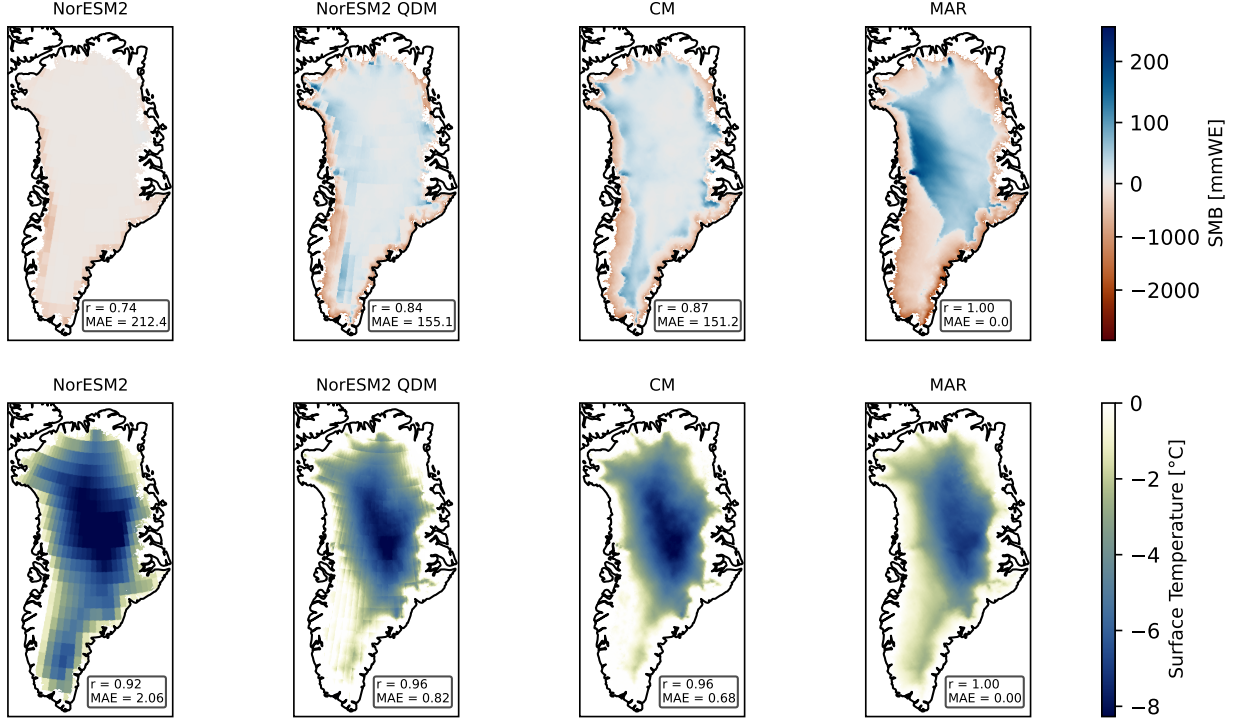


Figure C.1: **Exemplary downscaling of surface mass balance and temperature at surface for end of century month for NorESM2 SMB.** (a) NorESM2 SMB as sum of precipitation, melt and runoff. The SMB does not show any strong negative SMB at the margins. The MAE and correlation compared to the MAR SMB for the same month are denoted. (b) QDM-corrected NorESM2 SMB. The margins show a more negative SMB compared to non-bias corrected SMB. However, the SMB is still overestimated compared to the MAR SMB. c Downscaled SMB field with the CM model using hard constraints to 5 km resolution. d Unpaired high-resolution ground truth (GT) for this individual month as simulated by MAR. (e, f, g, h) Same as (a, b, c, d) but for the temperature at surface. The bias-corrected and downscaled T_s show stronger similarity with MAR than the SMB fields.

