The Cryosphere, 20, 4235–4253, 2026
https://doi.org/10.5194/tc-20-4235-2026


The Cryosphere Open Access EGU

# The Modèle Atmosphérique Régional – Intelligence Artificielle (MAR-IA): surface meltwater over Greenland

**Marco Tedesco**[1], **Racheet Matai**[1], **and Xavier Fettweis**[2]

[1]Lamont-Doherty Earth Observatory, Columbia Climate School, Columbia University, New York, USA
[2]University of Liege, Liege, Belgium

**Correspondence:** Marco Tedesco (mtedesco@ldeo.columbia.edu)

Received: 29 January 2026 – Discussion started: 10 February 2026
Revised: 7 May 2026 – Accepted: 10 June 2026 – Published: 30 July 2026

**Abstract.** Surface melting over the Greenland Ice Sheet has become one of the dominant sources of contemporary and projected global sea-level rise, with melt rates accelerating over recent decades. Understanding those processes and feedbacks that control Greenland's surface melt is central to improving projections of future mass loss and to clarifying how changes in surface energy balance components shape ice-sheet stability.

To this aim, we developed MAR-IA - a machine-learning emulator of the MAR regional climate model – designed to emulate daily surface meltwater production over Greenland and to enable attribution of melt drivers. We implement two complementary emulators: a high-fidelity MAR-IA trained on full MAR surface energy balance fields and a reanalysis-compatible MAR-IA-ERA trained on predictors available from products such as ERA5, thereby extending applicability beyond MAR-specific outputs. Both emulators employ gradient-boosted trees optimized via Bayesian hyperparameter search, achieving high test-set skill, with the best-performing final configuration reaching $R^2 = 0.987$, low root mean squared error ($<10\,\mathrm{mm\,w.e.\,d^{-1}}$), and negligible bias relative to MAR meltwater outputs. We apply an explainable artificial intelligence (AI) analysis based on Shapley Additive Explanations (SHAP) to quantify how the importance of surface energy balance components,e.g., albedo, shortwave and longwave radiation, etc.,evolves across space and time over Greenland. Our results reveal robust spatial and temporal patterns in the dominance of radiative versus non-radiative drivers and demonstrate long-term trends in the relative contribution of temperature, shortwave radiation, and albedo to melt variability. These findings show that emulators can be used as powerful tools to complement regional climate models by enabling computationally efficient ensemble simulations and physically interpretable attribution of past and future Greenland surface melt.

## 1 Introduction

Surface melting over the Greenland Ice Sheet (GrIS) has emerged as one of the dominant contributors to contemporary and projected sea-level rise, with melt rates accelerating markedly over recent decades (Box et al., 2012; van den Broeke et al., 2016; Fettweis et al., 2020; Tedesco et al., 2016; Zhang et al., 2025). In this regard, it is crucial to understand the physical processes and feedbacks driving increased melt for improving projections of future mass loss and for elucidating the interactions among components of the surface energy and mass balance (Noël et al., 2015; Tedesco et al., 2016; Hofer et al., 2017; Lenaerts et al., 2019).

Regional climate models (RCMs), including the *Modèle Atmosphérique Régional* (MAR, Fettweis et al., 2013), *Regional Atmospheric Climate Model* (RACMO Noël et al., 2018), and *HIRHAM* (a regional climate model based on the High Resolution Limited Area Model, HIRLAM, and ECHAM physics; Langen et al., 2017), have substantially advanced our ability to simulate surface melt and better capture the surface energy- and mass-balance processes that govern meltwater production. Nevertheless, disentangling the complex, non-linear relationships among surface energy balance (SEB) components, such as albedo, surface temperature, and shortwave and longwave radiation, and their influence on surface melt remains computationally demanding and formally ill-posed (Mioduszewski et al., 2014). This limits the feasi-

Published by Copernicus Publications on behalf of the European Geosciences Union.

bility of large-ensemble simulations and hinder the systematic attribution of melt variability to individual drivers.

Recent developments in machine learning (ML) and explainable artificial intelligence (XAI) offer promising pathways to address such limitations. ML-based emulators can replicate RCM outputs with high fidelity while substantially reducing computational costs, enabling ensemble simulations and supporting model-interpretation analyses (Reichstein et al., 2019; Doury et al., 2023). Tree-based algorithms such as eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), deep convolutional neural networks (LeCun et al., 2015), and ensemble approaches (Materia et al., 2024) have been increasingly applied in Earth system science. When trained on RCM outputs, such models have the potential to emulate melt-related behavior and support attribution analyses using interpretable frameworks such as SHapley Additive exPlanations (SHAP; Lundberg and Lee, 2017).

Recent cryosphere-focused studies illustrate several related but distinct uses of machine learning. Lütjens et al. (2025) developed a deep-learning framework for spatiotemporal downscaling of surface meltwater by combining remote sensing and physics-based model output over Helheim Glacier. Bochow et al. (2025) used a physics-constrained generative ML framework trained on MAR to downscale monthly Greenland surface mass balance and surface temperature fields to higher spatial resolution. A more closely related recent study is Schlager et al. (2026), who developed a neural-network emulator of daily Greenland surface melt trained on output from the polar regional climate model HIRHAM5 and its firn model. Related cryosphere emulation efforts also include the emulation of firn-hydrology-related processes in Antarctica, further illustrating the growing role of ML emulators in polar climate applications (Veldhuijsen et al., 2025). In contrast, our contribution lies in the direct emulation of daily surface meltwater production from MAR and in the interpretable attribution of the dominant melt drivers across Greenland in space and time. Here, we introduce a novel ML-based emulator, the *Modèle Atmosphérique Régional – Intelligence Artificielle* (MAR-IA), trained on daily meltwater production (mm w.e. $d^{-1}$) simulated by MAR for the period 1979–2024. The predictor set includes SEB components and near-surface meteorological variables. We implemented two complementary training strategies: (1) a high-fidelity emulator (MAR-IA) optimized to minimize errors relative to MAR outputs using full SEB predictors, and (2) a reanalysis-compatible emulator (MAR-IA-ERA) trained on predictors available from a widely used dataset, ERA5 (Hersbach et al., 2020), enhancing applicability beyond MAR-specific outputs. In the MAR-IA-ERA configuration, ERA5 variables are reprojected to the MAR grid but are not dynamically downscaled to MAR-consistent fields. We therefore view this setup as a practical reanalysis-compatible emulator input strategy rather than a substitute for dedicated downscaling and note that ongoing ML-based downscaling developments are highly relevant for improving such applications (Doury et al., 2023; Bochow et al., 2025; Lütjens et al., 2025). Following the development of the emulators, we apply SHAP-based model interpretation to quantify the emulator-attributed importance of SEB components, such as albedo and incoming shortwave and longwave radiation, over the historical period (1979–2024). This approach provides a way to assess whether the learned predictor–melt relationships are broadly consistent with known melt physics and demonstrates the potential for interpretable ML models in cryospheric research. Our findings highlight the potential of ML emulators to complement traditional modeling frameworks, enabling computationally efficient simulations and robust attribution of melt drivers under past and future climate conditions. A conceptual overview of the MAR-IA framework is shown in Fig. 1.

## 2 Methods and data

### 2.1 The MAR model

We use the outputs from the Modèle Atmosphérique Régional (MAR), a regional climate model designed to simulate atmosphere–surface interactions over the Greenland ice sheet (Fettweis et al., 2017; Tedesco et al., 2023). The model was developed for simulations of polar and mountainous climates, with a particular focus on the surface mass balance (SMB) of ice sheets and glaciers (Fettweis, 2007; Franco et al., 2012) and has been extensively used to study the Greenland and Antarctic Ice Sheets (Agosta et al., 2019; Smith et al., 2023). MAR incorporates a three-dimensional atmospheric model coupled with a multilayer snow model, which allows it to resolve energy and mass exchanges at the snow–atmosphere interface, including processes such as meltwater percolation, refreezing, and densification (Brun et al., 1992).

The atmospheric component of MAR is forced at the boundaries of the Greenland region using data from global reanalysis products. For the MAR simulations used here, MAR was forced by the ERA5 product, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Hersbach et al., 2020). ERA5 is a global reanalysis dataset that combines observations from satellites, ground stations, and radiosondes with a numerical weather prediction model to provide hourly estimates of atmospheric, land, and ocean variables since 1940 (Delhasse et al., 2020). With a horizontal resolution of approximately 31 km, 137 vertical levels, and a detailed vertical structure, ERA5 offers comprehensive coverage of variables such as geopotential height, temperature, humidity, wind speed, and radiative and turbulent surface fluxes (Hersbach et al., 2020; Delhasse et al., 2020).

In the context of surface meltwater prediction, variables of primary interest include surface albedo (modeled through snow grain properties in Crocus), which controls reflected

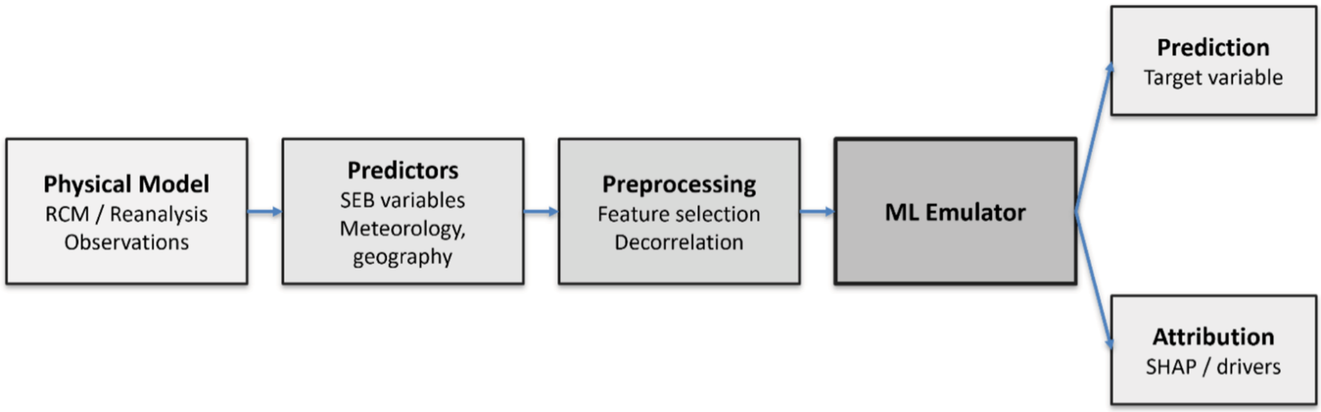



**Figure 1.** Conceptual overview of the MAR-IA framework.

versus absorbed solar radiation (Brun et al., 1992; Tedesco et al., 2016; Coléou and Lesaffre, 1998); and 2 m air temperature, a critical driver of melt energy. Radiative components such as downward shortwave and longwave radiation influence the surface energy balance, while turbulent fluxes of sensible and latent heat reflect energy exchange through convection and moisture transport (Hersbach et al., 2020; Delhasse et al., 2020; Wang et al., 2021). Topographic variables,elevation, latitude, and longitude,also modulate the local energy balance and atmospheric conditions (Tedesco et al., 2016a, b; Fettweis et al., 2017). These predictors are used to train our model, as explained in Sect. 2.3.

### 2.2 The XGBoost algorithm

Machine learning algorithms have become increasingly popular for extracting insights from complex datasets and making predictions from data. Among these algorithms, the eXtreme Gradient Boosting (XGBoost) algorithm has become one of the most popular due to its performance, accuracy, ability to handle a variety of data types, and capacity for processing large datasets (Chen and Guestrin 2016, Bentéjac et al., 2021). In this study, XGBoost was selected because our emulator is trained on a large tabular regression dataset composed of surface energy balance, meteorological, and geographic predictors. For this class of problem, gradient-boosted tree methods remain strong (Grinsztajn et al., 2022) and computationally efficient baselines, while also enabling efficient SHAP-based attribution analysis, which is a central objective of this work. We also tested a multilayer perceptron as a neural-network baseline, but it did not yield a meaningful improvement over XGBoost in predictive skill, while requiring substantially greater training time and tuning effort. XGBoost was therefore retained as the primary emulator model.

