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

M. Tedesco[1], R. Matai[1] and X. Fettweis [2]

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

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

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 mm w.e. day^{-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

Deleted: *model -*

Deleted: *variables*

Deleted: *Both emulators employ gradient-boosted trees optimized via Bayesian hyperparameter search, achieving test-set performance up to* $R^2 \approx 0.99$ *with low mean squared error and negligible bias relative to MAR meltwater outputs.*

Deleted: *We apply a SHAP-based explainable AI analysis*

Deleted: —

Deleted: —

Deleted: *Development of regional climate models should go hand in hand with ML-based tools.*

Deleted: a

Deleted: enhanced

Deleted: a

Deleted: RACMO (

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 feasibility 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. day^{-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 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

Deleted: HIRHAM (

Deleted: driving processes

Deleted: —

Deleted: —

Deleted: meltwater production

Deleted: ;Pirk et al. 2022

Deleted: promoting attribution analyses

Deleted: XGBoost

Deleted: b

Deleted: have demonstrated strong predictive capabilities in Earth system applications.

Deleted: When trained on RCM outputs, these models can emulate melt dynamics and quantify the relative importance of SEB drivers using interpretable frameworks such as SHapley Additive exPlanations (SHAP; Lundberg and Lee, 2017).

Deleted: Despite these advances, no study has systematically applied ML to emulate surface melt from a Greenland-focused RCM or to perform long-term attribution of melt drivers using interpretable ML techniques.¶

Deleted: Predictands

Deleted: are drawn from SEB components and near-surface meteorological fields.

Deleted: variables

Deleted: s such as

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.

Deleted: Following the development of the emulators, we apply SHAP-based attribution analysis to quantify the evolving importance of SEB components—such as albedo, incoming shortwave and longwave radiation—over the historical period (1979–2024).

Deleted: This approach provides new insights into the physical mechanisms governing melt variability and establishes a benchmark for the interpretability of ML models in cryospheric research.

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 versus absorbed solar radiation (Brun et al. 1992; M. Tedesco et al. 2016; Coléou and Lesaffre 1998); and 2-meter 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., 2016, 2016b; Fettweis et al. 2017). These predictors are used to train our model, as explained in Section 2.3.

Deleted: —

Deleted: —

Deleted: a

Deleted: variables

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, 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

Deleted: Machine learning algorithms have become increasingly popular for many tasks, such as extracting insights and making predictions from data (Bentéjac et al. 2021).

Deleted: a

Deleted: —

Deleted: a

Deleted: Two concepts are key to describing how XGBoost works: a) boosting and b) decision trees.

Deleted: a

Deleted: a

Deleted: a

Deleted: Boosting is a technique that iteratively improves weak models by adding new ones that correct previous errors (Mayr 2018). In boosting, each new model is trained on samples misclassified by the previous model, and the final prediction is a weighted average of all models, with more weight given to higher-accuracy models (Freund and Schapire 1997). Decision trees follow a tree-like structure, dividing data at each node based on thresholds of input variables (Bentéjac et al. 2021). Each branch reflects a potential outcome. While single trees are unstable to small data perturbations, XGBoost mitigates this by using an ensemble of many trees and averaging predictions (Chen and Guestrin 2016a). The algorithm also includes a regularization term in its objective function to prevent overfitting (Chen and Guestrin 2016a). A key advantage of XGBoost is its scalability—it handles datasets with millions of samples and thousands of features efficiently (Chen and Guestrin 2016a)

Deleted: Attribution analysis and Shapley coefficients

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

Formatted: Font: Italic

Deleted: Explainable AI (XAI) is a set of algorithms and tools designed to help interpret and understand ML models that are otherwise perceived as "black box". In this study, we employ one of such tools called SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) to "explain" the output of our MAR-IA models and gain insights into the drivers of surface melting, as well as their spatial and temporal distribution. The Shapley algorithm is named after Lloyd Shapley, who introduced it in 1951 and won the Nobel Memorial Prize in Economic Sciences in 2012 for it. The algorithm works as follows: a coalition of players cooperates and obtains a certain overall gain from that cooperation, which in our case is the minimization of the error between the original MAR output and the values simulated by MAR-IA. Since some players may contribute more to the end goal than others, it is legitimate to ask ourselves how important each player is to the overall cooperation, and what payoff can he or she reasonably expect? The Shapley value provides an answer to this question which, in ML terms, corresponds to explaining the predictions of a trained model by interpreting a model trained on a set of features as a value function on a coalition of players. Hence, Shapley values provide a natural way to compute which features contribute to a prediction (Lundberg and Lee 2017) or contribute to the uncertainty of a prediction (Watson et al. 2023). To determine the importance of each feature in our dataset, the Shapley algorithm retrains the model to be explained on feature subsets by systematically adding and removing each feature from the training dataset. The effect of including a feature is calculated based on the model prediction difference between the model variations with and without that feature. Because the effect of each feature depends on the presence of other features, the model is trained on all possible feature subsets to capture each feature's effect. The additive importance value of each feature is calculated as a weighted average of all possible differences in model variations. Features with a positive Shapley value positively impact the model prediction and vice versa. The Shapley algorithm has seen remarkable successes in the Earth and climate sciences related to predictions (Dikshit and Pradhan, 2021; Al-Najjar et al., 2023; Batunacun et al., 2021; Ghafarian et al., 2022) and classification (Descals et al., 2023). However, limited applications exist in the cryosphere sciences (Rohmer et al., 2022; Koo et al., 2023), with this paper being, to our knowledge, the first to use Shapley coefficients to stud ... [1]

Deleted: a

Deleted: feature

Deleted: We examined the distribution of the predictors

Deleted: 1

Deleted: 1

Deleted: 1

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 ~ 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. Fig. 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.

Deleted: In general, for training ML models, near-normal distributions reduce the risk of overfitting by avoiding bias toward dominant value ranges. From Fig. 1, we observe that most predictors were approximately normal, except latitude and longitude. Log–log transformations did not yield a significant improvement in accuracy, so no scaling was applied. ¶

Deleted: ands

Deleted: 2

Deleted: Having highly correlated predictands not only does not add any information to the training, but it can also impact the attribution analysis, in view of having predictands whose relative role is similar, but they are treated as separate variables. Hence, to mitigate multicollinearity, we removed highly correlated predictors (r > 0.9

Deleted: kin

Deleted: We decided to drop skin temperature it saturates at 0 °C during melt, whereas air temperature retains variability and might be able to provide information on melting

Deleted: kin

Deleted: 3

Deleted: 3

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 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.

Deleted: —

Deleted: —

Deleted: a

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 ≈ 1.082 mm w.e. day^{-1}, and bias ≈ -1.0×10^{-4} mm w.e. day^{-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. day^{-1}, and bias = 0.0002 mm w.e. day^{-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 ≈ 1.344 mm w.e. day^{-1}, and bias ≈ -0.0002 mm w.e. day^{-1}. This configuration retained most of the accuracy while keeping training and interpretation tractable and was applied consistently across all models.

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). Fig. 5 presents scatterplots of liquid 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.day^{-1}), with an low bias (−0.0002 mm w.e. day^{-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

Deleted: 17

Deleted: 062

Deleted: (

Deleted:)^2

Deleted: 82

Deleted: 349

Deleted: (

Deleted:)

Deleted: ^{2}

Deleted: 6

Deleted: We found the following optimal values: max_depth = 15, min_child_weight = 4, learning_rate = 0.05, and gamma = 0.…

Deleted: All machine-learning (ML) models were trained on 80% of the whole dataset, with 20 % reserved for validation and 10 % for testing.

Deleted: ure

Deleted: 4

Deleted: P

Deleted: R^2 quantifies the fraction of variance explained by the model (higher is better), MSE measures the average squared deviation between predictions and observations (lower is better, units in mm w.e.^{2}), and bias represents the mean difference between predictions and ground truth, where values near zero indicate negligible systematic error.

Deleted: 4

Deleted: 8

Deleted: extremely

Deleted: 8×10^{-4}

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. day^{-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. day^{-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. When 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.