This page contains no comments

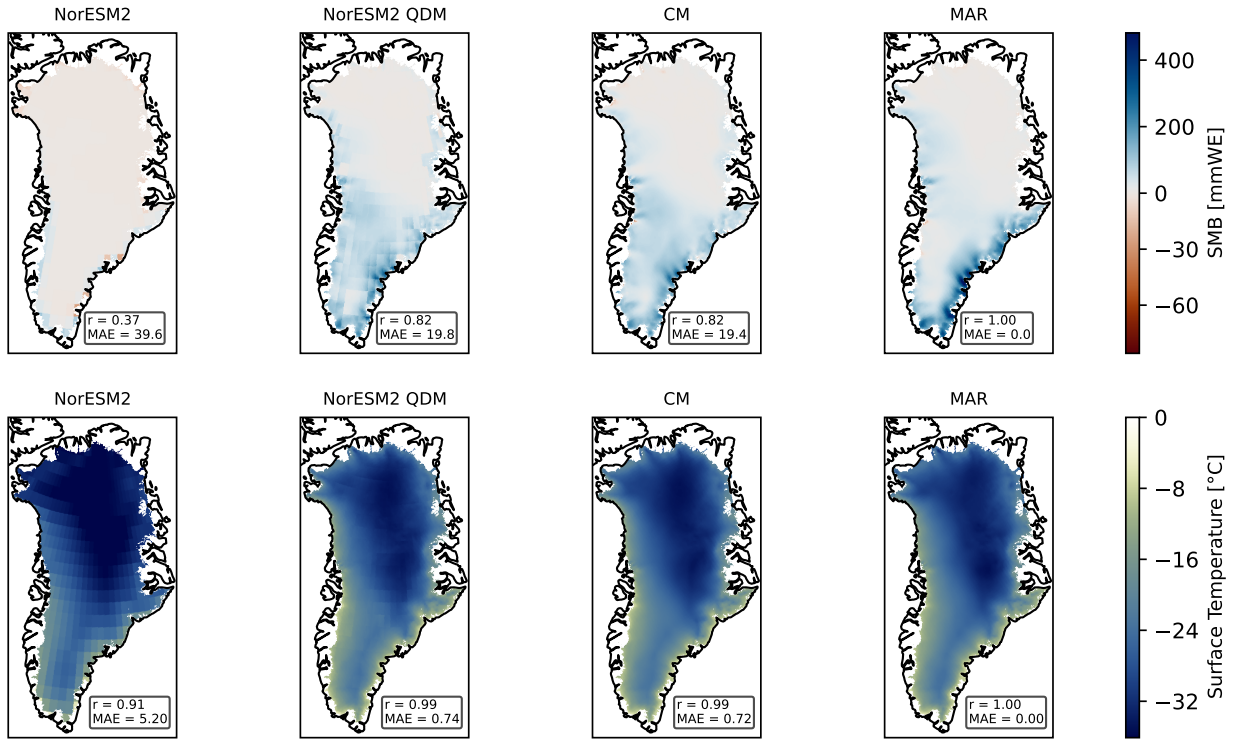


Figure C.2: **Exemplary downscaling of surface mass balance and temperature at surface for random month from test set.** (a) Random month from held-out test set pooled with a factor of 16 (80 km) resolution. This serves as input to our model. (b) Downscaled SMB field with the CM model using hard constraints to 5 km resolution. The small scale variability is recovered, especially at the ice-sheet margins. (c) High-resolution ground truth (GT) for this individual month. The GT and the downscaled field are visually indistinguishable. (d, e, f) Same as (a, b, c) but for the temperature at surface.