XGBoost has been successfully applied in several Earth- and climate-related studies such as risk assessment (Ma et al., 2021), classification and prediction (e.g. Zamani Joharestani et al., 2019). The algorithm combines the predictions of multiple decision trees to form a robust model with higher predictive accuracy (Chen and Guestrin 2016; Bentéjac et al., 2021). XGBoost combines two key ideas: boosting and decision trees. In boosting, trees are added sequentially, with each new tree trained to reduce the residual errors of the existing ensemble (Mayr et al., 2018). The final prediction is obtained by summing the contributions from all trees, with the model progressively improving as additional trees are added (Freund and Schapire, 1997; Chen and Guestrin, 2016). Each decision tree partitions the predictor space by applying a sequence of binary splits based on input-variable values, thereby grouping samples with similar predictor characteristics (Bentéjac et al., 2021). XGBoost also includes regularization terms that penalize overly complex trees, helping to reduce overfitting and improve generalization (Chen and Guestrin, 2016). A key practical advantage of XGBoost is that it scales well to large structured datasets and remains computationally tractable for tabular regression problems (Chen and Guestrin, 2016; Grinsztajn et al., 2022).

### 2.3 SHAP-based model interpretation

Explainable AI (XAI) comprises methods designed to help interpret machine-learning models that might otherwise be treated as "black boxes". In this study, we use SHAP (SHapley Additive exPlanations; Lundberg and Lee, 2017) to interpret the predictions of the MAR-IA models and to assess which predictors the emulator relies on most strongly, together with their spatial and temporal variability. SHAP is based on Shapley values from cooperative game theory, in which the contribution of each predictor is evaluated relative to the model prediction. Here, SHAP values are interpreted as model-based attributions of the trained emulator. They do not, by themselves, establish causal physical relationships.

In this framework, each predictor contributes to the final prediction. The SHAP value quantifies how much a given predictor contributes by comparing predictions from models that include or exclude that predictor across different predictor combinations. In this way, SHAP provides an additive measure of predictor contribution for each prediction. Positive SHAP values indicate that a predictor increases the predicted meltwater production relative to the model baseline, whereas negative values indicate that it decreases it. SHAP values were computed using the Python *shap* library applied to the trained XGBoost models. Because the emulator is tree-based, we used the TreeSHAP implementation for model-specific attribution.SHAP has been increasingly applied in Earth and climate sciences for both prediction (Dikshit and Pradhan, 2021; Al-Najjar et al., 2023; Batunacun et al., 2021; Ghafarian et al., 2022) and classification (Descals et al., 2023), although applications in cryosphere studies remain limited (Rohmer et al., 2022; Koo et al., 2023).

In this study, predictor importance is summarized primarily using mean absolute SHAP values, which capture the magnitude of contribution irrespective of sign. To provide a dominant directional tendency, we also compute a sign based on the Pearson correlation between predictor values and their SHAP values across samples. To assess long-term changes in emulator-attributed predictor importance, we further compute annual signed SHAP values for each variable and fit linear trends to the resulting time series. Because SHAP values depend on the trained model and on the predictor set provided to it, they are interpreted here as diagnostics of emulator behavior. In addition, correlations among predictors can redistribute attribution among related variables, even after removal of highly collinear inputs. For this reason, the SHAP results are discussed as model-based attributions that are compared against physical understanding, rather than as direct evidence of causality.

### 2.4 Training datasets and ML models

We tested multiple strategies for selecting training datasets for the machine-learning (ML) emulator. Table 1 lists the predictors included in each strategy. The first dataset, hereafter MAR-IA1, comprises variables from the MAR regional climate model (Table 1), including components of the surface energy balance (albedo, surface temperature, radiative fluxes, turbulent heat fluxes, e.g. Fettweis et al., 2017; Agosta et al., 2019). These predictors originate from MAR's internally consistent framework, providing a physically coherent reference for ML-based melt prediction (Tedesco et al., 2016, 2023). Conceptually, MAR-IA1 addresses how well the ML emulator can reproduce MAR outputs when forced by MAR-derived atmospheric fields – analogous to running MAR offline with its own atmospheric forcing.

To broaden applicability, we constructed a second dataset using similar predictors but sourced from ERA5 and reprojected to the MAR grid (MAR-IA2ERA). This enables use of MAR-IA by the wider scientific community without requiring MAR forcing. Predictor selection for MAR-IA2ERA was guided by correlation and SHAP based predictor – importance analysis from MAR-IA1, retaining variables such as surface and 2 m air temperature, downwelling and upwelling shortwave/longwave radiation, and sensible and latent heat fluxes. Sublimation was excluded due to its minor role in Greenland (Lenaerts et al., 2012).

We examined the value distributions of the predictor variables using kernel density estimates (Fig. 2). In general, for training ML models, predictor distributions that are not overly concentrated within narrow value ranges can help reduce bias toward dominant portions of the data distribution. From Fig. 2, we observe that the predictor distributions are generally broad, although several variables are clearly non-Gaussian. The purpose of Fig. 2 is to compare the overall distributional shapes of the predictors rather than to characterize tail behavior in detail. Log–log transformations did not yield a significant improvement in accuracy, so no scaling was applied. We also studied the correlation among predictors (Fig. 3). Highly correlated predictors do not add substantial independent information to the training. They can also affect the attribution analysis, since variables with similar roles may be treated as separate predictors. Hence, to mitigate multicollinearity, we removed highly correlated predictors. Specifically, we identified strong relationships between 2 m air and surface temperature ($r \sim 0.90$). We dropped surface temperature because it saturates at 0 °C during melt, whereas air temperature retains variability and may still provide information on melting. Upwelling shortwave and longwave radiation were also removed. The remaining predictors showed weak correlations with meltwater production. Based on these criteria, we defined MAR-IA2 (Table 1), a refined MAR-only dataset excluding surface temperature and upwelling radiation fields.

We also evaluated the distribution of ERA variables vs. those obtained with MAR, to assess their differences and address the potential impact on the emulator's performance. Figure 4 also shows the distribution of summer mean values for ERA predictors. While several variables within MAR and ERA show general agreement in their distribution and magnitude, differences exist for shortwave radiation and sensible heat flux. The largest discrepancy is observed in the case of albedo. In this case, ERA exhibits minimal spatial variability (Fig. 4), with values around 0.8. We expect this to have an impact on the model's performance, in view of the importance of albedo on surface melting. To assess this, we trained models without albedo and tested a variant in which ERA albedo was replaced with MAR albedo.

### 2.5 Hyperparameter optimization

Hyperparameter optimization refers to the process of selecting model parameters, such as learning rate, regularization strength, and tree depth, that are not learned during training but strongly influence model performance and generalization (Feurer et al., 2015). Common approaches include grid search, random search (Bergstra and Bengio, 2012), and

**Table 1.** Name of the models trained in this study and associated inputs used as predictors for the specific model.

| Model | Predictors | | | | | |
|---|---|---|---|---|---|---|
| | Ancillary | Temp. [K] | Radiation [$W\,m^{-2}$] | Fluxes [$W\,m^{-2}$] | Source | Albedo |
| MAR-IA1 | Lon., Lat., Elev. | Surf. Temp., 2 m Air Temp. | SW Rad., LW Rad., SW Up, LW Up | Sens. Heat Flux, Lat. Heat Flux | MAR | MAR |
| MAR-IA2 | Lon., Lat., Elev. | 2 m Air Temp. | SW Rad., LW Rad. | Sens. Heat Flux, Lat. Heat Flux | MAR | MAR |
| MAR-IA2_no_alb | Lon., Lat., Elev. | 2 m Air Temp. | SW Rad., LW Rad. | Sens. Heat Flux, Lat. Heat Flux | MAR | None |
| MAR-IA2ERA | Lon., Lat., Elev. | 2 m Air Temp. ERA | SW Rad. ERA, LW Rad. ERA | Sens. Heat Flux ERA, Lat. Heat Flux ERA | ERA | ERA |
| MAR-IA2ERA_no_alb | Lon., Lat., Elev. | 2 m Air Temp. ERA | SW Rad. ERA, LW Rad. ERA | Sens. Heat Flux ERA, Lat. Heat Flux ERA | ERA | None |
| MAR-IA2ERA_alb_mar | Lon., Lat., Elev. | 2 m Air Temp. ERA | SW Rad. ERA, LW Rad. ERA | Sens. Heat Flux ERA, Lat. Heat Flux ERA | ERA | MAR |

Bayesian optimization (Snoek et al., 2012), the latter using probabilistic models to explore the parameter space efficiently.

For this study, we applied Bayesian optimization with five-fold cross-validation, in which the training data are divided into five subsets and the model is trained repeatedly using four subsets for fitting and one for validation, to tune a subset of key XGBoost parameters known to affect predictive accuracy (Chen and Guestrin, 2016). Bayesian optimization constructs a surrogate model of the cross-validated error and iteratively selects new candidates using an acquisition function that balances exploration of uncertain regions and exploitation of promising areas. This approach reduces the number of evaluations required to identify near-optimal settings. The search space was defined over the following discrete hyperparameter values: n_estimators = {100, 200, 500, 700, 1000, 1500, 2000, 2500}, max_depth = {2, 3, 5, 10, 15}, learning_rate = {0.05, 0.10, 0.15, 0.20}, min_child_weight = {1, 2, 3, 4} , and gamma = {0, 0.25, 0.5}. This corresponds to 1920 possible combinations in the full discrete search space. Using Bayesian optimization, we evaluated 50 candidate configurations (n_iter = 50) rather than exhaustively testing all combinations.

The optimization objective was to minimize mean squared error (MSE) across folds. We used the BayesSearchCV implementation from the scikit-optimize package (Head et al., 2018) to tune five XGBoost hyperparameters that jointly govern model capacity and regularization: n_estimators (number of boosting rounds) sets the overall number of additive trees (larger values can improve fit but also increase computation and, if not counterbalanced by regularization, may amplify overfitting); max_depth controls the depth of individual trees (deeper trees can represent higher-order feature interactions but tend to increase variance); learning_rate scales the contribution of each tree (smaller values typically promote smoother, more stable optimization); min_child_weight specifies the minimum summed instance weight required to create a new child node; finally, gamma defines the minimum loss reduction required to introduce an additional split. We performed hyperparameter tuning on the MAR-IA1 model, which uses all MAR variables, to maximize agreement with MAR outputs and provide a baseline for other configurations. Model performance is reported using the coefficient of determination ($R^2$), which quantifies the fraction of variance explained by the model (higher is better); mean squared error (MSE), which measures the average squared deviation between predictions and observations (lower is better); root mean square error (RMSE); and bias, which represents the mean difference between predictions and ground truth, where values near zero indicate negligible systematic error. Starting from n_estimators = 100, performance was $R^2 \approx 0.938$, RMSE $\approx 1.082$ mm w.e. $d^{-1}$, and bias $\approx -1.0 \times 10^4$ mm w.e. $d^{-1}$. The global optimum occurred at n_estimators = 2500, max_depth = 15, min_child_weight = 4, learning_rate = 0.05, and gamma = 0, achieving $R^2 = 0.996$, RMSE = 0.249 mm w.e. $d^{-1}$, and bias = 0.0002 mm w.e. $d^{-1}$. Beyond several hundred trees, the accuracy plateaued while the computational cost – particularly for SHAP-based attribution – increased substantially. As a compromise, we fixed non-iterative hyperparameters and used n_estimators = 500, yielding $R^2 = 0.987$, RMSE $\approx 1.344$ mm w.e. $d^{-1}$, and bias $\approx -0.0002$ mm w.e. $d^{-1}$. This configuration retained most of the accuracy while keeping training and interpretation tractable and was applied consistently across all models.