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. day^{-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. Fig. 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), 2m air temperature shows the highest SHAP value (4.14 mm w.e. day^{-1}), followed by the downward shortwave radiation (3.94 mm w.e. day^{-1}), the albedo (3.57 mm w.e. day^{-1}) and the sensible heat flux (2.16 mm w.e.day^{-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

Deleted: kin

Deleted: 7

Deleted: 4

Deleted: 746

Deleted: 4

Deleted: 4

Deleted: 4

Deleted: 4

Deleted: 5

Deleted: Feature Importance and testing the model's explanatory nature…

Deleted: ure

Deleted: 6

Deleted: We applied the SHAP algorithm to our models to assess their explanatory nature and physical consistency. In simple terms, feature importance analysis allows us to assess which variables drive the evolution of liquid water content. This is a crucial aspect for properly understanding whether the ML model is obtaining the right answer for the right reasons. To mitigate the computational cost of running Shapley for all cases, we randomly selected a subset of 1000 points and ran the SHAP algorithm 10 times. Fig. 6 shows the mean absolute Shapley values obtained for the different models. Red denotes features with positive SHAP contributions, whereas blue denotes that higher feature with negative contribution. Error bars show the standard deviation of SHAP values across the 10 cross-validation runs, reflecting variability in attribution stability.¶

Deleted: 6

Deleted: Shapley

Deleted: As expected, albedo is negatively correlated with surface meltwater production, where there is a positive relationship between shortwave radiation, 2m air temperature and downward longwave radiation.

Deleted: kin

Deleted: kin

saturates to 0ºC 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 2m air temperature (5.17 mm w.e. day^{-1}), surface elevation (3.68 mm w.e. day^{-1}) and downward shortwave (2.71 mm w.e. day^{-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. day^{-1}, with an RMSE of 1.52 mm w.e. day^{-1}, a maximum of 5.97 mm w.e. day^{-1}, and a standard deviation of 0.96 mm w.e. day^{-1}. The model performs comparably well at the K-transect S6 station, yielding a mean difference of 0.91 mm w.e. day^{-1}, an RMSE of 1.31 mm w.e. day^{-1}, a maximum of 6.54 mm w.e. day^{-1}, and a standard deviation of 0.93 mm w.e. day^{-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.

Deleted: Shapley

Deleted: 6

Deleted: 6

Deleted: the model is still capable of identifying the main drivers, consistent with our knowledge of the physical processes.…

Deleted: are the top three dominant drivers.

Deleted: 6

Deleted: Shapley

Deleted: 6

Deleted: Shapley coefficients

Deleted: with albedo gaining a dominant role as a driver

Formatted: Indent: First line: 0"

Deleted: 7

Deleted: 7

Deleted: At the Swiss Camp station, the mean absolute difference between the machine learning model and MAR is 1.17 mm w.e. day^{-1}, with a maximum of 5.97 mm w.e. day^{-1} and a standard deviation of 0.96 mm w.e. Day^{-1}

Deleted: The model performs comparably well at the K-transect S6 station, yielding a mean difference of 0.91 mm w.e. day^{-1}, a maximum of 6.54 mm w.e. day^{-1}, and a standard deviation of 0.93 mm w.e. day^{-1}.

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.

Fig. 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}, $p < 0.01$). 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 example 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 m , 1000 – 2000 m 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 m 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 (~ 6 mm w.e.) than for the region between 1000 m and 2000 m (~ 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 2000m (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

Deleted: After assessing the skills of the MAR-IA emulator or replicating the outputs of the MAR model, we turn our attention on using the ML model to study how different quantities have been driving the surface meltwater production and how such drivers have been changing

Deleted: ure

Deleted: 8

Deleted: Shapley

Deleted: Shapley

Deleted: ure

Deleted: 9

Deleted: values for Shapley coefficients

Deleted: Shapley

Deleted: 0

Deleted: [

Deleted: 0

Deleted: the main drivers are still albedo, air temperature and shortwave downward radiation.

Deleted: Shapley

Deleted: 0

Deleted: Shapley

Deleted: 0

Deleted: downward longwave radiation becomes the third most important driver 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 Fig. 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 shows the – on the left y-axis – the corresponding daily values of the predictors. 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 (Figure 12).

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. day^{-1}, confirming the emulator's fidelity. MAR-IA2 retains strong performance with $R^2 = 0.975$ and RMSE = 1.882 mm w.e. day^{-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. day^{-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. day^{-1}. These results demonstrate that while ERA-based models enable broader applicability, predictor quality—particularly albedo—

Deleted: ures
Deleted: Shapley
Deleted: ure
Deleted: ands
Deleted: kin
Deleted: 9
Deleted: 79
Deleted: (
Deleted: 2.87
Deleted: 7
Deleted: (
Deleted: 6
Deleted: 4
Deleted: (
Deleted: 2
Deleted: 22
Deleted: (

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 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 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

Formatted: Font: Not Bold

Deleted: Explainable AI techniques using SHAP (Shapley Additive Explanations) provide a rigorous framework for interpreting machine learning predictions by quantifying the contribution of each input feature to the model output. We applied SHAP to the MAR-IA emulators to assess whether the model reproduces meltwater dynamics for the right (physical) reasons. The analysis revealed that albedo, downward shortwave radiation, and 2 m air temperature consistently emerge as the dominant drivers of surface meltwater production, in agreement with known physical processes governing the surface energy balance of the GrIS. As expected, albedo exhibits the strongest negative influence, meaning that lower reflectivity amplifies melt, while shortwave radiation and air temperature showed positive contributions, enhancing meltwater generation as their values increase. Longwave radiation also played a significant role, though secondary to shortwave fluxes and temperature. Importantly, SHAP analysis confirmed that the emulator’s predictions are not only accurate but physically interpretable: the ranking and directional influence of predictors align with theory and MAR outputs.¶

Deleted: Albedo remains a high-magnitude control, comparable to temperature in summer mean influence, yet its long-term trend is small and not statistically significant—consistent with the view that recent Greenland albedo variability is substantial but not necessarily expressed as a simple linear intensification in attribution metrics (Feng et al., 2023).

Deleted: Shapley

Deleted: Shapley

Deleted: Above 2000 m, downward longwave radiation becomes comparatively more important, supporting the interpretation that episodic intrusions of warm, moist air and associated longwave enhancement can be critical for high-elevation melt events—again consistent with prior evidence from extreme years such as 2012 (Nghiem et al., 2012). Finally, station-scale daily analyses (K-transect S6 and Swiss Camp in 2012) demonstrate that melt drivers are highly state-dependent: Shapley values change sign and magnitude through the season as boundary-layer structure, radiative conditions, and surface state co-evolve.

Deleted: Shapley

Formatted: Font color: Auto

Deleted: predictands

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.

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.

Code/Data availability

The MAR-IA models and the training datasets can be found here - https://zenodo.org/records/17942115. The code is available at: https://github.com/racheetmatai/MAR_emulator

Author contribution

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

Deleted: Moreover, the MAR-IA emulator allows running past and future scenarios without forcing the atmospheric model in MAR with fields that are not always available for past periods and future simulations. ¶

Deleted: . W

Deleted: This can accelerate the pace of discovery and allows analysis – such as Shapley or similar – that can synthesize the climate model skills, reducing computational cost and promoting synthetic experiments that can help improve our understanding of the past while improving future estimates. ¶

One of the authors is an editor at the Cryosphere.

References.

Agosta, Cécile, Charles Amory, Christoph Kittel, et al. 2019. "Estimation of the Antarctic Surface Mass Balance Using the Regional Climate Model MAR (1979–2015) and Identification of Dominant Processes." *The Cryosphere* 13 (1): 281–96. https://doi.org/10.5194/tc-13-281-2019.

Al-Najjar, Husam AH, Biswajeet Pradhan, Ghassan Beydoun, Raju Sarkar, Hyuck-Jin Park, and Adbullah Alamri. "A novel method using explainable artificial intelligence (XAI)-based Shapley Additive Explanations for spatial landslide prediction using Time-Series SAR dataset." Gondwana Research 123 (2023): 107-124.

Batunacun, Ralf Wieland, Tobia Lakes, and Claas Nendel. "Using Shapley additive explanations to interpret extreme gradient boosting predictions of grassland degradation in Xilingol, China." *Geoscientific model development* 14, no. 3 (2021): 1493-1510.