This page contains no comments

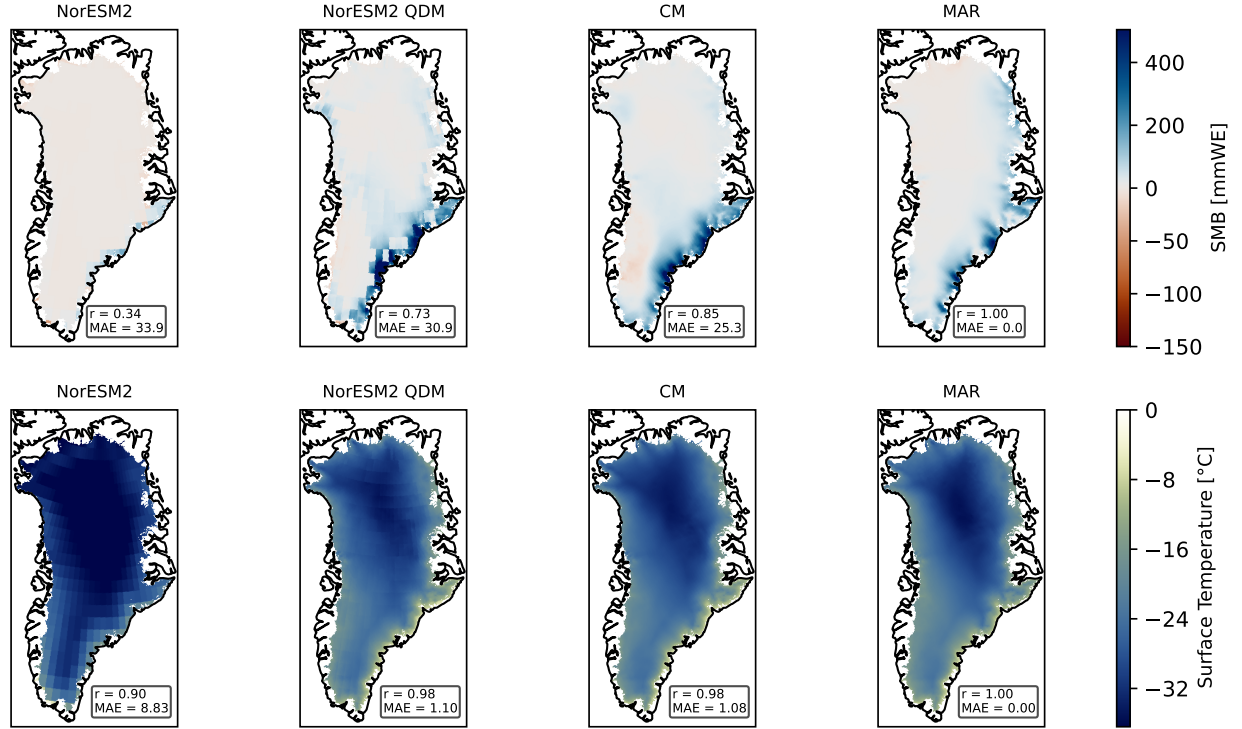


Figure C.3: **Exemplary downscaling of surface mass balance and temperature at surface for random month from test set.** (a) Random month from held-out test set pooled with a factor of 16 (80 km) resolution. This serves as input to our model. (b) Downscaled SMB field with the CM model using hard constraints to 5 km resolution. The small scale variability is recovered, especially at the ice-sheet margins. (c) High-resolution ground truth (GT) for this individual month. The GT and the downscaled field are visually indistinguishable. (d, e, f) Same as (a, b, c) but for the temperature at surface.