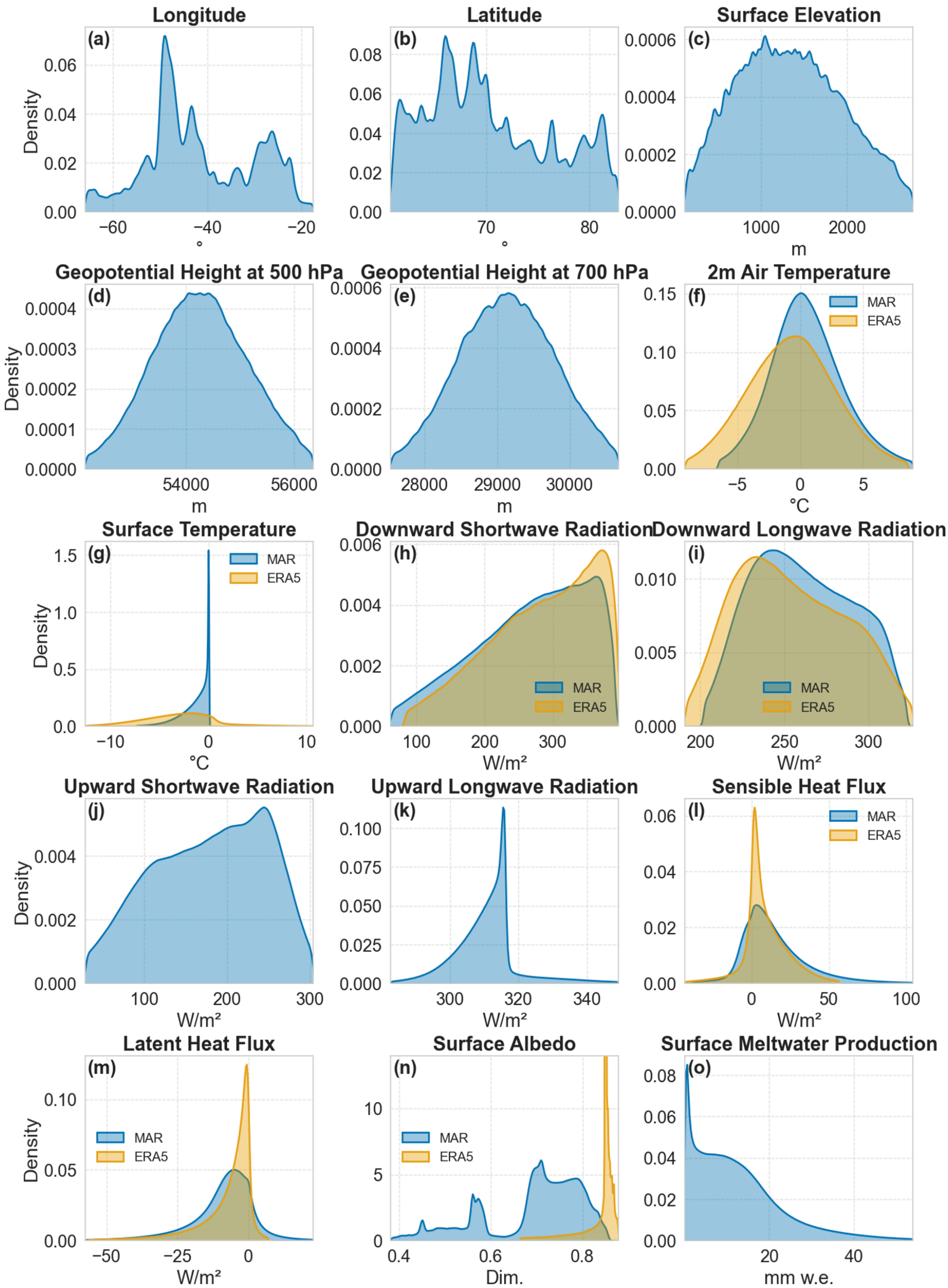



**Figure 2.** Distributions of MAR (blue) and ERA-5 (yellow) variables used for training the different models in this study.

## 3 Results

### 3.1 Models' performances

All machine-learning (ML) models were developed using a randomly shuffled split of the full dataset. Specifically, 70 % of the samples were used for training, 20 % for validation, and 10 % for testing. This split is a widely adopted heuristic in ML, providing sufficient training samples while retaining independent validation and performance assessment subsets (Gholamy et al., 2018; Sivakumar et al., 2024; Li et al., 2023; Xu et al., 2024). Figure 5 presents scatterplots of liq-

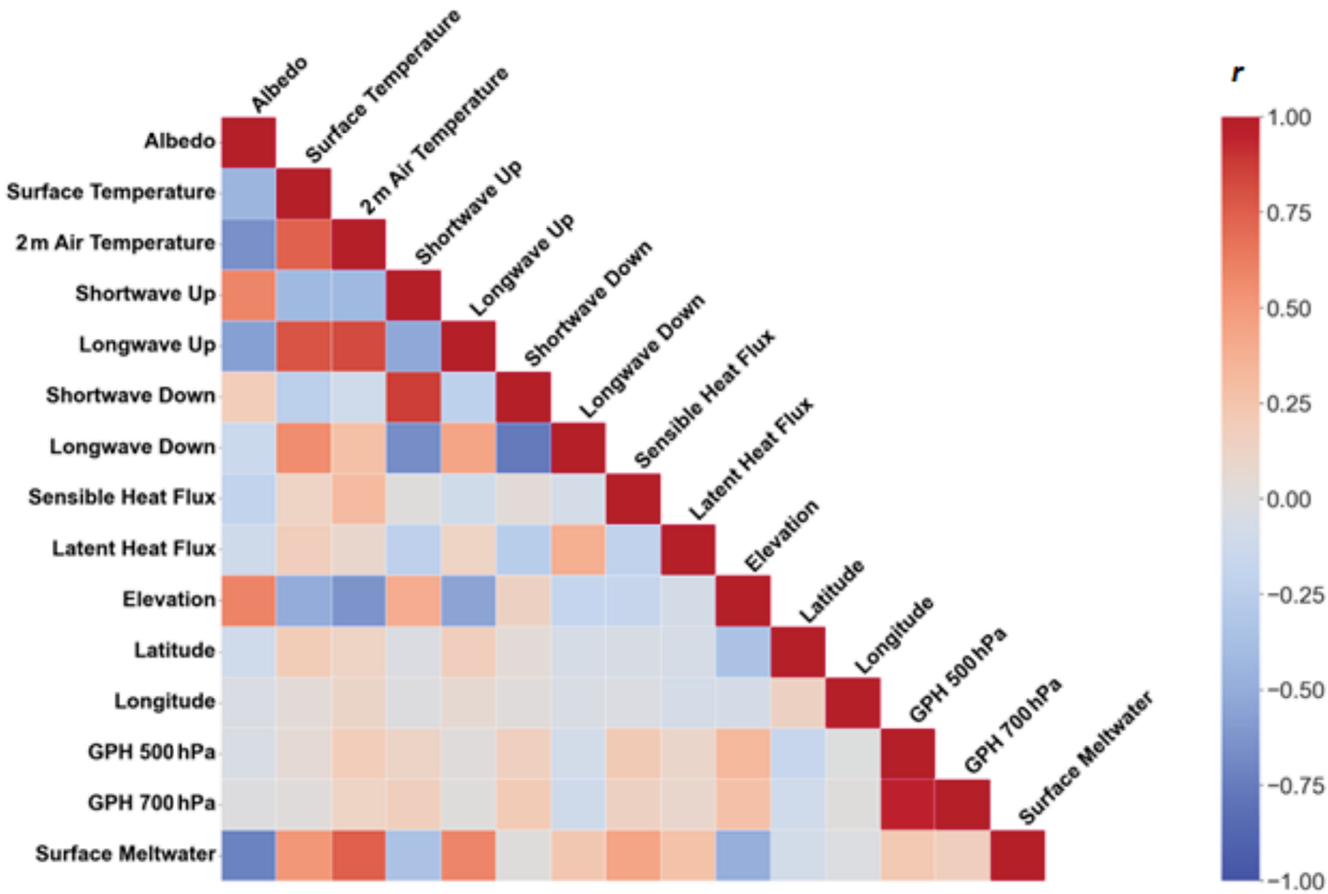



**Figure 3.** Heatmap of correlation coefficient ($r$) between the predictors and the predictand for the datasets used in the models' training.

uid meltwater production in the uppermost metre (here used as a predictor) obtained from the original MAR model ($x$-axis) with those obtained from the ML emulators ($y$-axis). Training, validation, and test data are shown, respectively, as red, green, and blue dots. Test-set performance metrics are summarised in Table 2 using the coefficient of determination ($R^2$), mean squared error (MSE), root mean squared (RMSE) and prediction bias. Table 2 also reports 95 % confidence intervals for RMSE and bias, which quantify uncertainty in these test-set performance estimates.

As expected, the MAR-IA1 model (Fig. 5a), which includes the full predictor set, achieves the highest accuracy ($R^2 = 0.987$, RMSE $= 1.344$ mm w.e. $d^{-1}$), with an low bias ($-0.0002$ mm w.e. $d^{-1}$). The narrow 95 % confidence intervals for RMSE and bias (Table 2) indicate stable test-set performance across the evaluated models. For comparison, an ordinary least-squares linear regression trained on the same predictor set achieved $R^2 = 0.90$ indicating a clear gain in predictive skill from the non-linear emulator. Excluding surface temperature and upwelling longwave and shortwave radiation (MAR-IA2) has little effect on $R^2$ (0.975) but results in a higher RMSE of 1.882 mm w.e. $d^{-1}$, reflected in the greater spread for low melt values (<20 mm w.e.; Fig. 5b). Using ERA-based predictors (MAR-IA2-ERA) further reduces performance ($R^2 = 0.839$, RMSE $= 4.737$ mm w.e. $d^{-1}$), with larger errors also concentrated at low melt rates (Fig. 5d). The small difference between MAR-IA2-ERA (Fig. 5d) and MAR-IA2-ERA-no_alb (Fig. 5e, where albedo from ERA is removed as a predictand) suggests a limited role of albedo in ERA-based configurations. This (as well as the deterioration in the ERA-based model's performance) is very likely due to the lack of spatial variability in the albedo product, as we anticipated in the previous section. Substituting ERA albedo with MAR albedo in the ERA dataset markedly improves accuracy (Fig. 5f), confirming the deterioration of the ML emulator because of the lack of granular information on albedo within the ERA product.

To assess sensitivity to temporal autocorrelation, we also performed a chronology-preserving experiment in which the models were trained on data up to 2014 and evaluated on data from 2015–2024. For MAR-IA2 and MAR-IA2-ERA, the resulting $R^2$ values were almost unchanged (0.970 and 0.838, respectively), indicating that the main performance conclusions are robust to this alternative split strategy.

**Table 2.** Test-set performance metrics ($R^2$, MSE, RMSE, bias, and 95 % confidence intervals for RMSE and bias) for each trained model using the predictor configurations described in Table 1.

| | $R^2$ | MSE/RMSE | RMSE 95 % CI | Bias | Bias 95 % CI |
|---|---|---|---|---|---|
| MAR-IA1 | 0.987 | 1.805/1.344 | (1.340, 1.347) | −0.0002 | (−0.0011, 0.0007) |
| MAR-IA2 | 0.975 | 3.543/1.882 | (1.877, 1.887) | −0.0010 | (−0.0022, 0.0002) |
| MAR-IA2_no_alb | 0.915 | 11.847/3.442 | (3.437, 3.447) | 0.0007 | (−0.0015, 0.0030) |
| MAR-IA2ERA | 0.839 | 22.435/4.737 | (4.731, 4.742) | 0.0021 | (−0.0010, 0.0054) |
| MAR-IA2ERA_no_alb | 0.824 | 24.467/4.946 | (4.941, 4.952) | 0.0023 | (−0.0010, 0.0057) |
| MAR-IA2ERA_mar_alb | 0.911 | 12.380/3.519 | (3.512, 3.525) | −0.0004 | (−0.0028, 0.0019) |

The spatial distribution of mean annual meltwater differences between the ML emulator and MAR is shown in Fig. 6 for the MAR-IA2 and MAR-IA2ERA configurations. The domain-mean difference for MAR-IA2 is 0.015 mm w.e. $yr^{-1}$ with a spatial standard deviation of 0.182 mm w.e. $yr^{-1}$, while MAR-IA2ERA exhibits a larger mean difference of 0.125 mm w.e. $yr^{-1}$ and a spatial standard deviation of 0.310 mm w.e. $yr^{-1}$. For reference, the standard deviation of MAR meltwater production is 1.97 mm w.e. $d^{-1}$.