Bentéjac, Candice, Anna Csörgő, and Gonzalo Martínez-Muñoz. 2021. "A Comparative Analysis of Gradient Boosting Algorithms." *Artificial Intelligence Review* 54 (3): 1937–67.

Bergstra, James, and Yoshua Bengio. 2012. "Random Search for Hyper-Parameter Optimization." *The Journal of Machine Learning Research* 13 (1): 281–305.

Bochow, Nils, Philipp Hess, and Alexander Robinson. "Physics-constrained generative machine learning-based high-resolution downscaling of Greenland's surface mass balance and surface temperature." *arXiv preprint arXiv:2507.22485* (2025).

Box, J. E., X. Fettweis, et al. 2012. "Greenland Ice Sheet Albedo Feedback: Thermodynamics and Atmospheric Drivers." *The Cryosphere* 6 (4): 821–39. https://doi.org/10.5194/tc-6-821-2012.

Brun, Eric, Paul David, Marcel Sudul, and Gilles Brunot. 1992. "A Numerical Model to Simulate Snow-Cover Stratigraphy for Operational Avalanche Forecasting." *Journal of Glaciology* 38 (128): 13–22. Chen, Tianqi, and Carlos Guestrin. 2016b. "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 13, 785–94. https://doi.org/10.1145/2939672.2939785.

Coléou, Cécile, and Bernard Lesaffre. 1998. "Irreducible Water Saturation in Snow: Experimental Results in a Cold Laboratory." *Annals of Glaciology* 26: 64–68.

Delhasse, Alison, Christoph Kittel, Charles Amory, et al. 2020. "Brief Communication: Evaluation of the near-Surface Climate in ERA5 over the Greenland Ice Sheet." *The Cryosphere* 14 (3): 957–65.

Descals, Adrià, Aleixandre Verger, Gaofei Yin, Iolanda Filella, and Josep Peñuelas. "Local interpretation of machine learning models in remote sensing with SHAP: the case of global climate constraints on photosynthesis phenology." *International Journal of Remote Sensing* 44, no. 10 (2023): 3160-3173.

Dikshit, Abhirup, and Biswajeet Pradhan. "Interpretable and explainable AI (XAI) model for spatial drought prediction." *Science of the Total Environment* 801 (2021): 149797.

Doury, Antoine, Samuel Somot, Sebastien Gadat, Aurélien Ribes, and Lola Corre. "Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach." *Climate Dynamics* 60, no. 5 (2023): 1751-1779.

Feng, Shunan, Joseph Mitchell Cook, Alexandre Magno Anesio, Liane G. Benning, and Martyn Tranter. "Long time series (1984–2020) of albedo variations on the Greenland ice sheet from harmonized Landsat and Sentinel 2 imagery." *Journal of Glaciology* 69, no. 277 (2023): 1225-1240.

Fettweis, X. 2007. “Reconstruction of the 1979–2006 Greenland Ice Sheet Surface Mass Balance Using the Regional Climate Model MAR.” *The Cryosphere* 1 (1): 21–40. https://doi.org/10.5194/tc-1-21-2007.

Fettweis, Xavier, B. Franco, M. Tedesco, et al. 2013. “Estimating Greenland Ice Sheet Surface Mass Balance Contribution to Future Sea Level Rise Using the Regional Atmospheric Climate Model MAR.” *The Cryosphere* 7: 469–89. https://doi.org/10.5194/tc-7-469-2013.

Fettweis, Xavier, Jason E. Box, Cécile Agosta, et al. 2017. “Reconstructions of the 1900–2015 Greenland Ice Sheet Surface Mass Balance Using the Regional Climate MAR Model.” *The Cryosphere* 11 (2): 1015–33. https://doi.org/10.5194/tc-11-1015-2017.

Fettweis, Xavier, Stefan Hofer, Uta Krebs-Kanzow, et al. 2020. *GrSMBMIP: Intercomparison of the Modelled 1980–2012 Surface Mass Balance over the Greenland Ice Sheet*. Preprint. Ice sheets/Greenland. https://doi.org/10.5194/tc-2019-321.

Feurer, Matthias, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum, and Frank Hutter. 2015. “Efficient and Robust Automated Machine Learning.” *Advances in Neural Information Processing Systems* 28.

Franco, B., X. Fettweis, C. Lang, and M. Erpicum. 2012. “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 (3): 695–711. https://doi.org/10.5194/tc-6-695-2012.

Freund, Yoav, and Robert E Schapire. 1997. “A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting.” *Journal of Computer and System Sciences* 55 (1): 119–39.

Ghafarian, Fatemeh, Ralf Wieland, Dietmar Lüttschwager, and Claas Nendel. "Application of extreme gradient boosting and Shapley Additive explanations to predict temperature regimes inside forests from standard open-field meteorological data." *Environmental Modelling & Software* 156 (2022): 105466.

Gholamy, Afshin, Vladik Kreinovich, and Olga Kosheleva. 2018. *Why 70/30 or 80/20 Relation between Training and Testing Sets: A Pedagogical Explanation*.

Grinsztajn, Léo, Edouard Oyallon, and Gaël Varoquaux. "Why do tree-based models still outperform deep learning on typical tabular data?." *Advances in neural information processing systems* 35 (2022): 507-520.

Head, Tim, Gilles Louppe, Iaroslav Shcherbatyi, et al. 2018. “Scikit-Optimize/Scikit-Optimize: V0. 5.2.” *Zenodo*.

Hersbach, Hans, Bill Bell, Paul Berrisford, et al. 2020. “The ERA5 Global Reanalysis.” *Quarterly Journal of the Royal Meteorological Society* 146 (730): 1999–2049. https://doi.org/10.1002/qj.3803.

Hofer, Stefan, Andrew J. Tedstone, Xavier Fettweis, and Jonathan L. Bamber. "Decreasing cloud cover drives the recent mass loss on the Greenland Ice Sheet." Science Advances 3, no. 6 (2017): e1700584.

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 72(6), 558-569

Koo, Younghyun, Hongjie Xie, Nathan T. Kurtz, Stephen F. Ackley, and Wei Wang. "Sea ice surface type classification of ICESat-2 ATL07 data by using data-driven machine learning model: Ross Sea, Antarctic as an example." *Remote Sensing of Environment* 296 (2023): 113726.

Langen, Peter L, Robert S Fausto, Baptiste Vandecrux, Ruth H Mottram, and Jason E Box. 2017. "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.

Larour, E., Seroussi, H., Morlighem, M., & Rignot, E. (2012). Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM). *Journal of Geophysical Research: Earth Surface*, 117(F1).

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44.

Lenaerts, JTM, MR Van den Broeke, JH Van Angelen, E Van Meijgaard, and SJ Déry. 2012. "Drifting Snow Climate of the Greenland Ice Sheet: A Study with a Regional Climate Model." *The Cryosphere* 6 (4): 891–99.

Lenaerts, Jan TM, Brooke Medley, Michiel R. van den Broeke, and Bert Wouters. "Observing and modeling ice sheet surface mass balance." Reviews of Geophysics 57, no. 2 (2019): 376-420.

Li, Kexuan, Fangfang Wang, Lingli Yang, and Ruiqi Liu. 2023. "Deep Feature Screening: Feature Selection for Ultra High-Dimensional Data via Deep Neural Networks." *Neurocomputing* 538: 126186.

Lundberg, Scott, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *arXiv:1705.07874 [Cs, Stat]*, November 24. http://arxiv.org/abs/1705.07874.

Lütjens, Björn, Patrick Alexander, Raf Antwerpen, Til Widmann, Guido Cervone, and Marco Tedesco. "MeltwaterBench: Deep learning for spatiotemporal downscaling of surface meltwater." *arXiv preprint arXiv:2512.12142* (2025).

Ma, Meihong, Gang Zhao, Bingshun He, et al. 2021. "XGBoost-Based Method for Flash Flood Risk Assessment." *Journal of Hydrology* 598: 126382.

Materia, Stefano, Lluís Palma García, Chiem van Straaten, et al. 2024. "Artificial Intelligence for Climate Prediction of Extremes: State of the Art, Challenges, and Future Perspectives." *Wiley Interdisciplinary Reviews: Climate Change* 15 (6): e914.

Mayr, A.; Binder, H.; Gefeller, O.; Schmid, M. 2018. "The Evolution of Boosting Algorithms." *From Machine Learning to Statistical Modelling* 53 (06): 419–27. https://doi.org/10.3414/ME13-01-0122.