This page contains no comments

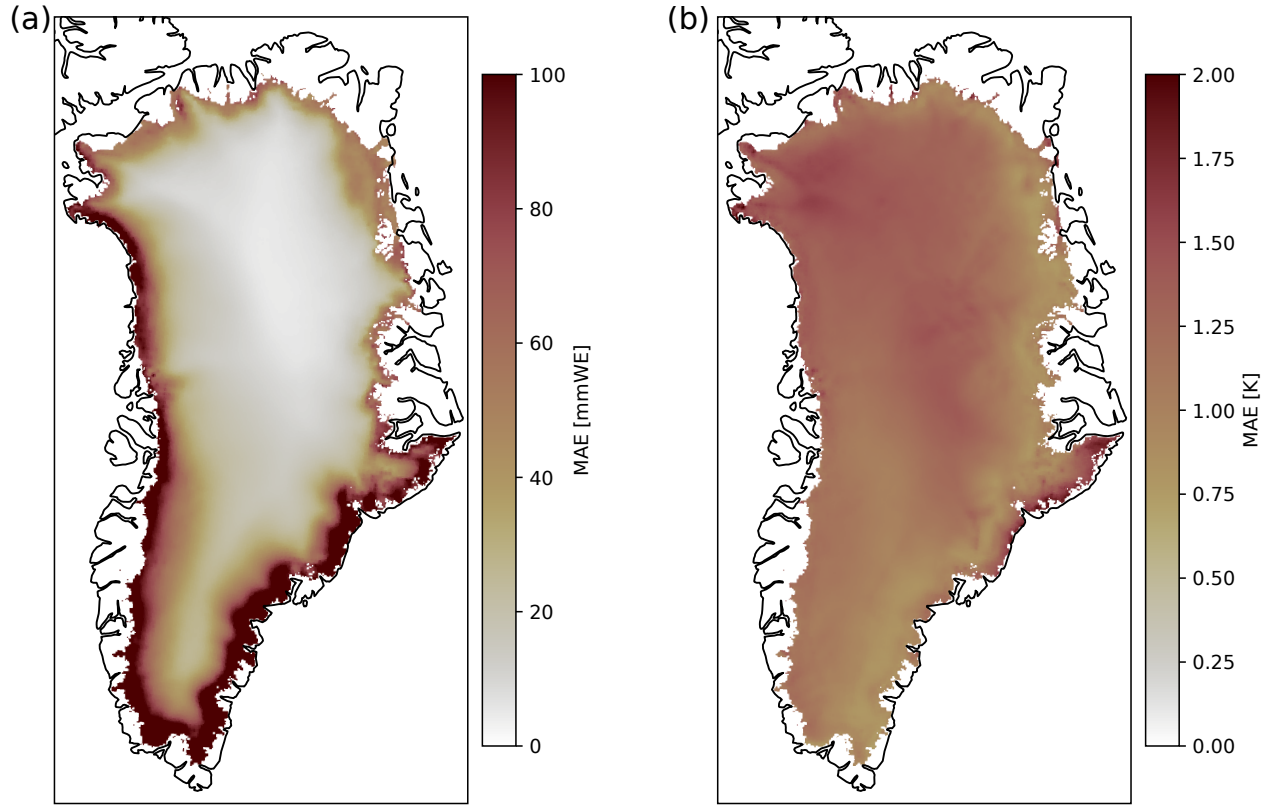


Figure C.4: **Mean absolute error over whole time period (2015-2100) for QDM-corrected NorESM2 SMB for SSP-5.85 scenario.** (a) Temporal MAE for each grid cell of downscaled SMB from bias-corrected SMB field derived from NorESM2 SSP-5.85 run. The margins show the largest error. (b) Same as a but for surface temperature.

This page contains no comments

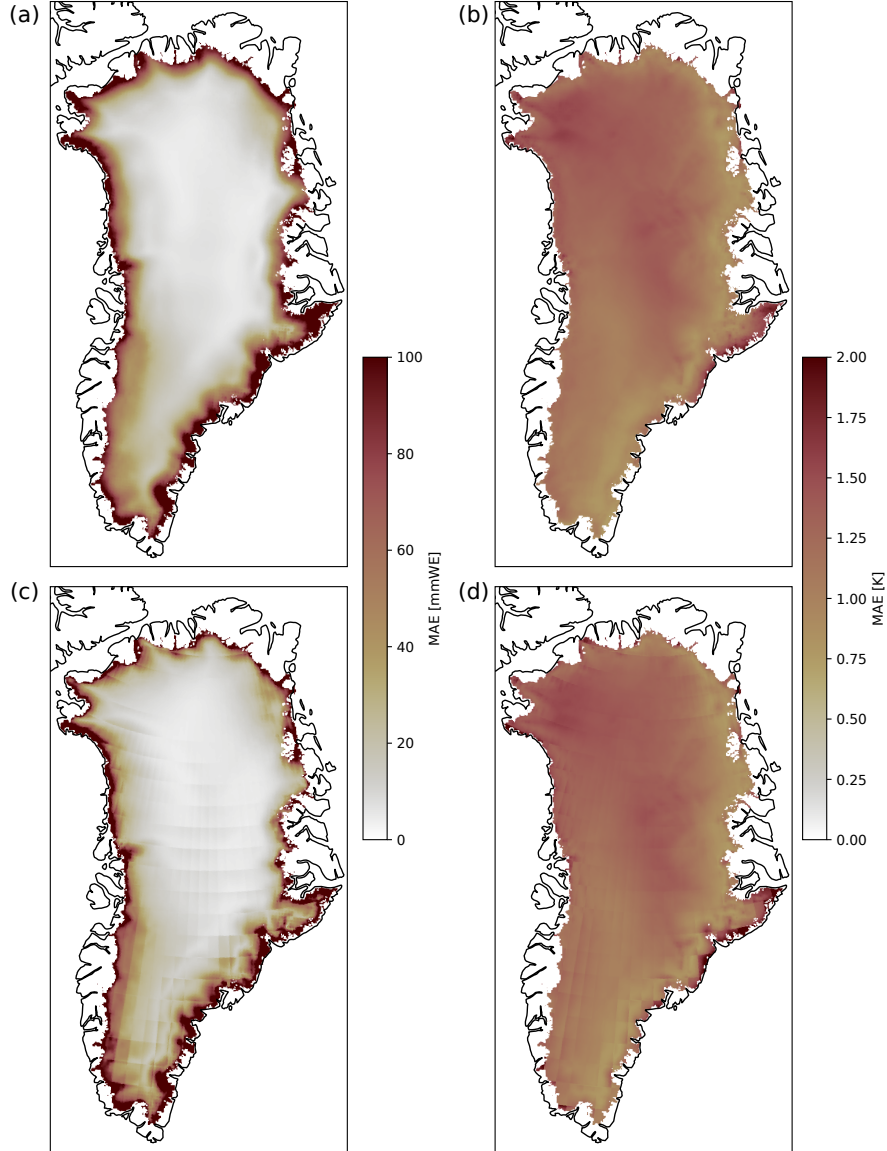


Figure C.5: **Mean absolute error over whole time period (2015-2100) for PDD NorESM2 SMB for SSP-5.85 scenario.** (a) Temporal MAE for each grid cell of downscaled SMB from bias-corrected PDD derived SMB field forced by NorESM2 SSP-5.85 run. The margins show the largest error. (b) Same as a but for surface temperature. (c,d) Same as a,b but directly for bias corrected SMB and T_s fields without downscaling.