### 3.2 SHAP-based interpretation of emulator predictions

We applied SHAP to the trained emulators to interpret which predictors contribute most strongly to their meltwater predictions and to assess whether these learned relationships are broadly consistent with known melt processes. Here, SHAP values are interpreted as model-based attributions rather than direct causal effects. To reduce computational cost, we randomly selected a subset of 1000 points and repeated the SHAP analysis 10 times. Figure 7 shows the mean absolute SHAP values for the different models. Red denotes predictors with positive SHAP contributions, whereas blue denotes predictors with negative SHAP contributions. Error bars show the standard deviation across the 10 runs, reflecting attribution variability within this subsampling procedure.In the case of MAR-IA1 (Fig. 7a), 2 m air temperature shows the highest SHAP value (4.14 mm w.e. $d^{-1}$), followed by the downward shortwave radiation (3.94 mm w.e. $d^{-1}$), the albedo (3.57 mm w.e. $d^{-1}$) and the sensible heat flux (2.16 mm w.e. $d^{-1}$). In the emulator, higher albedo is associated with negative SHAP contributions, whereas higher shortwave radiation, 2 m air temperature, and downward longwave radiation are associated with positive SHAP contributions. Latitude and surface meltwater production are negatively correlated, meaning that meltwater production decreases from south to north. Surface (skin) temperature and meltwater production are negatively correlated. Nevertheless, the SHAP value in the case of surface temperature is relatively small and the relationship can be explained by the fact that the surface temperature saturates to 0 °C when melting occurs, hence providing no sensitivity to melting and showing negative values when no melting occurs.

The SHAP values for the MAR-IA2 (Fig. 7b) model are similar to those for MAR-IA1, with albedo, air temperature and shortwave radiation still being the dominant terms. The importance of albedo in the model is shown by the results from the MAR-IA2_no_alb model. Moreover, the results (Fig. 7c) show that, also in the case of an ML model without albedo, the model still assigns the largest SHAP importance to physically plausible predictors, consistent with current understanding of melt processes. Indeed, air temperature becomes the dominant driver, followed by latitude and shortwave/longwave radiation.

Similar results are obtained in the case of the MAR-IA2-ERA, where 2 m air temperature (5.17 mm w.e. $d^{-1}$), surface elevation (3.68 mm w.e. $d^{-1}$) and downward shortwave (2.71 mm w.e. $d^{-1}$) are the three predictors with the largest SHAP importance. We point out the model's lack of sensitivity to albedo, given the spatio-temporal granularity of this variable in the ERA product. This hypothesis is also supported by the results obtained with MAR-IA2ERA_no_alb. In this case (Fig. 7e), we observe that the SHAP values are only mildly affected by the removal of the albedo. Furthermore, Fig. 7f shows the SHAP values obtained with the MAR-IA2ERA_mar_alb model, which uses all ERA-5 variables as predictors with the exception of albedo, which comes from MAR. In this case, the results are consistent with those obtained with the MAR-IA1 and MAR-IA2 models, with albedo gaining a dominant role in the emulator attribution.

### 3.3 Assessment of the ML model at selected sites

We also evaluated the performance of MAR-IA2 against the conventional MAR model at two well-established monitoring sites: the K-transect S6 station (Fig. 8a) and Swiss Camp (Fig. 8b), during the summer of 2012. These sites provide high-quality observational data, making them ideal benchmarks for assessing model accuracy. The comparison revealed strong agreement between MAR-IA2 and MAR in simulating surface energy fluxes, with only minor deviations across key variables, including net radiation, sensible heat flux, and latent heat flux. At the Swiss Camp station, the mean absolute difference between the machine learning model and MAR is 1.17 mm w.e. $d^{-1}$, with an RMSE

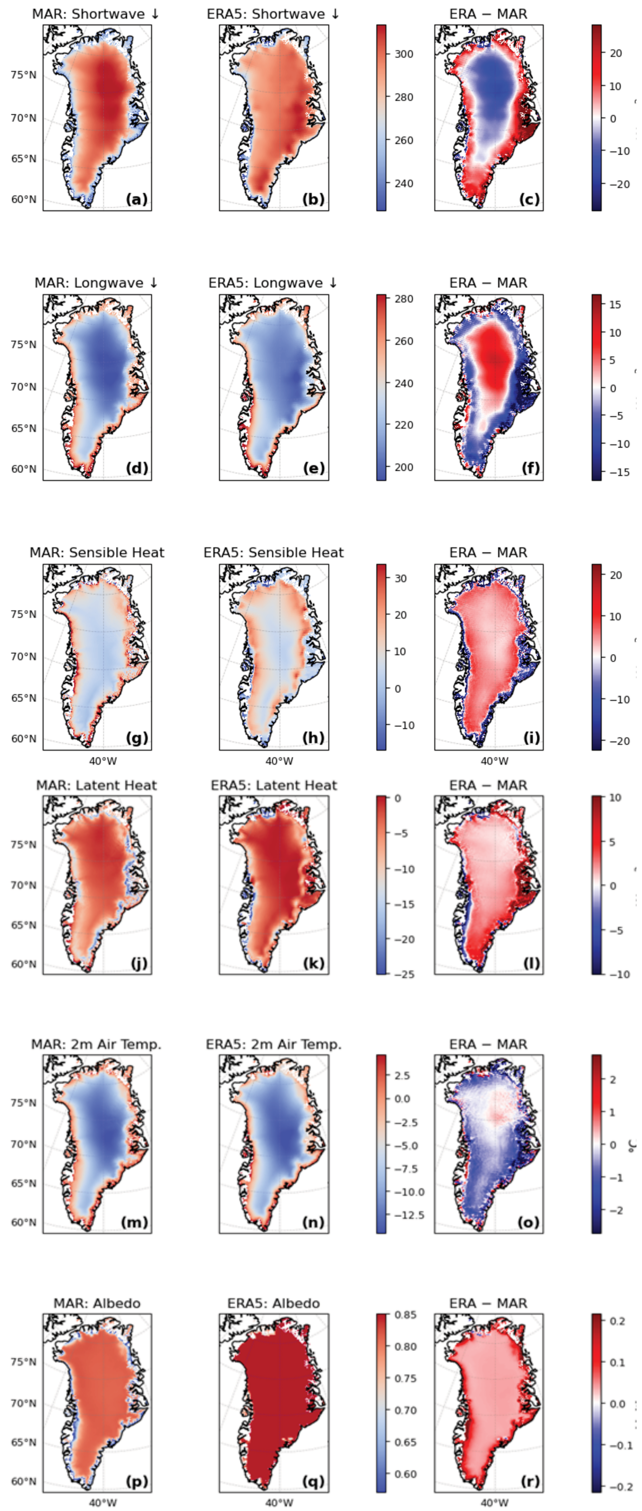



**Figure 4.** 1979–2024 summer (June–July–August, JJA) averaged values for the predictors used in the models' training (left and center columns) and differences between the MAR and ERA-5 values (right column).

of 1.52 mm w.e. $d^{-1}$, a maximum of 5.97 mm w.e. $d^{-1}$, and a standard deviation of 0.96 mm w.e. $d^{-1}$.The model performs comparably well at the K-transect S6 station, yielding a mean difference of 0.91 mm w.e. $d^{-1}$, an RMSE of 1.31 mm w.e. $d^{-1}$, a maximum of 6.54 mm w.e. $d^{-1}$, and a standard deviation of 0.93 mm w.e. $d^{-1}$.This result underscores the reliability of the ML-based emulator as a robust emulator for the MAR model. Peaks associated with key melt events are also well captured, as are the general magnitude and patterns of melt variability. For reference, near peak melt (top 20 percentile of melt), MAR-IA2 has a maximum absolute percentage error of 1.81 % at S6 station and 2.48 % at the Swiss camp site. Minor differences that do occur are generally modest lags or amplitude differences around peak melt events.

## 4 Discussion

After assessing the skill of the MAR-IA emulator in reproducing MAR meltwater output, we turn to the SHAP analysis to examine which predictors the emulator relies on most strongly and how those emulator-based attributions vary spatially and temporally.

Figure 9 shows the summer (JJA) mean SHAP values for the (a) non-flux and (b) flux variables for the period 1979–2024 for the MAR-IA2 model. Linear trends are also reported as dashed lines with the same color as the corresponding variables. Our results show that the relative role of the 2 m air temperature has increased over the past decades (slope = 0.0081 $yr^{-1}$, $p<0.01$, Table 3). Shortwave radiation also shows a significant increase in its influence (0.0051 $yr^{-1}$, $p \approx 0.01$), reflecting the growing dominance of radiative forcing. Sensible heat flux displays a similarly strong and statistically significant positive trend (0.0043 $yr^{-1}$, $p<0.01$), pointing to enhanced turbulent heat transfer. Albedo shows summer values similar (in magnitude) to those obtained in the case of air temperature. However, in this case, the trend is small and not statistically significant. This is consistent with the recent albedo variability observed over Greenland (Feng et al., 2023). Latent heat flux exhibits a moderate but significant upward trend (0.0028 $yr^{-1}$, $p \approx 0.03$), suggesting a rising role of moisture-related surface processes, while longwave radiation shows no meaningful directional trend ($p \approx 0.63$). Among the non-flux variables, elevation shows a strong negative trend (−0.0016 $yr^{-1}$, $p<0.01$), indicating that low-elevation regions increasingly dominate the spatial pattern of melt variability. Latitude also shows a statistically significant negative trend (−0.0022 $yr^{-1}$, $p<0.01$), whereas longitude exhibits no significant change over time. We note that geographic variables such as latitude and longitude should not be interpreted as direct physical melt drivers. Rather, within the emulator they may act as spatial proxies for regional gradients or unresolved covarying conditions. As a further exam-

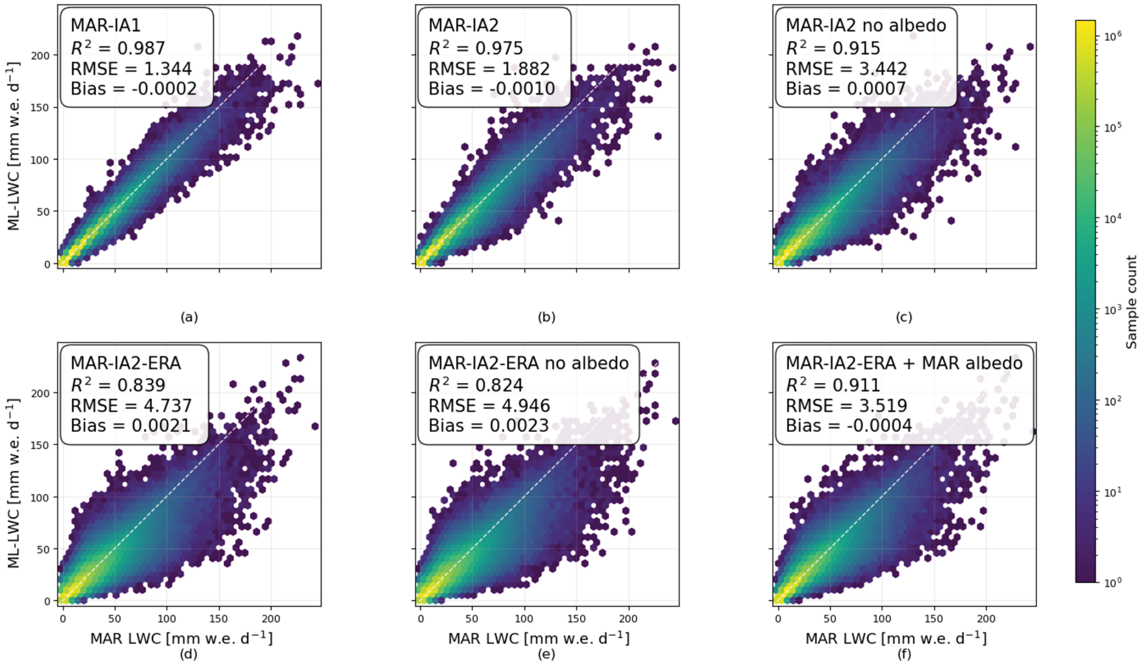



**Figure 5.** Scatterplot of the MAR ($x$-axis) and MAR-IA simulated ($y$-axis) surface meltwater production in the case of MAR-IA1 **(a)**, MAR-IA2 **(b)**, MAR-IA2_no_alb **(c)**, MAR-IA2ERA **(d)**, MAR-IA2ERA_no_alb **(e)**, MAR-IA2ERA_mar_alb **(f)**. Note only the test data statistics are reported and plotted.