Mioduszewski, J. R., A. K. Rennermalm, D. A. Robinson, and T. L. Mote. "Attribution of snowmelt onset in Northern Canada." *Journal of Geophysical Research: Atmospheres* 119, no. 16 (2014): 9638-9653.

Neff, William, Gilbert P. Compo, F. Martin Ralph, and Matthew D. Shupe. "Continental heat anomalies and the extreme melting of the Greenland ice surface in 2012 and 1889." *Journal of Geophysical Research: Atmospheres* 119, no. 11 (2014): 6520-6536.

Nghiem, S. V., D. K. Hall, T. L. Mote, Marco Tedesco, M. R. Albert, K. Keegan, C. A. Shuman, N. E. DiGirolamo, and G. Neumann. "The extreme melt across the Greenland ice sheet in 2012." *Geophysical Research Letters* 39, no. 20 (2012).

Noël, B., W. J. van de Berg, E. van Meijgaard, P. Kuipers Munneke, R. S. W. van de Wal, and M. R. van den Broeke. 2015. "Summer Snowfall on the Greenland Ice Sheet: A Study with the Updated Regional Climate Model RACMO2.3." *The Cryosphere Discuss.* 9 (1): 1177–208. https://doi.org/10.5194/tcd-9-1177-2015.

Noël, Brice, Willem Jan van de Berg, J. Melchior van Wessem, et al. 2018. "Modelling the Climate and Surface Mass Balance of Polar Ice Sheets Using RACMO2 – Part 1: Greenland (1958–2016)." *The Cryosphere* 12 (3): 811–31. https://doi.org/10.5194/tc-12-811-2018.

Reichstein, Markus, Gustau Camps-Valls, Bjorn Stevens, et al. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566 (7743): 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Rohmer, Jeremy, Remi Thieblemont, Goneri Le Cozannet, Heiko Goelzer, and Gael Durand. "Improving interpretation of sea-level projections through a machine-learning-based local explanation approach." *The Cryosphere* 16, no. 11 (2022): 4637-4657.

Schlager, Elke, Sebastian Scher, Ruth H. Mottram, and Peter L. Langen. "Learning to melt: Emulating Greenland surface melt from a polar RCM with machine learning." *EGUsphere* 2026 (2026): 1-24.

Sivakumar, Muthuramalingam, Sudhaman Parthasarathy, and Thiyagarajan Padmapriya. 2024. "Trade-off between Training and Testing Ratio in Machine Learning for Medical Image Processing." *PeerJ Computer Science* 10: e2245.

Smith, Benjamin E, Brooke Medley, Xavier Fettweis, et al. 2023. "Evaluating Greenland Surface-Mass-Balance and Firn-Densification Data Using ICESat-2 Altimetry." *The Cryosphere* 17 (2): 789–808.

Snoek, Jasper, Hugo Larochelle, and Ryan P Adams. 2012. "Practical Bayesian Optimization of Machine Learning Algorithms." *Advances in Neural Information Processing Systems* 25.

Tedesco, M., S. Doherty, X. Fettweis, P. Alexander, J. Jeyaratnam, and J. Stroeve. 2016. "The Darkening of the Greenland Ice Sheet: Trends, Drivers and Projections (1981-2100)." *The Cryosphere* 10: 477–96. https://doi.org/10.5194/tc-10-477-2016.

Tedesco, Marco, T Mote, Xavier Fettweis, et al. 2016. "Arctic Cut-off High Drives the Poleward Shift of a New Greenland Melting Record." *Nature Communications* 7 (1): 11723.

Tedesco, Marco, Paolo Colosio, Xavier Fettweis, and Guido Cervone. 2023. "A Computationally Efficient Statistically Downscaled 100 m Resolution Greenland Product from the Regional Climate Model MAR." *The Cryosphere Discussions* 2023: 1–27.

Van den Broeke, Michiel R, Ellyn M Enderlin, Ian M Howat, et al. 2016. "On the Recent Contribution of the Greenland Ice Sheet to Sea Level Change." *The Cryosphere* 10 (5): 1933–46.

Veldhuijsen, Sanne BM, Willem Jan Van De Berg, Peter Kuipers Munneke, Nicolaj Hansen, Fredrik Boberg, Christoph Kittel, Charles Amory, and Michiel R. Van Den Broeke. "Emulating the expansion of Antarctic perennial firn aquifers in the 21st century." *The Cryosphere* 19, no. 10 (2025): 5157-5173.

Wang, Wenshan, Charles S. Zender, Dirk van As, Robert S. Fausto, and Matthew K. Laffin. "Greenland surface melt dominated by solar and sensible heating." Geophysical Research Letters 48, no. 7 (2021): e2020GL090653.

Xu, Jingjing, Caesar Wu, Yuan-Fang Li, Gregoire Danoy, and Pascal Bouvry. 2024. "Survey and Taxonomy: The Role of Data-Centric AI in Transformer-Based Time Series Forecasting." *arXiv Preprint arXiv:2407.19784*.

Zamani Joharestani, Mehdi, Chunxiang Cao, Xiliang Ni, Barjeece Bashir, and Somayeh Talebiesfandarani. 2019. “PM2. 5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data.” *Atmosphere* 10 (7): 373. Zhang, Qing-Lin, Ming-Hu Ding, Michiel R. Van Den Broeke, Brice NoËl, Xavier Fettweis, Sai Wang, Wei-Jun Sun et al. "Variations in Greenland surface melt and extreme events from 1958 to 2023." *Advances in Climate Change Research* 16, no. 5 (2025): 910-921.

Figures and tables

Model	Predictors					
	Ancillary	Temp. [K]	Radiation [W/m2]	Fluxes [w/m2]	Source	Albedo
MAR-IA1	Lon., Lat., Elev.	Surf. Temp., 2m Air Temp.	SW Rad., LW Rad., SW Up, LW Up	Sens. Heat Flux, Lat. Heat Flux	MAR	MAR
MAR-IA2	Lon., Lat., Elev.	2m Air Temp.	SW Rad., LW Rad.	Sens. Heat Flux, Lat. Heat Flux	MAR	MAR
MAR-IA2_no_alb	Lon., Lat., Elev.	2m Air Temp.	SW Rad., LW Rad.	Sens. Heat Flux, Lat. Heat Flux	MAR	None
MAR-IA2ERA	Lon., Lat., Elev.	2m 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.	2m 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.	2m Air Temp. ERA	SW Rad. ERA, LW Rad. ERA	Sens. Heat Flux ERA, Lat. Heat Flux ERA	ERA	MAR

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

	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)

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.

Variable	Group	Slope (per year)	R²	p-value
2m 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

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.

Commented [1]: updated

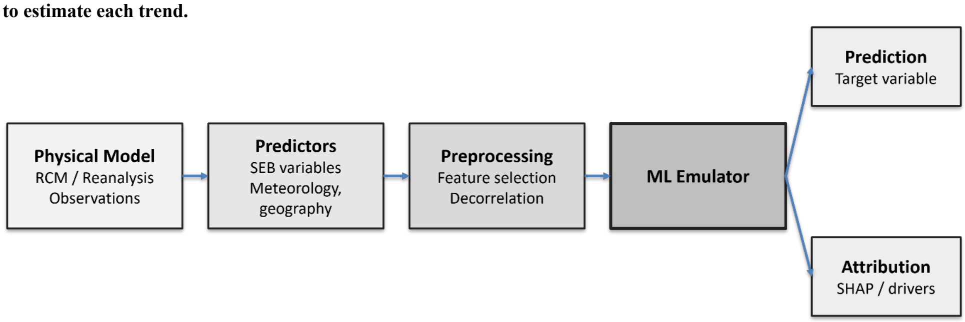


Figure 1: Conceptual overview of the MAR-IA framework.