This page contains no comments

D Training and validation loss

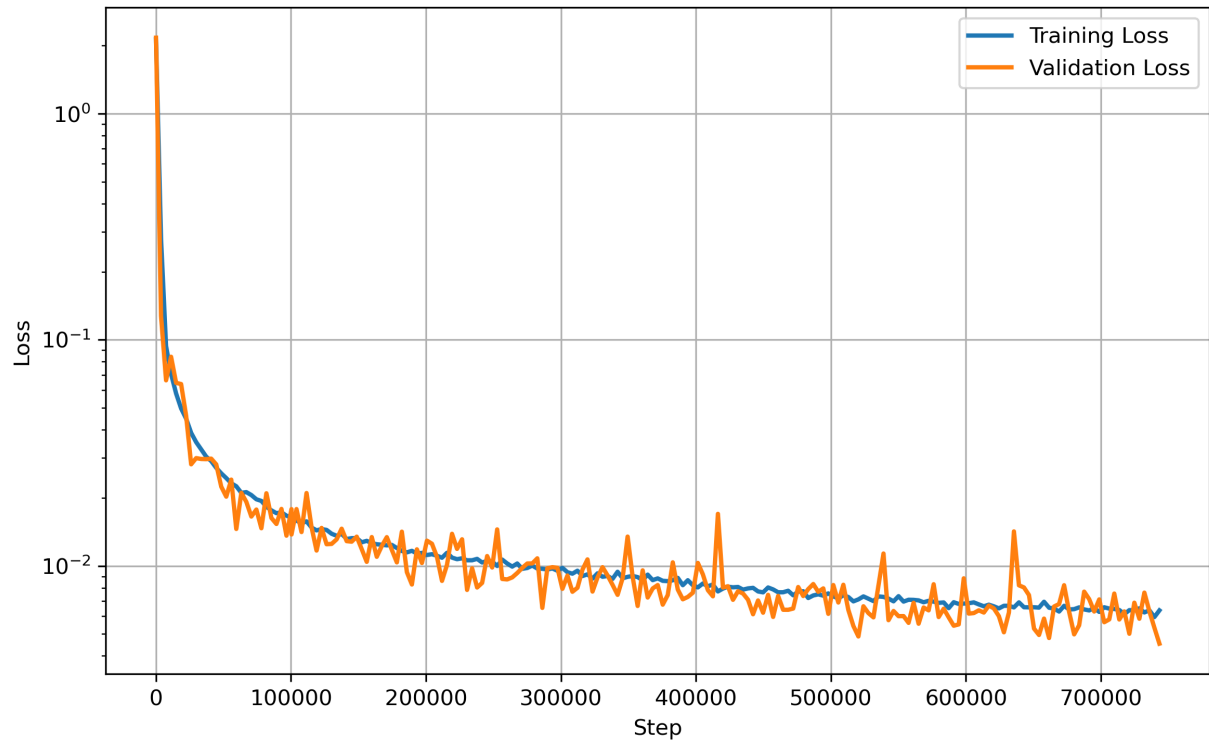


Figure D.1: Validation and training loss of consistency model.