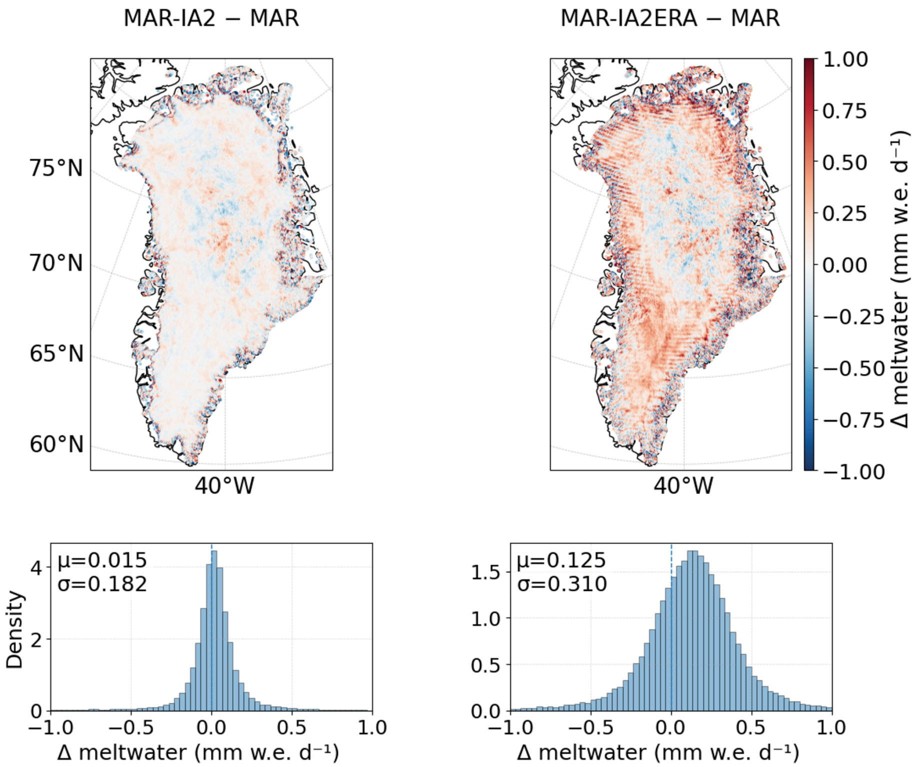



**Figure 6.** Mean difference in meltwater production between the machine-learning emulator and MAR for (left) MAR-IA2 and (right) MAR-IA2ERA configurations. Colors show the spatial distribution of the mean daily meltwater difference (ML − MAR) at each MAR grid point, obtained by averaging over all available days and years (1979–2024). The lower panels show the corresponding spatial distributions of the meltwater differences across the domain.

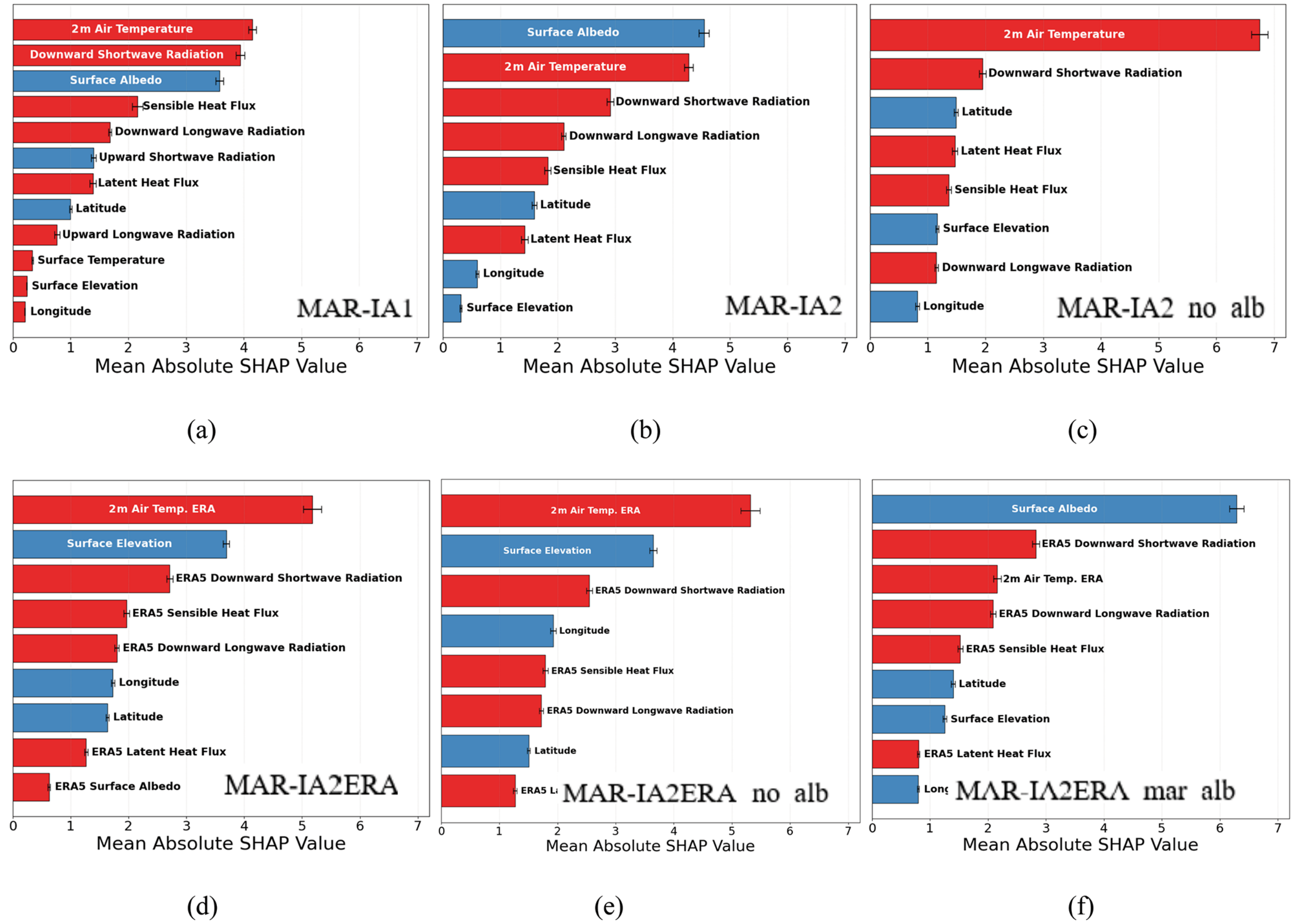
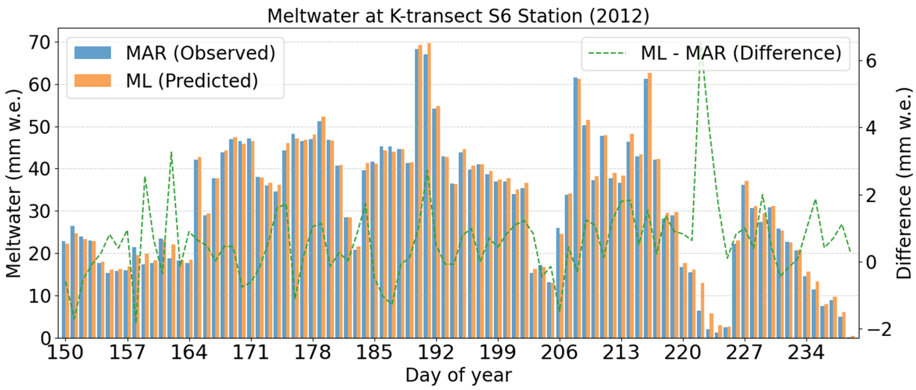



**Figure 7.** Mean absolute Shapley values obtained for the different models using the SHAP algorithm. Predictors that are negatively (positively) correlated with surface meltwater production are plotted as blue (red) bars.

ple of the relationships between SHAP values and the surface energy balance quantities, Fig. 10 shows the distribution of the averaged summed SHAP values, albedo and the MAR predicted meltwater for the years 1992 and 2012. These two years are characterized by opposite conditions in terms of surface melting, with 1992 being a low year, as a consequence of the Mount Pinatubo eruptions, and 2012 being an extremely high year in terms of meltwater production (e.g., Nghiem et al., 2012).

To better understand the relative role of drivers on surface melting and explore the potential of the MAR-IA2 emulator, we performed SHAP analysis for different elevation bands (Fig. 11). Specifically, we divide the ice sheet into three elevations bands, 0–1000, 1000–2000 and above 2000 m. Our choice of the elevation bands is partially driven by the elevation of the equilibrium line altitude (ELA, e.g., where runoff equals accumulation), which has been estimated to be fluctuating between 1200 and 1500 m. Our results in Fig. 11a shows that for areas between 0 and 1000 m and those between 1000 and 2000 m the largest SHAP contributions remain associated with albedo, air temperature, and downward shortwave radiation. This is similar to the results obtained when considering the whole ice sheet. A difference between the two elevation bands is that for areas at lower elevations the magnitude of the mean SHAP values is higher ($\sim 6$ mm w.e.) than for the region between 1000 and 2000 m ($\sim 4$ mm w.e.). For this region (Fig. 11b), albedo becomes the dominant driver, though the SHAP values are comparable to those for air temperature. For areas above 2000 m (Fig. 11c), downward longwave radiation becomes the third most important predictor in the SHAP ranking, still after albedo and surface temperature. This is consistent with recent studies (Nghiem et al., 2012; Neff et al., 2014) showing that the intrusion of warm, moist air to the top of the ice sheet was responsible for surface melting.

Lastly, in Figs. 12 and 13 we show the temporal evolution of daily SHAP values at the K-transect S6 (Fig. 12) and Swiss Camp (Fig. 13) stations for the year 2012 for the predictors used in MAR-IA2. The figures also show the – on the left $y$-axis – the corresponding daily values of the predic-




(a)

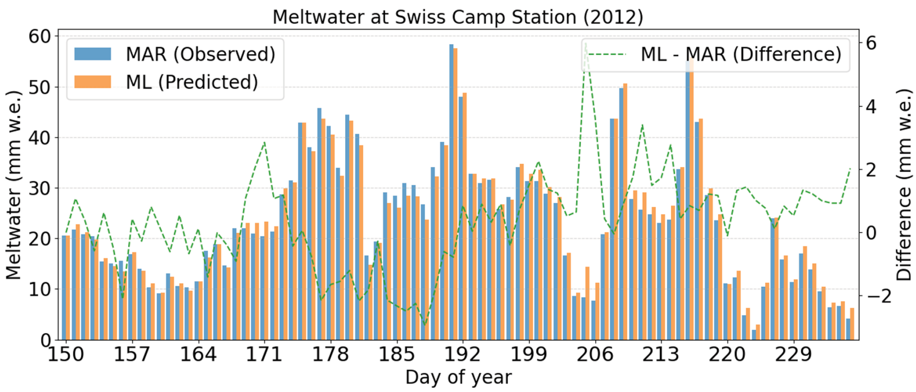



**Figure 8.** Comparison of daily MAR-IA2 Machine Learning Model and MAR Model Predictions at **(a)** K-transect S6 and **(b)** Swiss Camp Stations in Summer 2012.