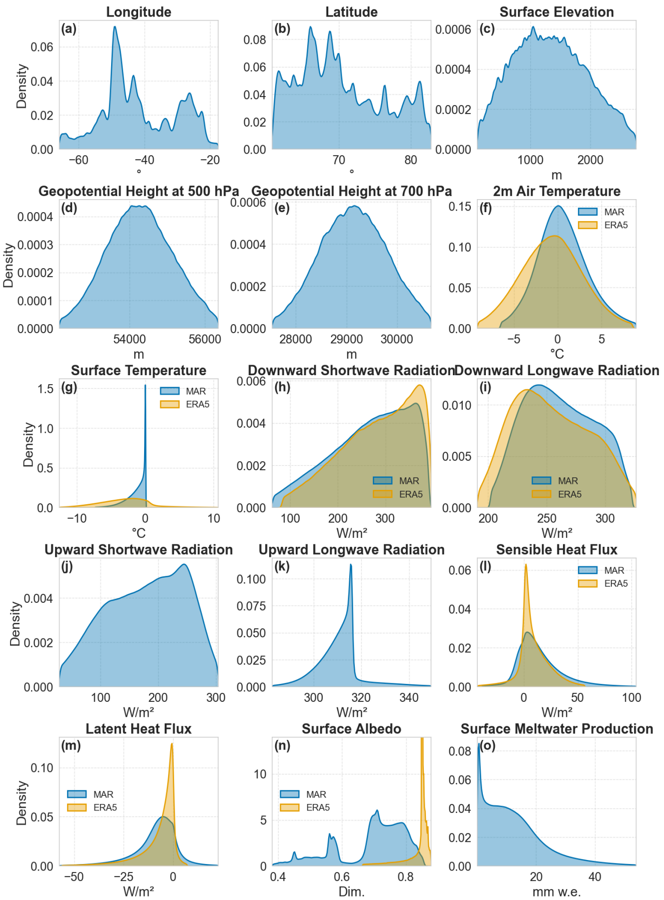


Figure 2: Distributions of MAR (blue) and ERA-5 (yellow) variables used for training the different models in this study.

Commented [2]: updated

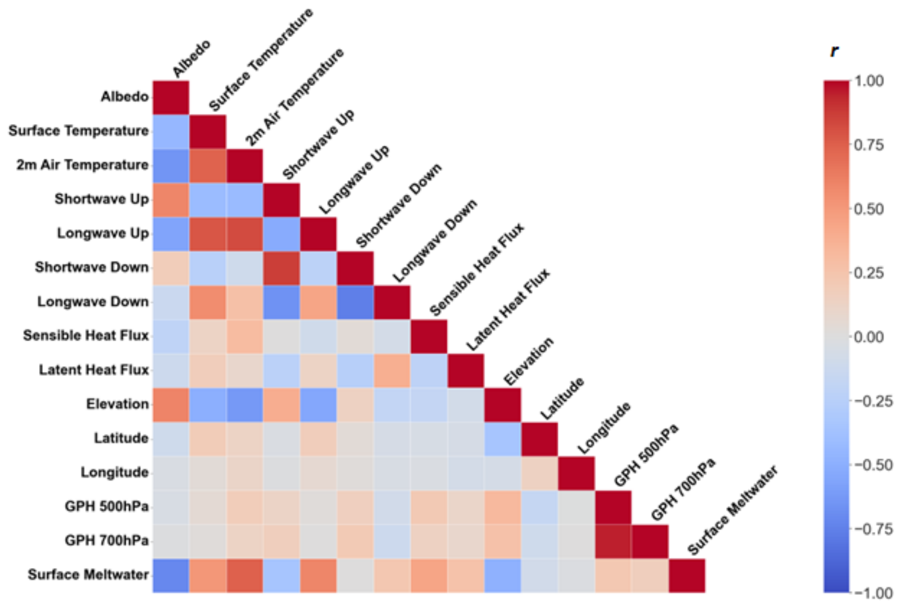
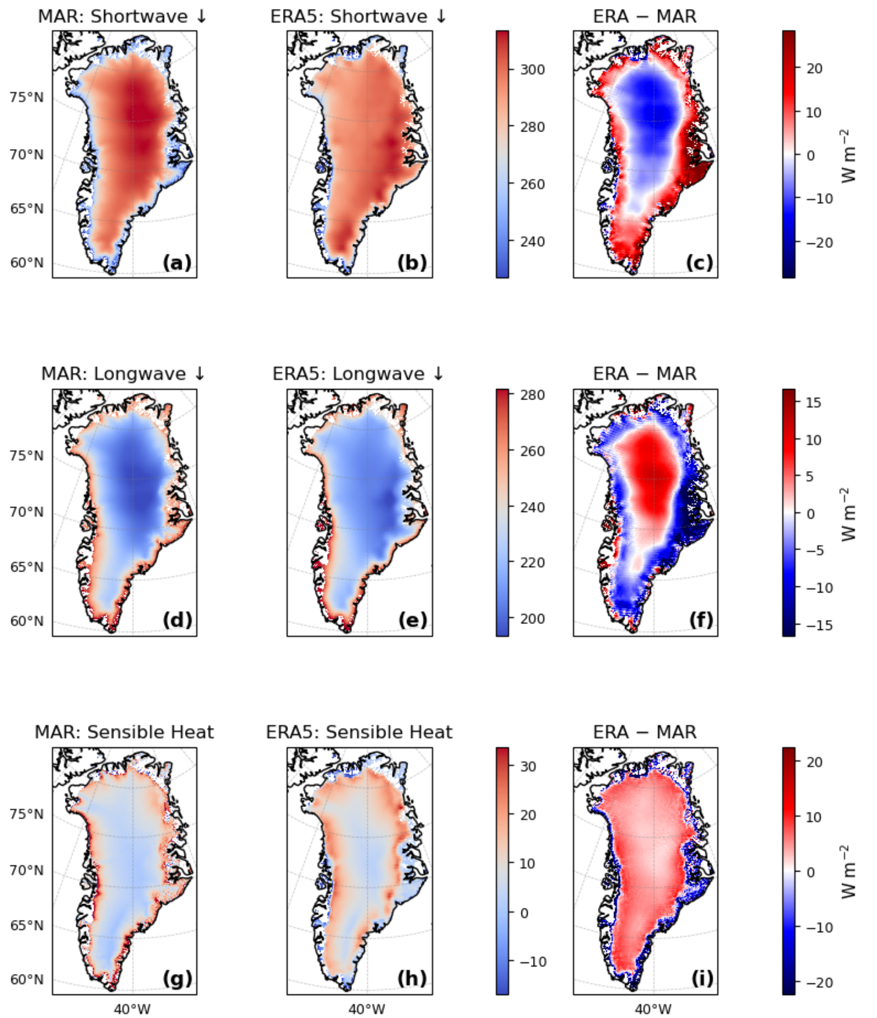


Figure 3: Heatmap of correlation coefficient (r) between the predictors and the predictand for the datasets used in the models' training.



MAR: Shortwave ↓
ERA5: Shortwave ↓
ERA − MAR
MAR: Longwave ↓
ERA5: Longwave ↓
ERA − MAR
MAR: Sensible Heat
ERA5: Sensible Heat
ERA − MAR
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
75°N
70°N
65°N
60°N
40°W
W m−2

1036

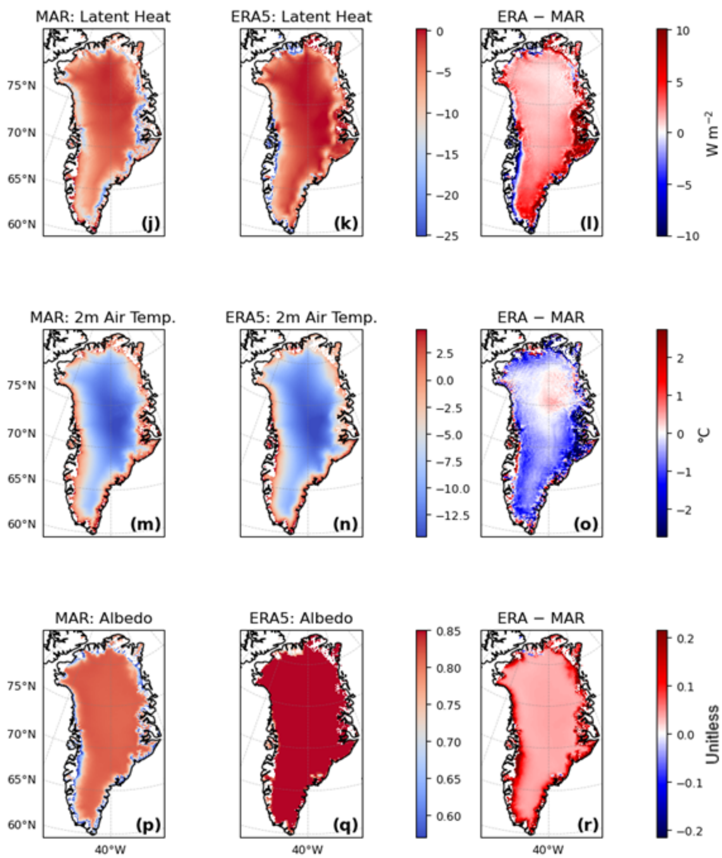


Figure 4: 1979 - 2024 summer (June - July - August, JJA) averaged values for the predictors used in the models' training (left and center columns) and differences between the MAR and ERA-5 values (right column).