This page contains no comments

References

1. Morlighem, M. *et al.* BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation. en. *Geophysical Research Letters* **44**, 11, 051–11, 061. ISSN: 1944-8007. doi:10.1002/2017GL074954 (2017).
2. Trusel, L. D. *et al.* Nonlinear rise in Greenland runoff in response to post-industrial Arctic warming. en. *Nature* **564**, 104–108. ISSN: 1476-4687. doi:10.1038/s41586-018-0752-4 (Dec. 2018).
3. Van den Broeke, M. R. *et al.* On the recent contribution of the Greenland ice sheet to sea level change. English. *The Cryosphere* **10**, 1933–1946. ISSN: 1994-0416. doi:10.5194/tc-10-1933-2016 (Sept. 2016).
4. Choi, Y., Morlighem, M., Rignot, E. & Wood, M. Ice dynamics will remain a primary driver of Greenland ice sheet mass loss over the next century. en. *Communications Earth & Environment* **2**, 1–9. ISSN: 2662-4435. doi:10.1038/s43247-021-00092-z (Feb. 2021).
5. Goelzer, H. *et al.* The future sea-level contribution of the Greenland ice sheet: a multi-model ensemble study of ISMIP6. English. *The Cryosphere* **14**, 3071–3096. ISSN: 1994-0416. doi:10.5194/tc-14-3071-2020 (Sept. 2020).
6. Payne, A. J. *et al.* Future Sea Level Change Under Coupled Model Intercomparison Project Phase 5 and Phase 6 Scenarios From the Greenland and Antarctic Ice Sheets. en. *Geophysical Research Letters* **48**, e2020GL091741. ISSN: 1944-8007. doi:10.1029/2020GL091741 (2021).
7. Lucas-Picher, P. *et al.* Very high resolution regional climate model simulations over Greenland: Identifying added value. en. *Journal of Geophysical Research: Atmospheres* **117**. ISSN: 2156-2202. doi:10.1029/2011JD016267 (2012).
8. Mottram, R. *et al.* Surface mass balance of the Greenland ice sheet in the regional climate model HIRHAM5: Present state and future prospects. *Low Temperature Science* **75**, 105–115 (Mar. 2017).
9. Fettweis, X. *et al.* Estimating the Greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model MAR. English. *The Cryosphere* **7**, 469–489. ISSN: 1994-0416. doi:10.5194/tc-7-469-2013 (Mar. 2013).
10. Fettweis, X. *et al.* Reconstructions of the 1900–2015 Greenland ice sheet surface mass balance using the regional climate MAR model. English. *The Cryosphere* **11**, 1015–1033. ISSN: 1994-0416. doi:10.5194/tc-11-1015-2017 (Apr. 2017).
11. Van Dalum, C. T. *et al.* First results of the polar regional climate model RACMO2.4. English. *The Cryosphere* **18**, 4065–4088. ISSN: 1994-0416. doi:10.5194/tc-18-4065-2024 (Sept. 2024).
12. Aschwanden, A. *et al.* Contribution of the Greenland Ice Sheet to sea level over the next millennium. *Science Advances* **5**, eaav9396. doi:10.1126/sciadv.aav9396 (June 2019).
13. Garbe, J., Albrecht, T., Levermann, A., Donges, J. F. & Winkelmann, R. The hysteresis of the Antarctic Ice Sheet. en. *Nature* **585**, 538–544. ISSN: 1476-4687. doi:10.1038/s41586-020-2727-5 (Sept. 2020).
14. Beckmann, J. & Winkelmann, R. Effects of extreme melt events on ice flow and sea level rise of the Greenland Ice Sheet. English. *The Cryosphere* **17**, 3083–3099. ISSN: 1994-0416. doi:10.5194/tc-17-3083-2023 (July 2023).
15. Zeitz, M., Reese, R., Beckmann, J., Krebs-Kanzow, U. & Winkelmann, R. Impact of the melt–albedo feedback on the future evolution of the Greenland Ice Sheet with PISM-dEBM-simple. English. *The Cryosphere* **15**, 5739–5764. ISSN: 1994-0416. doi:10.5194/tc-15-5739-2021 (Dec. 2021).
16. Bochow, N. *et al.* Overshooting the critical threshold for the Greenland ice sheet. en. *Nature* **622**, 528–536. ISSN: 1476-4687. doi:10.1038/s41586-023-06503-9 (Oct. 2023).
17. Bochow, N., Poltronieri, A. & Boers, N. Projections of precipitation and temperatures in Greenland and the impact of spatially uniform anomalies on the evolution of the ice sheet. English. *The Cryosphere* **18**, 5825–5863. ISSN: 1994-0416. doi:10.5194/tc-18-5825-2024 (Dec. 2024).
18. Seroussi, H. *et al.* ISMIP6 Antarctica: a multi-model ensemble of the Antarctic ice sheet evolution over the 21st century. English. *The Cryosphere* **14**, 3033–3070. ISSN: 1994-0416. doi:10.5194/tc-14-3033-2020 (Sept. 2020).
19. Nowicki, S. *et al.* Experimental protocol for sea level projections from ISMIP6 stand-alone ice sheet models. English. *The Cryosphere* **14**, 2331–2368. ISSN: 1994-0416. doi:10.5194/tc-14-2331-2020 (July 2020).
20. Sellevold, R. *et al.* Surface mass balance downscaling through elevation classes in an Earth system model: application to the Greenland ice sheet. English. *The Cryosphere* **13**, 3193–3208. ISSN: 1994-0416. doi:10.5194/tc-13-3193-2019 (Dec. 2019).