**Table 3.** Trend (per year) of the mean summer Shapley coefficients, together with the $R^2$ of the linear regression used to estimate each trend.

| Variable | Group | Slope (per year) | $R^2$ | *p*-value |
|---|---|---|---|---|
| 2 m Air Temp. | Non-Flux Variables | 0.008067 | 0.33 | <0.01 |
| SW Rad. | Flux Variables | 0.005056 | 0.12 | 0.01 |
| Sens. Heat Flux | Flux Variables | 0.004325 | 0.21 | <0.01 |
| Lat. Heat Flux | Flux Variables | 0.002844 | 0.12 | 0.03 |
| LW Rad. | Flux Variables | −0.000349 | 0.01 | 0.63 |
| Elev. | Non-Flux Variables | −0.001553 | 0.44 | <0.01 |
| Lat. | Non-Flux Variables | −0.002175 | 0.15 | <0.01 |
| Lon. | Non-Flux Variables | −0.002276 | 0.01 | 0.69 |
| Albedo | Non-Flux Variables | −0.003355 | 0.04 | 0.16 |

tors. Over the course of the season, the magnitude and sign of individual SHAP values fluctuate, reflecting changing physical conditions and the covariate dependency of melt drivers. Most SHAP attributions covary with the predicted meltwater value, and the relationship between SHAP and predictand values changes dynamically. Obviously, in these examples, latitude, longitude, and elevation remain the same, though their relative ratios fluctuate. Albedo shows a less evident relationship between the SHAP value and the predictand, due to the saturation of low albedo values during intensive and prolonged melting periods. In the case of the K-transect, for example, we observe the albedo values of ∼ 0.4 lasting for several days starting at the end of May (day 180). Despite the albedo value remaining relatively constant, the SHAP values fluctuate to account for the changing conditions due to the variability of other forcings. Similar considerations apply to Swiss Camp (Fig. 12).


**Figure 9.** Summer (JJA) mean Shapley values for the **(a)** non-flux and **(b)** flux variables for the period 1979–2024. Linear trends are reported also as dashed lines with the same color as the corresponding variables. Note this is for MAR-IA2.

## 5 Conclusions

In this study, we introduced MAR-IA, a machine learning (ML) emulator designed to replicate surface meltwater production over the Greenland Ice Sheet (GrIS) as simulated by the MAR regional climate model (Fettweis et al., 2017). MAR-IA leverages XGBoost, a tree-based ML algorithm, trained on MAR outputs from 1979–2024 using predictors such as albedo, air temperature, radiative fluxes, and turbulent heat fluxes (MAR-IA1). Predictor selection was guided by both physical relevance and statistical analysis. Initially, we considered all variables associated with the surface energy balance, including albedo, surface and near-surface temperatures, shortwave and longwave radiation, and sensible and latent heat fluxes. Highly correlated variables, such as surface temperature and upwelling radiation fields, were removed to avoid redundancy and multicollinearity (Kim, 2019), which can obscure feature importance and inflate variance to train an alternative model (MAR-IA2). A third configuration, MAR-IA2-ERA, was trained on ERA5-based predictors (Hersbach et al., 2020) to facilitate wider usability, although its performance is affected by the limited spatial variability of ERA's albedo product. Experiments substituting ERA albedo with MAR albedo demonstrated substantial improvements in accuracy, underscoring the critical role of albedo in meltwater prediction and the need for improving albedo estimates within the ERA dataset.

Performance assessments reveal that MAR-IA1 achieved high agreement with MAR outputs on the test set, with $R^2 = 0.987$ and RMSE $= 1.344$ mm w.e. d$^{-1}$, confirming the emulator's fidelity. MAR-IA2 retains strong performance with $R^2 = 0.975$ and RMSE $= 1.882$ mm w.e. d$^{-1}$, despite using fewer predictors. In contrast, MAR-IA2-ERA shows reduced accuracy, with $R^2 = 0.839$ and RMSE $= 4.737$ mm w.e. d$^{-1}$, primarily due to ERA's lack of albedo representation. When MAR albedo replaces ERA albedo, performance improves markedly to $R^2 = 0.911$ and RMSE $= 3.519$ mm w.e. d$^{-1}$. These results demonstrate that while ERA-based models enable broader applicability, predictor quality – particularly albedo – remains essential for accurate meltwater emulation. Overall, MAR-IA provides a computationally efficient alternative to physically based models, supporting large-scale simulations and long-term attribution studies.

SHAP provides a useful framework for interpreting the predictions of the MAR-IA emulators by quantifying the contribution of each input predictor to the model output. In this study, we used SHAP to assess which predictors the emulator relies on most strongly and whether those learned predictor–melt relationships are broadly consistent with current physical understanding of surface melt over Greenland. The analysis showed that albedo, downward shortwave radiation, and 2 m air temperature consistently receive the largest SHAP importance, in agreement with established knowledge of the surface energy balance. At the same time, we interpret these results as model-based attributions, not as direct causal proof. In particular, SHAP values depend on the trained emulator and can be influenced by remaining correlations among

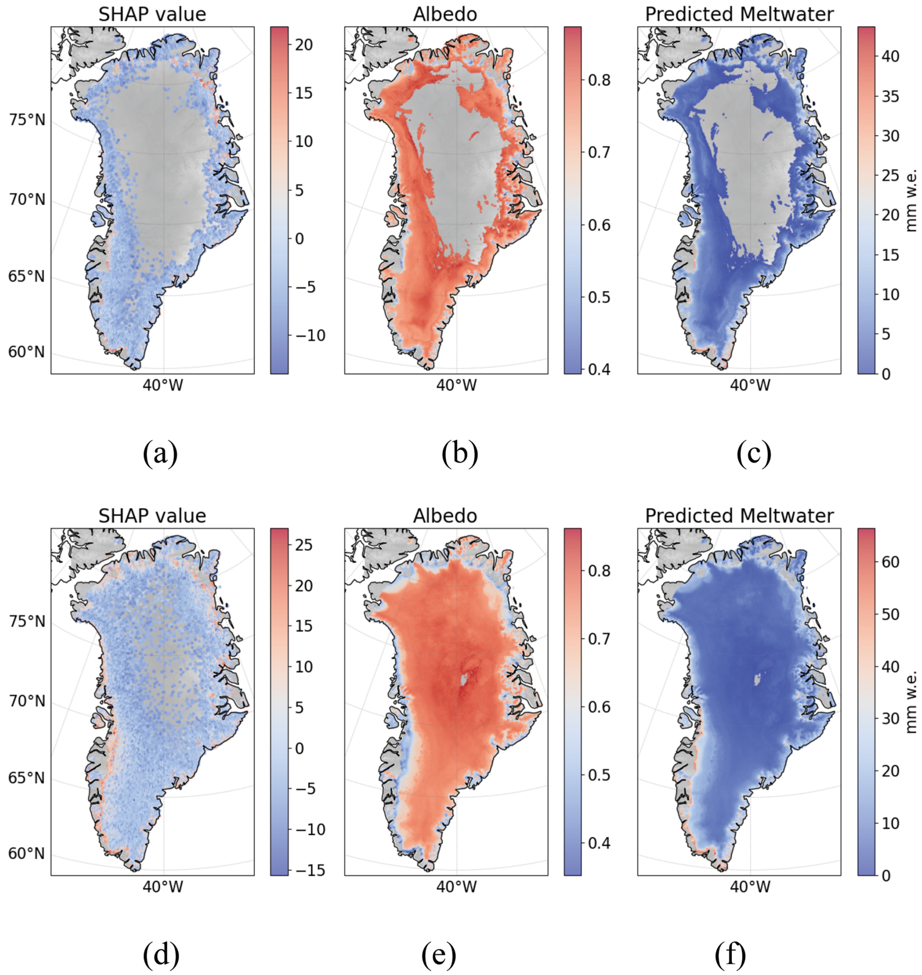



**Figure 10.** Summer-averaged spatial distribution of Shapley **(a, d)**, albedo **(b, e)** and ML-estimated meltwater **(c, f)** for the 1992 **(a, b, c)** and 2012 **(d, e, f)** for the MAR-IA2 model.

predictors. We therefore view the SHAP analysis as a physically informed interpretation of the emulator, rather than a replacement for direct process-based diagnosis from MAR itself.We studied how the relative role of the different predictors has been changing over the past decades over Greenland. We found an increasing contribution of 2 m air temperature, together with a rise in the influence of downward shortwave radiation, being consistent with a melt regime progressively dominated by atmospheric warming and enhanced radiative forcing. In parallel, the growing role of sensible heat flux might point to strengthened turbulent heat transfer as a co-evolving driver of melt variability. Albedo remains an important control on surface melt, with a summer mean influence comparable in magnitude to that of air temperature. However, its long-term trend is small and not statistically significant. This is consistent with recent studies showing substantial Greenland albedo variability, but not necessarily a simple monotonic increase in its contribution to melt over time (Feng et al., 2023). The comparison of SHAP summer maps for low-melt (1992) and extreme-melt (2012) years further illustrates how the emulator and SHAP decomposition can provide physically interpretable results across regimes and extremes: 1992 reflects externally forced radiative suppression following the Mount Pinatubo eruption, whereas 2012 captures the compounded influence of strong melt-favorable energy inputs and feedbacks associated with a well-documented pan-Greenland melt episode (Nghiem et al., 2012).

Attribution stratified by elevation bands further underscored that “dominant drivers” are not spatially uniform. While albedo, air temperature, and shortwave radiation remain primary controls below 2000 m, their absolute contributions are stronger at the lowest elevations, consistent with higher melt sensitivity. Above 2000 m, downward longwave radiation becomes comparatively more important. This supports the interpretation that episodic intrusions of warm, moist air, and the associated enhancement of longwave radiation, can be critical for high elevation melt events. This is again consistent with prior evidence from extreme years such as 2012 (Nghiem et al., 2012). Finally, station-scale daily

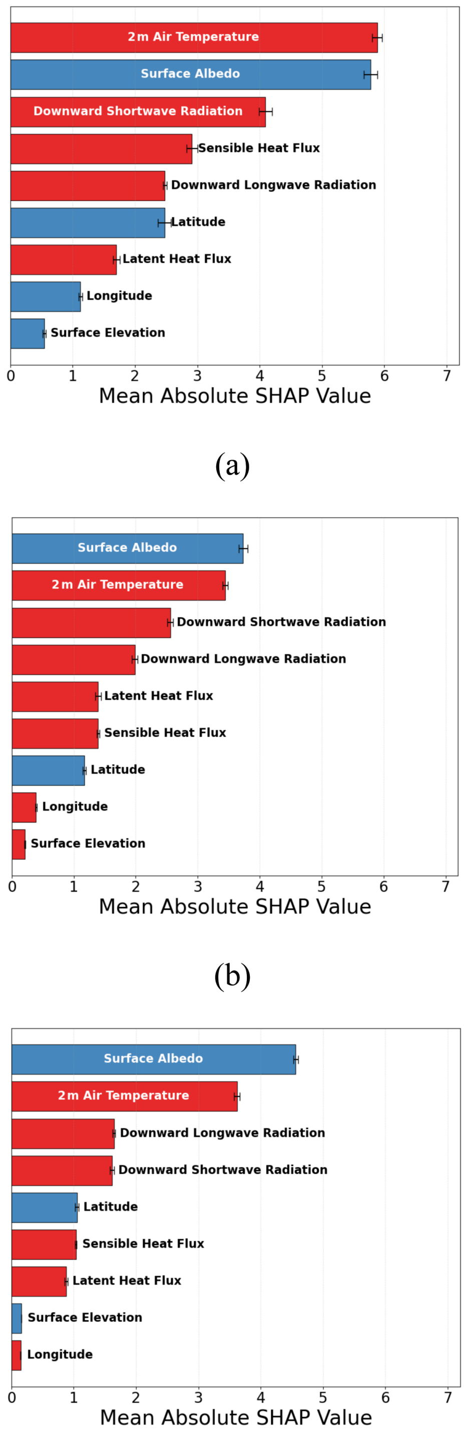



**Figure 11.** Mean absolute Shapley values obtained for the MAR-IA2 model for areas **(a)** below 1000 m, **(b)** between 1000 and 2000 m nd **(c)** above 2000 m.

analyses at K-transect S6 and Swiss Camp in 2012 demonstrate that melt drivers are highly state-dependent. SHAP values change sign and magnitude through the season as boundary-layer structure, radiative conditions, and surface state co-evolve. The results concerning albedo values and their SHAP attribution during prolonged melt – reflecting albedo saturation at low values – highlight how the MAR-IA model captures marginal contributions in context, not merely the instantaneous magnitude of a variable.