Commented [3]: updated

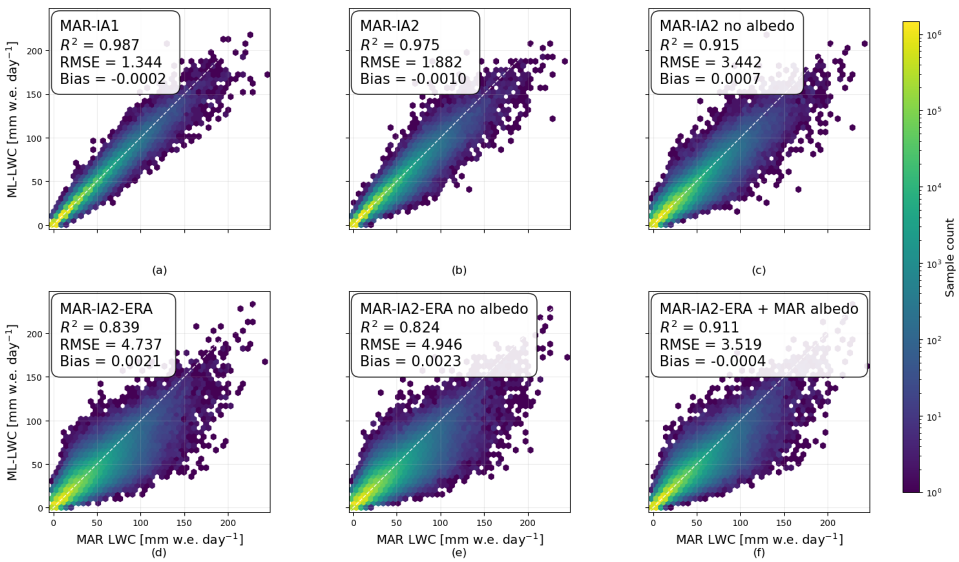


Figure5: Scatterplot of the MAR (x-axis) and MAR-IA simulated (y-axis) surface meltwater production in the case of MAR-IA1 (a), MAR-IA2 (b), MAR-IA2_no_alb (c), MAR-IA2ERA (d), MAR-IA2ERA_no_alb (e), MAR-IA2ERA_mar_alb (f). Note only the test data statistics are reported and plotted.

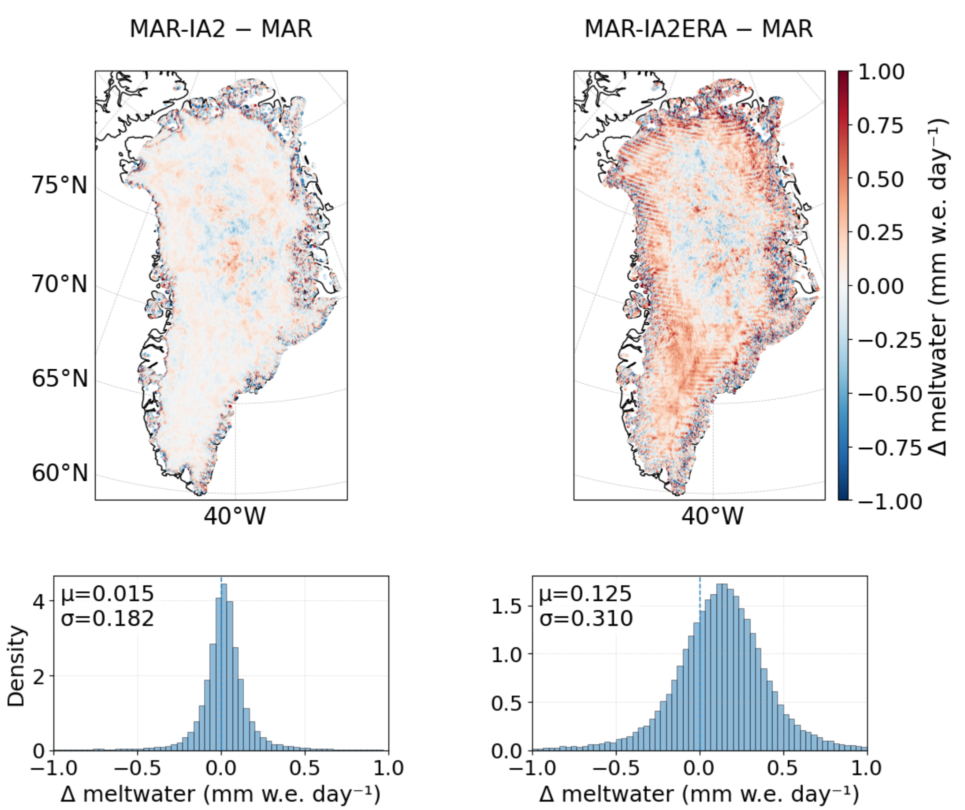
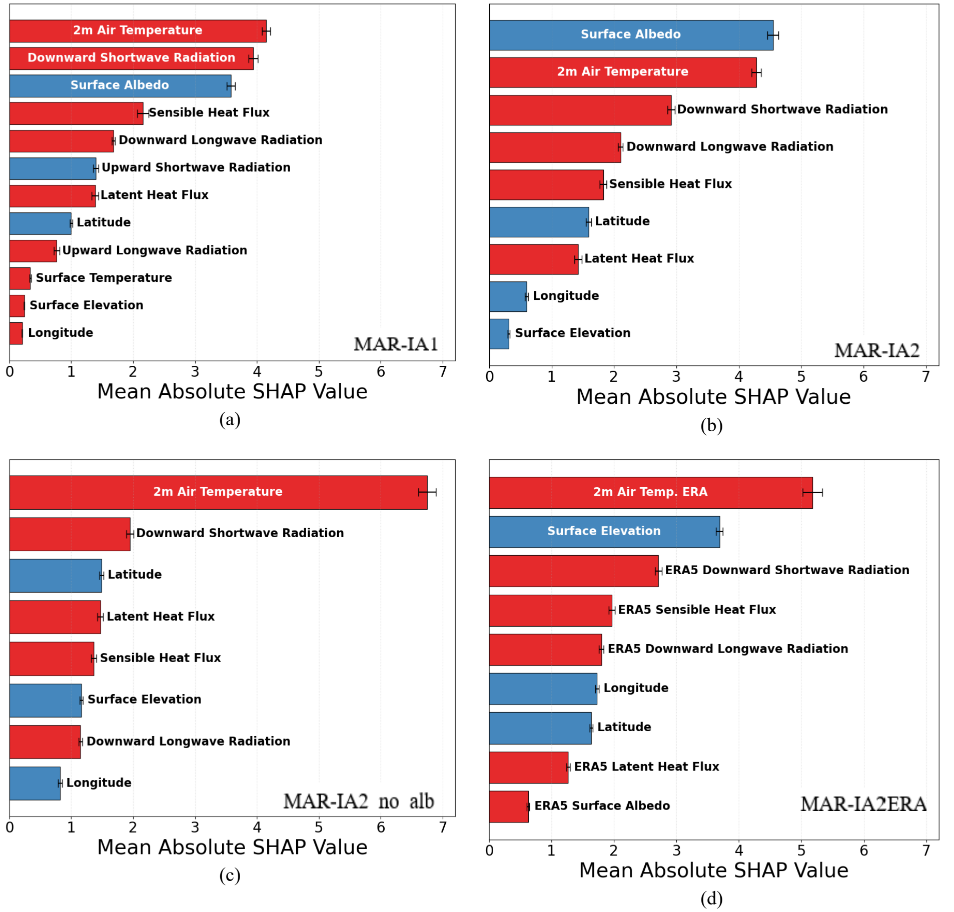
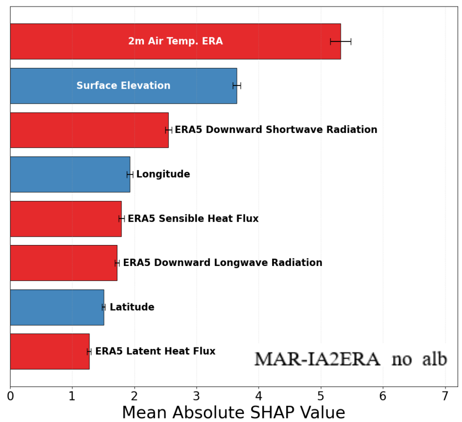