This page contains no comments

21. Noël, B. *et al.* A daily, 1 km resolution data set of downscaled Greenland ice sheet surface mass balance (1958–2015). English. *The Cryosphere* **10**, 2361–2377. ISSN: 1994-0416. doi:10.5194/tc-10-2361-2016 (Oct. 2016).
22. Tedesco, M., Colosio, P., Fettweis, X. & Cervone, G. A computationally efficient statistically downscaled 100&thinspm resolution Greenland product from the regional climate model MAR. English. *The Cryosphere* **17**, 5061–5074. ISSN: 1994-0416. doi:10.5194/tc-17-5061-2023 (Nov. 2023).
23. Bochow, N., Poltronieri, A., Rypdal, M. & Boers, N. Reconstructing historical climate fields with deep learning. *Science Advances* **11**, eadp0558. doi:10.1126/sciadv.adp0558 (Apr. 2025).
24. Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421. doi:10.1126/science.ad2336 (Dec. 2023).
25. Kochkov, D. *et al.* Neural general circulation models for weather and climate. en. *Nature* **632**, 1060–1066. ISSN: 1476-4687. doi:10.1038/s41586-024-07744-y (Aug. 2024).
26. Hess, P., Aich, M., Pan, B. & Boers, N. Fast, scale-adaptive and uncertainty-aware downscaling of Earth system model fields with generative machine learning. en. *Nature Machine Intelligence* **7**, 363–373. ISSN: 2522-5839. doi:10.1038/s42256-025-00980-5 (Mar. 2025).
27. Bi, K. *et al.* Accurate medium-range global weather forecasting with 3D neural networks. en. *Nature* **619**, 533–538. ISSN: 1476-4687. doi:10.1038/s41586-023-06185-3 (July 2023).
28. Price, I. *et al.* Probabilistic weather forecasting with machine learning. en. *Nature* **637**, 84–90. ISSN: 1476-4687. doi:10.1038/s41586-024-08252-9 (Jan. 2025).
29. Watt-Meyer, O. *et al.* ACE2: accurately learning subseasonal to decadal atmospheric variability and forced responses. en. *npj Climate and Atmospheric Science* **8**, 205. ISSN: 2397-3722. doi:10.1038/s41612-025-01090-0 (May 2025).
30. Juvet, G. *et al.* Deep learning speeds up ice flow modelling by several orders of magnitude. en. *Journal of Glaciology* **68**, 651–664. ISSN: 0022-1430, 1727-5652. doi:10.1017/jog.2021.120 (Aug. 2022).
31. Juvet, G. & Cordonnier, G. Ice-flow model emulator based on physics-informed deep learning. en. *Journal of Glaciology* **69**, 1941–1955. ISSN: 0022-1430, 1727-5652. doi:10.1017/jog.2023.73 (Dec. 2023).
32. Wang, Y., Lai, C.-Y., Prior, D. J. & Cowen-Breen, C. Deep learning the flow law of Antarctic ice shelves. *Science* **387**, 1219–1224. doi:10.1126/science.adp3300 (Mar. 2025).
33. Rosier, S. H. R., Bull, C. Y. S., Woo, W. L. & Gudmundsson, G. H. Predicting ocean-induced ice-shelf melt rates using deep learning. English. *The Cryosphere* **17**, 499–518. ISSN: 1994-0416. doi:10.5194/tc-17-499-2023 (Feb. 2023).
34. Bolibar, J. *et al.* Deep learning applied to glacier evolution modelling. English. *The Cryosphere* **14**, 565–584. ISSN: 1994-0416. doi:10.5194/tc-14-565-2020 (Feb. 2020).
35. Van der Meer, M. *et al.* A minimal machine-learning glacier mass balance model. English. *The Cryosphere* **19**, 805–826. ISSN: 1994-0416. doi:10.5194/tc-19-805-2025 (Feb. 2025).
36. Steidl, V., Bamber, J. L. & Zhu, X. X. Physics-aware machine learning for glacier ice thickness estimation: a case study for Svalbard. English. *The Cryosphere* **19**, 645–661. ISSN: 1994-0416. doi:10.5194/tc-19-645-2025 (Feb. 2025).
37. Verjans, V. & Robel, A. Accelerating Subglacial Hydrology for Ice Sheet Models With Deep Learning Methods. en. *Geophysical Research Letters* **51**, e2023GL105281. ISSN: 1944-8007. doi:10.1029/2023GL105281 (2024).
38. Harris, L., McRae, A. T. T., Chantry, M., Dueben, P. D. & Palmer, T. N. A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts. en. *Journal of Advances in Modeling Earth Systems* **14**, e2022MS003120. ISSN: 1942-2466. doi:10.1029/2022MS003120 (2022).
39. Hess, P., Lange, S., Schötz, C. & Boers, N. Deep Learning for Bias-Correcting CMIP6-Class Earth System Models. en. *Earth's Future* **11**, e2023EF004002. ISSN: 2328-4277. doi:10.1029/2023EF004002 (2023).
40. Aich, M. *et al.* Conditional diffusion models for downscaling & bias correction of Earth system model precipitation Apr. 2024. doi:10.48550/arXiv.2404.14416.
41. Arjovsky, M. & Bottou, L. *Towards Principled Methods for Training Generative Adversarial Networks* Jan. 2017. doi:10.48550/arXiv.1701.04862.
42. Bischoff, T. & Deck, K. Unpaired Downscaling of Fluid Flows with Diffusion Bridges. EN. *Artificial Intelligence for the Earth Systems* **3**. ISSN: 2769-7525. doi:10.1175/AIES-D-23-0039.1 (May 2024).
43. Song, Y., Dhariwal, P., Chen, M. & Sutskever, I. *Consistency Models* May 2023. doi:10.48550/arXiv.2303.01469.
44. Aich, M., Bathiany, S., Hess, P., Huang, Y. & Boers, N. *Diffusion models for probabilistic precipitation generation from atmospheric variables* Apr. 2025. doi:10.48550/arXiv.2504.00307.