The development of MAR-IA opens several avenues for future research and practical applications. First, expanding the emulator to include additional target variables or predictors such as snowpack properties, cloud microphysics, and atmospheric circulation indices could improve its ability to capture complex melt dynamics under changing climate conditions and offer an insight into the relative role of dynamic and thermodynamic drivers. Our current work is to export the model to other regions – such as Antarctica, the Himalaya, and the European Alps – as well as extend our method to SMB. We also suggest that the low computational cost of the ML emulators allows running synthetic experiments to quantify uncertainty, conducting sensitivity experiments, and coupling the emulator with other models, focusing, for example, on modelling the total mass loss of the Greenland and Antarctic ice sheets, such as the ISSM (Larour et al., 2012). Moreover, the emulator offers a practical alternative when the full MAR atmospheric model cannot be run. MAR requires lateral boundary forcing over Greenland with meteorological variables at several vertical levels, and these inputs are not always available with sufficient completeness for historical reconstructions or future climate simulations. The emulator circumvents this requirement by predicting meltwater directly from a reduced set of atmospheric and surface predictors, without integrating the full atmospheric model. Its use for past or future applications is therefore promising, but should be restricted to conditions that remain within, or close to, the range represented in the training data, with further validation needed for true extrapolative cases.Tools like MAR-IA can accelerate large-scale attribution studies, sensitivity experiments, and multi-model intercomparisons, promoting collaborations across glaciology, climate science, and machine learning fields. We are not arguing that fundamental climate models and physically based knowledge of processes should be fully replaced by machine learning tools or methods. On the contrary, we argue that the real benefits arise from the proper alignment of fundamental research leading to the understanding of the processes and feedbacks driving the observed processes and machine learning tools. This can accelerate the pace of discovery and enable model-interpretation analyses via, for example, SHAP-based attribution, helping summarize emulator behavior at lower computational cost and supporting synthetic experiments that improve understanding of past changes while refining future estimates.

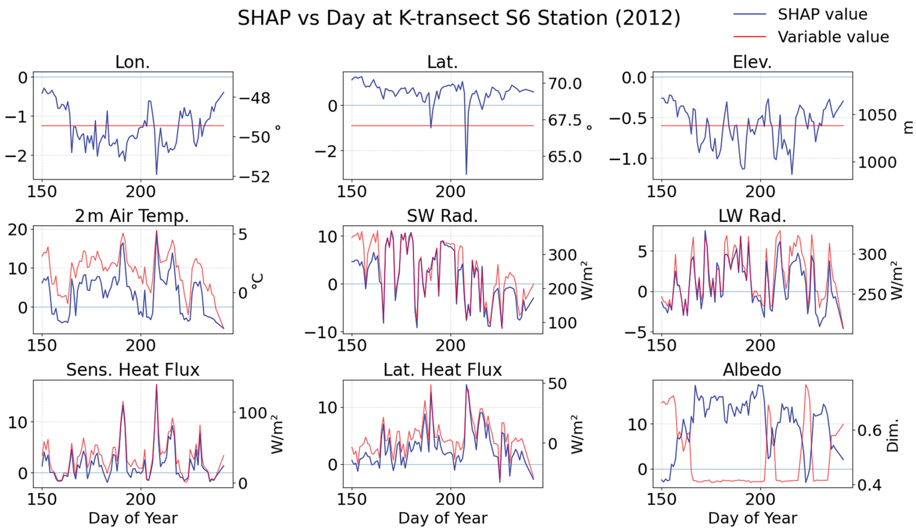



**Figure 12.** Daily Shapley values for the different predictors obtained from the model MAR-IA1 and associated trends (dashed red line) for the pixel containing the K-transect S6 station.

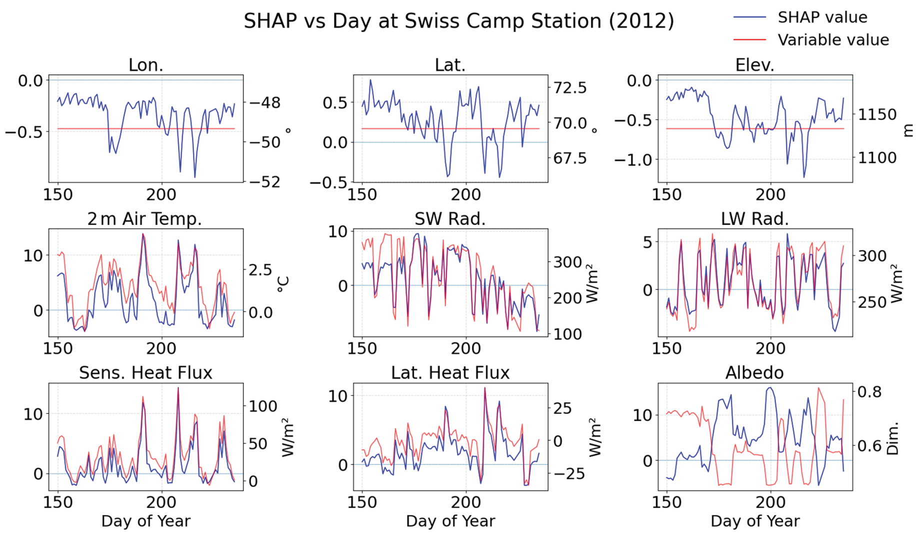



**Figure 13.** Same as Fig. 8 but for the Swiss Camp location.

*Code and data availability.* The MAR-IA models and the training datasets can be found here: https://doi.org/10.5281/zenodo.17942115 (Tedesco et al., 2025). The code is available at: https://github.com/racheetmatai/MAR_emulator (last access: 30 June 2026; https://doi.org/10.5281/zenodo.20838627, racheetmatai, 2026).

*Author contributions.* MT conceived the study and performed the analysis of the results. XF provided MAR outputs. RM supported the development and analysis of ML results.

*Competing interests.* At least one of the (co-)authors is a member of the editorial board of *The Cryosphere*. The peer-review process was guided by an independent editor, and the authors also have no other competing interests to declare.

*Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

*Acknowledgements.* We acknowledge funding from NSF through the Learning the Earth with Artificial Intelligence and Physics

(LEAP) Science and Technology Center (STC) (Award number 2019625).

We acknowledge the use of AI tools to revise grammar and suggest corrections.

*Financial support.* This research has been supported by the Heising-Simons Foundation (grant no. HSFOUND 2019-1160), the Directorate for Geosciences, Office of Polar Programs (grant nos. OPP1901603 and OPP2136938), and the National Aeronautics and Space Administration, NASA Earth Science Technology Office (grant no. 80NSSC19K0663).

*Review statement.* This paper was edited by Alexander Robinson and reviewed by two anonymous referees.

## References

Agosta, C., Amory, C., Kittel, C., Orsi, A., Favier, V., Gallée, H., van den Broeke, M. R., Lenaerts, J. T. M., van Wessem, J. M., van de Berg, W. J., and Fettweis, X.: Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979–2015) and identification of dominant processes, The Cryosphere, 13, 281–296, https://doi.org/10.5194/tc-13-281-2019, 2019.

Al-Najjar, H. A. H., Pradhan, B., Beydoun, G., Sarkar, R., Park, H.-J., and Alamri, A.: A novel method using explainable artificial intelligence (XAI)-based Shapley Additive Explanations for spatial landslide prediction using Time-Series SAR dataset, Gondwana Res., 123, 107–124, https://doi.org/10.1016/j.gr.2022.08.004, 2023.

Batunacun, Wieland, R., Lakes, T., and Nendel, C.: Using Shapley additive explanations to interpret extreme gradient boosting predictions of grassland degradation in Xilingol, China, Geosci. Model Dev., 14, 1493–1510, https://doi.org/10.5194/gmd-14-1493-2021, 2021.

Bentéjac, C., Csörgő, A., and Martínez-Muñoz, G.: A comparative analysis of gradient boosting algorithms, Artif. Intell. Rev., 54, 1937–1967, https://doi.org/10.1007/s10462-020-09896-5, 2021.

Bergstra, J. and Bengio, Y.: Random search for hyper-parameter optimization, J. Mach. Learn. Res., 13, 281–305, 2012.

Bochow, N., Hess, P., and Robinson, A.: Physics-constrained generative machine learning-based high-resolution downscaling of Greenland's surface mass balance and surface temperature, arXiv [preprint], https://doi.org/10.48550/arXiv.2507.22485, 2025.

Box, J. E., Fettweis, X., Stroeve, J. C., Tedesco, M., Hall, D. K., and Steffen, K.: Greenland ice sheet albedo feedback: thermodynamics and atmospheric drivers, The Cryosphere, 6, 821–839, https://doi.org/10.5194/tc-6-821-2012, 2012.

Brun, E., David, P., Sudul, M., and Brunot, G.: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting, J. Glaciol., 38, 13–22, https://doi.org/10.3189/S0022143000009552, 1992.

Chen, T. and Guestrin, C.: XGBoost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Coléou, C. and Lesaffre, B.: Irreducible water saturation in snow: experimental results in a cold laboratory, Ann. Glaciol., 26, 64–68, https://doi.org/10.3189/1998AoG26-1-64-68, 1998.

Delhasse, A., Kittel, C., Amory, C., Hofer, S., van As, D., S. Fausto, R., and Fettweis, X.: Brief communication: Evaluation of the near-surface climate in ERA5 over the Greenland Ice Sheet, The Cryosphere, 14, 957–965, https://doi.org/10.5194/tc-14-957-2020, 2020.

Descals, A., Verger, A., Yin, G., Filella, I., and Peñuelas, J.: Local interpretation of machine learning models in remote sensing with SHAP: the case of global climate constraints on photosynthesis phenology, Int. J. Remote Sens., 44, 3160–3173, https://doi.org/10.1080/01431161.2023.2217982, 2023.

Dikshit, A. and Pradhan, B.: Interpretable and explainable AI (XAI) model for spatial drought prediction, Sci. Total Environ., 801, 149797, https://doi.org/10.1016/j.scitotenv.2021.149797, 2021.

Doury, A., Somot, S., Gadat, S., Ribes, A., and Corre, L.: Regional climate model emulator based on deep learning: concept and first evaluation of a novel hybrid downscaling approach, Clim. Dynam., 60, 1751–1779, https://doi.org/10.1007/s00382-022-06343-9, 2023.

Feng, S., Cook, J. M., Anesio, A. M., Benning, L. G., and Tranter, M.: Long time series (1984–2020) of albedo variations on the Greenland ice sheet from harmonized Landsat and Sentinel 2 imagery, J. Glaciol., 69, 1225–1240, https://doi.org/10.1017/jog.2023.13, 2023.

Fettweis, X.: Reconstruction of the 1979–2006 Greenland ice sheet surface mass balance using the regional climate model MAR, The Cryosphere, 1, 21–40, https://doi.org/10.5194/tc-1-21-2007, 2007.

Fettweis, X., Franco, B., Tedesco, M., van Angelen, J. H., Lenaerts, J. T. M., van den Broeke, M. R., and Gallée, H.: Estimating the Greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model MAR, The Cryosphere, 7, 469–489, https://doi.org/10.5194/tc-7-469-2013, 2013.

Fettweis, X., Box, J. E., Agosta, C., Amory, C., Kittel, C., Lang, C., van As, D., Machguth, H., and Gallée, H.: Reconstructions of the 1900–2015 Greenland ice sheet surface mass balance using the regional climate MAR model, The Cryosphere, 11, 1015–1033, https://doi.org/10.5194/tc-11-1015-2017, 2017.