Figure 6: Mean difference in meltwater production between the machine-learning emulator and MAR for (left) MAR-IA2 and (right) MAR-IA2ERA configurations. Colors show the spatial distribution of the mean daily meltwater difference (ML − MAR) at each MAR grid point, obtained by averaging over all available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

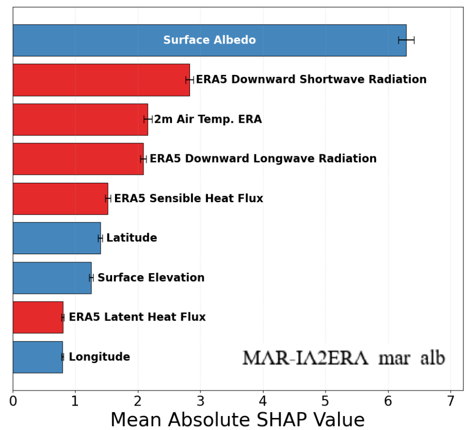
Commented [U4]: Some of these are aligned to the center and some to the left. Not sure what the paper requirements are. Pls check.


2m Air Temperature
Downward Shortwave Radiation
Surface Albedo
Sensible Heat Flux
Downward Longwave Radiation
Upward Shortwave Radiation
Latent Heat Flux
Latitude
Upward Longwave Radiation
Surface Temperature
Surface Elevation
Longitude
MAR-IA1
Mean Absolute SHAP Value
Surface Albedo
2m Air Temperature
Downward Shortwave Radiation
Downward Longwave Radiation
Sensible Heat Flux
Latitude
Latent Heat Flux
Longitude
Surface Elevation
MAR-IA2
Mean Absolute SHAP Value
2m Air Temperature
Downward Shortwave Radiation
Latitude
Latent Heat Flux
Sensible Heat Flux
Surface Elevation
Downward Longwave Radiation
Longitude
MAR-IA2 no alb
Mean Absolute SHAP Value
2m Air Temp. ERA
Surface Elevation
ERA5 Downward Shortwave Radiation
ERA5 Sensible Heat Flux
ERA5 Downward Longwave Radiation
Longitude
Latitude
ERA5 Latent Heat Flux
ERA5 Surface Albedo
MAR-IA2ERA
Mean Absolute SHAP Value

(a) (b) (c) (d)



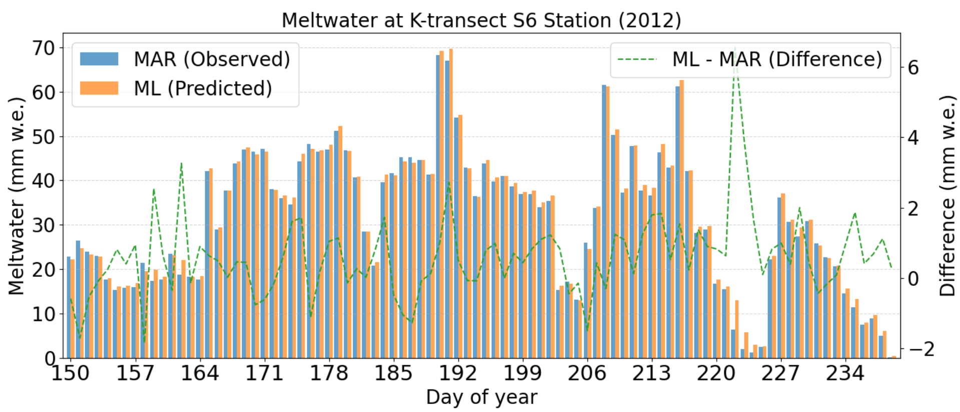
(e)



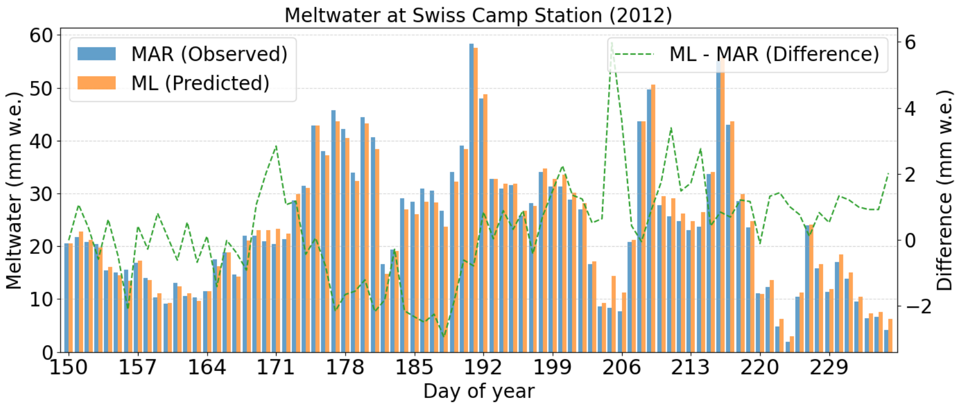
(f)

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.

Commented [5]: updated



(a)



(b)

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.

Commented [6]: Coordinates:
----
location_name = "Swiss Camp Station"
lat = 69.5644
long = -49.3272

closest points in simulations:
Latitude : 69.5390396118164
Longitude : -49.3537713989257805
----
location_name = "K-transect S6 Station"
lat = 67.079
long = -49.407

closes points in simulation
Latitude : 67.07711791992188
Longitude : -49.445999145507805

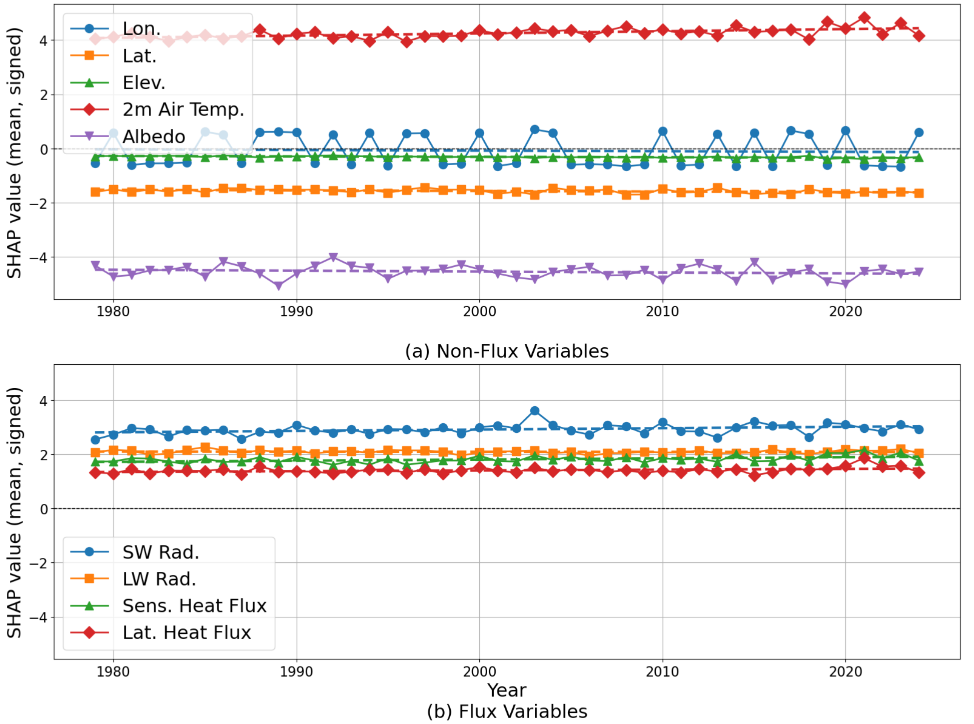


Figure 9: Summer (JJA) mean Shapley values for the (a) non-flux and (b) flux variables for the period 1979 - 2024. Linear trends are reported also as dashed lines with the same color as the corresponding variables. Note this is for MAR-IA2.