This page contains no comments

45. Glaude, Q. *et al.* A Factor Two Difference in 21st-Century Greenland Ice Sheet Surface Mass Balance Projections From Three Regional Climate Models Under a Strong Warming Scenario (SSP5-8.5). en. *Geophysical Research Letters* **51**, e2024GL111902. ISSN: 1944-8007. doi:10.1029/2024GL111902 (2024).
46. Harder, P. *et al.* Hard-constrained deep learning for climate downscaling. *J. Mach. Learn. Res.* **24**, 365:17534–365:17573. ISSN: 1532-4435 (July 2023).
47. Delhasse, A., Beckmann, J., Kittel, C. & Fettweis, X. Coupling MAR (Modèle Atmosphérique Régional) with PISM (Parallel Ice Sheet Model) mitigates the positive melt–elevation feedback. English. *The Cryosphere* **18**, 633–651. ISSN: 1994-0416. doi:10.5194/tc-18-633-2024 (Feb. 2024).
48. Ho, J., Jain, A. & Abbeel, P. *Denoising Diffusion Probabilistic Models* Dec. 2020. doi:10.48550/arXiv.2006.11239.
49. Song, Y. *et al.* *Score-Based Generative Modeling through Stochastic Differential Equations* Feb. 2021. doi:10.48550/arXiv.2011.13456.
50. Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric* Apr. 2018. doi:10.48550/arXiv.1801.03924.
51. Bourgault, P. *et al.* xclim: xarray-based climate data analytics. en. *Journal of Open Source Software* **8**, 5415. ISSN: 2475-9066. doi:10.21105/joss.05415 (May 2023).
52. Cannon, A. J., Sobie, S. R. & Murdock, T. Q. Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? EN. *Journal of Climate* **28**, 6938–6959. ISSN: 0894-8755, 1520-0442. doi:10.1175/JCLI-D-14-00754.1 (Sept. 2015).
53. Seguinot, J. *PyPDD: a positive degree day model for glacier surface mass balance* Oct. 2019. doi:10.5281/zenodo.3467639.
54. Seguinot, J. Spatial and seasonal effects of temperature variability in a positive degree-day glacier surface mass-balance model. en. *Journal of Glaciology* **59**, 1202–1204. ISSN: 0022-1430, 1727-5652. doi:10.3189/2013JoG13J081 (Jan. 2013).
55. Calov, R. & Greve, R. A semi-analytical solution for the positive degree-day model with stochastic temperature variations. en. *Journal of Glaciology* **51**, 173–175. ISSN: 0022-1430, 1727-5652. doi:10.3189/172756505781829601 (Jan. 2005).
56. Crameri, F., Shephard, G. E. & Heron, P. J. The misuse of colour in science communication. en. *Nature Communications* **11**, 5444. ISSN: 2041-1723. doi:10.1038/s41467-020-19160-7 (Oct. 2020).

This page contains no comments