Fettweis, X., Hofer, S., Krebs-Kanzow, U., Amory, C., Aoki, T., Berends, C. J., Born, A., Box, J. E., Delhasse, A., Fujita, K., Gierz, P., Goelzer, H., Hanna, E., Hashimoto, A., Huybrechts, P., Kapsch, M.-L., King, M. D., Kittel, C., Lang, C., Langen, P. L., Lenaerts, J. T. M., Liston, G. E., Lohmann, G., Mernild, S. H., Mikolajewicz, U., Modali, K., Mottram, R. H., Niwano, M., Noël, B., Ryan, J. C., Smith, A., Streffing, J., Tedesco, M., van de Berg, W. J., van den Broeke, M., van de Wal, R. S. W., van Kampenhout, L., Wilton, D., Wouters, B., Ziemen, F., and Zolles, T.: GrSMBMIP: intercomparison of the modelled 1980–2012 surface mass balance over the Greenland Ice Sheet, The Cryosphere, 14, 3935–3958, https://doi.org/10.5194/tc-14-3935-2020, 2020.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., and Hutter, F.: Efficient and robust automated machine learning, Adv. Neur. In., 28, 2962–2970, 2015.

Franco, B., Fettweis, X., Lang, C., and Erpicum, M.: Impact of spatial resolution on the modelling of the Greenland ice sheet surface mass balance between 1990–2010, using the regional climate model MAR, The Cryosphere, 6, 695–711, https://doi.org/10.5194/tc-6-695-2012, 2012.

Freund, Y. and Schapire, R. E.: A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci., 55, 119–139, https://doi.org/10.1006/jcss.1997.1504, 1997.

Ghafarian, F., Wieland, R., Lüttschwager, D., and Nendel, C.: Application of extreme gradient boosting and Shapley Additive explanations to predict temperature regimes inside forests from standard open-field meteorological data, Environ. Modell. Softw., 156, 105466, https://doi.org/10.1016/j.envsoft.2022.105466, 2022.

Gholamy, A., Kreinovich, V., and Kosheleva, O.: Why 70/30 or 80/20 relation between training and testing sets: a pedagogical explanation, Departmental Technical Reports (CS), University of Texas at El Paso, https://scholarworks.utep.edu/cgi/viewcontent.cgi?article=2202&context=cs_techrep (last access: 21 June 2026), 2018.

Grinsztajn, L., Oyallon, E., and Varoquaux, G.: Why do tree-based models still outperform deep learning on typical tabular data?, Adv. Neur. In., 35, 507–520, 2022.

Head, T., MechCoder, G., Shcherbatyi, I., F, Zé Vinícius, C. Schröder, N. Campos, Young, T., Cereda, S., Fan, T. , Shi, K. (KJ), Schwabedal, J., Hvass-Labs, Pak, M., and Fabisch, A.: scikit-optimize/scikit-optimize: v0.5.2, Zenodo [code], https://doi.org/10.5281/zenodo.1207017, 2018.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hofer, S., Tedstone, A. J., Fettweis, X., and Bamber, J. L.: Decreasing cloud cover drives the recent mass loss on the Greenland Ice Sheet, Science Advances, 3, e1700584, https://doi.org/10.1126/sciadv.1700584, 2017.

Kim, J. H.: Multicollinearity and misleading statistical results, Korean Journal of Anesthesiology, 72, 558–569, https://doi.org/10.4097/kja.19087, 2019.

Koo, Y., Xie, H., Kurtz, N. T., Ackley, S. F., and Wang, W.: Sea ice surface type classification of ICESat-2 ATL07 data by using data-driven machine learning model: Ross Sea, Antarctic as an example, Remote Sens. Environ., 296, 113726, https://doi.org/10.1016/j.rse.2023.113726, 2023.

Langen, P. L., Fausto, R. S., Vandecrux, B., Mottram, R. H., and Box, J. E.: Liquid water flow and retention on the Greenland ice sheet in the regional climate model HIRHAM5: local and large-scale impacts, Frontiers in Earth Science, 4, 110, https://doi.org/10.3389/feart.2016.00110, 2017.

Larour, E., Seroussi, H., Morlighem, M., and Rignot, E.: Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM), J. Geophys. Res.-Earth, 117, F01022, https://doi.org/10.1029/2011JF002140, 2012.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/10.1038/nature14539, 2015.

Lenaerts, J. T. M., van den Broeke, M. R., van Angelen, J. H., van Meijgaard, E., and Déry, S. J.: Drifting snow climate of the Greenland ice sheet: a study with a regional climate model, The Cryosphere, 6, 891–899, https://doi.org/10.5194/tc-6-891-2012, 2012.

Lenaerts, J. T. M., Medley, B., van den Broeke, M. R., and Wouters, B.: Observing and modeling ice sheet surface mass balance, Rev. Geophys., 57, 376–420, https://doi.org/10.1029/2018RG000622, 2019.

Li, K., Wang, F., Yang, L., and Liu, R.: Deep feature screening: feature selection for ultra high-dimensional data via deep neural networks, Neurocomputing, 538, 126186, https://doi.org/10.1016/j.neucom.2023.03.047, 2023.

Lundberg, S. and Lee, S.-I.: A unified approach to interpreting model predictions, Adv. Neur. In., 30, 4765–4774, 2017.

Lütjens, B., Alexander, P., Antwerpen, R., Widmann, T., Cervone, G., and Tedesco, M.: MeltwaterBench: deep learning for spatiotemporal downscaling of surface meltwater, arXiv [preprint], https://doi.org/10.48550/arXiv.2512.12142, 2025.

Ma, M., Zhao, G., He, B., Li, Q., Dong, H., Wang, S., and Wang, Z.: XGBoost-based method for flash flood risk assessment, J. Hydrol., 598, 126382, https://doi.org/10.1016/j.jhydrol.2021.126382, 2021.

Materia, S., Palma García, L., van Straaten, C., Sungmin, O., Mamalakis, A., Cavicchia, L., Coumou, D., de Luca, P., Kretschmer, M., and Donat, M.: Artificial intelligence for climate prediction of extremes: state of the art, challenges, and future perspectives, Wiley Interdisciplinary Reviews: Climate Change, 15, e914, https://doi.org/10.1002/wcc.914, 2024.

Mayr, A., Binder, H., Gefeller, O., and Schmid, M.: The evolution of boosting algorithms, Method. Inform. Med., 53, 419–427, https://doi.org/10.3414/ME13-01-0122, 2018.

Mioduszewski, J. R., Rennermalm, A. K., Robinson, D. A., and Mote, T. L.: Attribution of snowmelt onset in Northern Canada, J. Geophys. Res.-Atmos., 119, 9638–9653, https://doi.org/10.1002/2013JD021024, 2014.

Neff, W., Compo, G. P., Ralph, F. M., and Shupe, M. D.: Continental heat anomalies and the extreme melting of the Greenland ice surface in 2012 and 1889, J. Geophys. Res.-Atmos., 119, 6520–6536, https://doi.org/10.1002/2014JD021470, 2014.

Nghiem, S. V., Hall, D. K., Mote, T. L., Tedesco, M., Albert, M. R., Keegan, K., Shuman, C. A., DiGirolamo, N. E., and Neumann, G.: The extreme melt across the Greenland ice sheet in 2012, Geophys. Res. Lett., 39, L20502, https://doi.org/10.1029/2012GL053611, 2012.

Noël, B., van de Berg, W. J., van Meijgaard, E., Kuipers Munneke, P., van de Wal, R. S. W., and van den Broeke, M. R.: Evaluation of the updated regional climate model RACMO2.3: summer snowfall impact on the Greenland Ice Sheet, The Cryosphere, 9, 1831–1844, https://doi.org/10.5194/tc-9-1831-2015, 2015.

Noël, B., van de Berg, W. J., van Wessem, J. M., van Meijgaard, E., van As, D., Lenaerts, J. T. M., Lhermitte, S., Kuipers Munneke, P., Smeets, C. J. P. P., van Ulft, L. H., van de Wal, R. S. W., and van den Broeke, M. R.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2 –

Part 1: Greenland (1958–2016), The Cryosphere, 12, 811–831, https://doi.org/10.5194/tc-12-811-2018, 2018.

racheetmatai: racheetmatai/MAR_emulator: MAR-IA v1.0.0 – Code accompanying the MAR-IA manuscript (v1.0.0), Zenodo [code], https://doi.org/10.5281/zenodo.20838627, 2026.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Rohmer, J., Thieblemont, R., Le Cozannet, G., Goelzer, H., and Durand, G.: Improving interpretation of sea-level projections through a machine-learning-based local explanation approach, The Cryosphere, 16, 4637–4657, https://doi.org/10.5194/tc-16-4637-2022, 2022.

Schlager, E., Scher, S., Mottram, R. H., and Langen, P. L.: Learning to melt: Emulating Greenland surface melt from a polar RCM with machine learning, The Cryosphere, 20, 3313–3343, https://doi.org/10.5194/tc-20-3313-2026, 2026.

Sivakumar, M., Parthasarathy, S., and Padmapriya, T.: Trade-off between training and testing ratio in machine learning for medical image processing, PeerJ Computer Science, 10, e2245, https://doi.org/10.7717/peerj-cs.2245, 2024.

Smith, B. E., Medley, B., Fettweis, X., Sutterley, T., Alexander, P., Porter, D., and Tedesco, M.: Evaluating Greenland surface-mass-balance and firn-densification data using ICESat-2 altimetry, The Cryosphere, 17, 789–808, https://doi.org/10.5194/tc-17-789-2023, 2023.

Snoek, J., Larochelle, H., and Adams, R. P.: Practical Bayesian optimization of machine learning algorithms, Adv. Neur. In., 25, 2951–2959, 2012.

Tedesco, M., Doherty, S., Fettweis, X., Alexander, P., Jeyaratnam, J., and Stroeve, J.: The darkening of the Greenland ice sheet: trends, drivers, and projections (1981–2100), The Cryosphere, 10, 477–496, https://doi.org/10.5194/tc-10-477-2016, 2016a.

Tedesco, M., Mote, T., Fettweis, X., Hanna, E., Jeyaratnam, J., Booth, J. F., Datta, R., and Briggs, K.: Arctic cut-off high drives the poleward shift of a new Greenland melting record, Nat. Commun., 7, 11723, https://doi.org/10.1038/ncomms11723, 2016b.

Tedesco, M., Colosio, P., Fettweis, X., and Cervone, G.: A computationally efficient statistically downscaled 100 m resolution Greenland product from the regional climate model MAR, The Cryosphere, 17, 5061–5074, https://doi.org/10.5194/tc-17-5061-2023, 2023.

Tedesco, M., Matai, R., and Fettweis, X.: MAR_ERA5_dataset_1979_2024_Greenland, Zenodo [data set], https://doi.org/10.5281/zenodo.17942115, 2025.

van den Broeke, M. R., Enderlin, E. M., Howat, I. M., Kuipers Munneke, P., Noël, B. P. Y., van de Berg, W. J., van Meijgaard, E., and Wouters, B.: On the recent contribution of the Greenland ice sheet to sea level change, The Cryosphere, 10, 1933–1946, https://doi.org/10.5194/tc-10-1933-2016, 2016.

Veldhuijsen, S. B. M., van de Berg, W. J., Kuipers Munneke, P., Hansen, N., Boberg, F., Kittel, C., Amory, C., and van den Broeke, M. R.: Emulating the expansion of Antarctic perennial firn aquifers in the 21st century, The Cryosphere, 19, 5157–5173, https://doi.org/10.5194/tc-19-5157-2025, 2025.

Wang, W., Zender, C. S., van As, D., Fausto, R. S., and Laffin, M. K.: Greenland surface melt dominated by solar and sensible heating, Geophys. Res. Lett., 48, e2020GL090653, https://doi.org/10.1029/2020GL090653, 2021.

Xu, J., Wu, C., Li, Y.-F., Danoy, G., and Bouvry, P.: Survey and taxonomy: the role of data-centric AI in transformer-based time series forecasting, arXiv [preprint], https://doi.org/10.48550/arXiv.2407.19784, 2024.

Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., and Talebiesfandarani, S.: PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data, Atmosphere, 10, 373, https://doi.org/10.3390/atmos10070373, 2019.

Zhang, Q.-L., Ding, M.-H., van den Broeke, M. R., Noël, B., Fettweis, X., Wang, S., Sun, W.-J., You, Q.-L., Xiao, C.-D., Qin, D.-H., and Huai, B.-J.: Variations in Greenland surface melt and extreme events from 1958 to 2023, Advances in Climate Change Research, 16, 910–921, https://doi.org/10.1016/j.accre.2025.05.004, 2025.