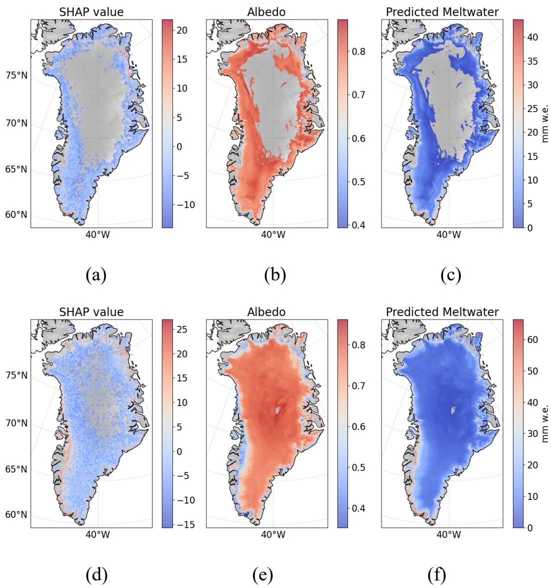


Figure 10: Summer-averaged spatial distribution of Shapley (a,d), albedo (b,e) and ML-estimated meltwater (c,f) for the 1992 (a,b,c) and 2012 (d, e, f) for the MAR-IA2 model.

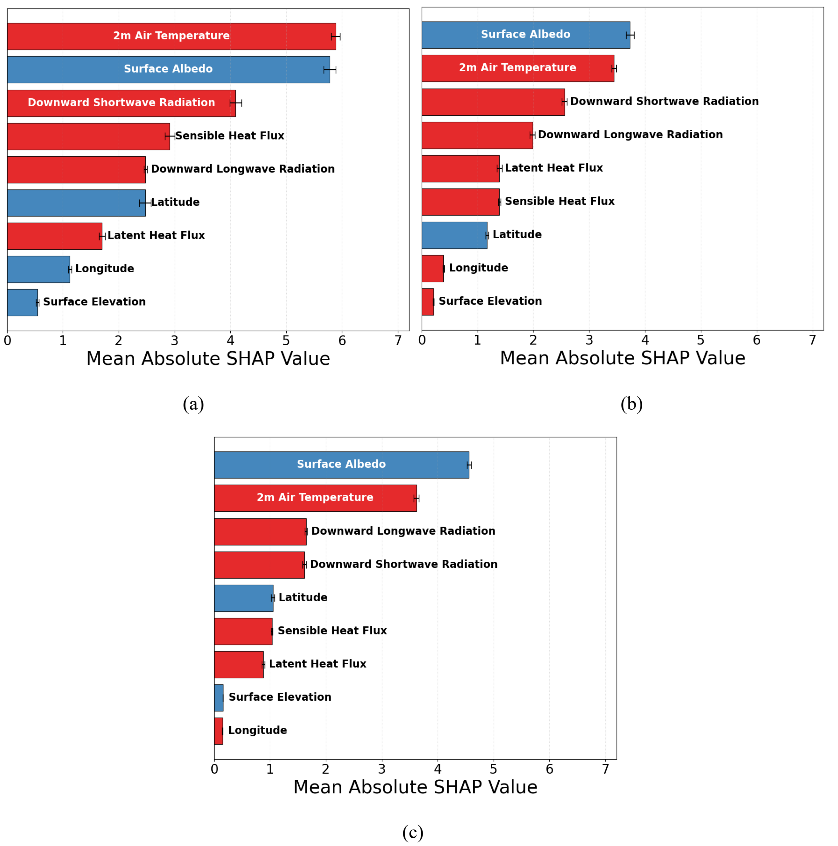


Figure 11: Mean absolute Shapley values obtained for the MAR-IA2 model for areas a) below 1000 m, b) between 1000 and 2000 m nd c) above 2000 m.

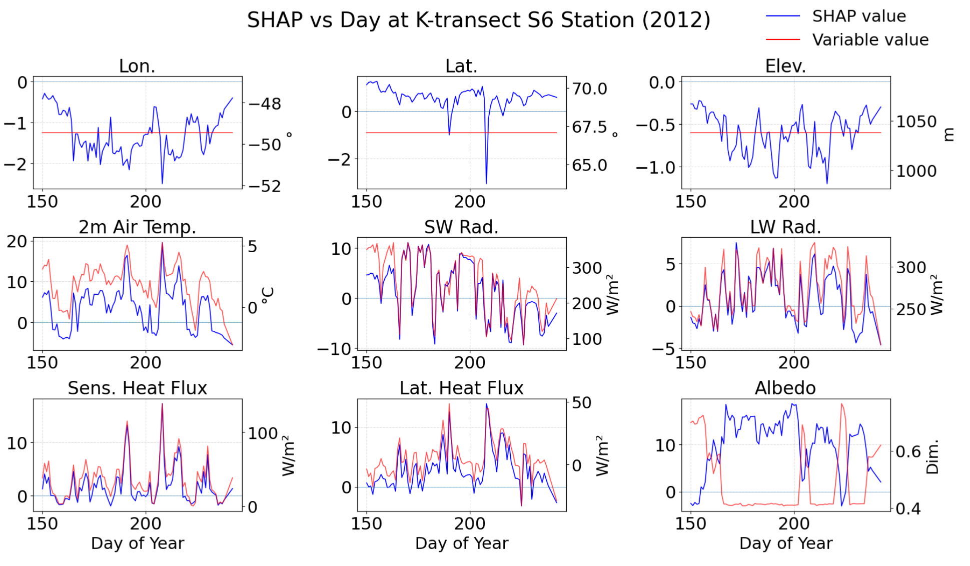


Fig.12: Daily Shapley values for the different predictors obtained from the model MAR-IA1 and associated trends (dashed red line) for the pixel containing the K-transect S6 station.

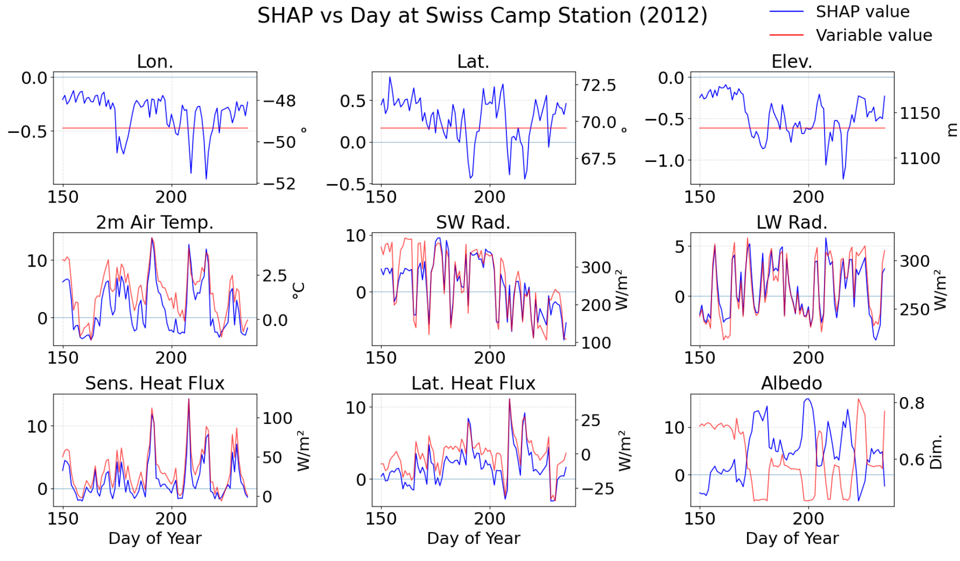


Fig. 13: Same as Fig. 8 but for the Swiss Camp location.

1098

Page 5: [1] Deleted	Desktop_2	3/14/26 3:45:00 PM