https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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, 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 variables 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 test-set performance up to $R^2 = 0.99$ *with low mean squared error and negligible bias relative to MAR meltwater outputs. We apply a SHAP-based explainable AI analysis 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. Development of regional climate models should go hand in hand with ML-based tools.*

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., 2016a). In this regard, it is crucial to understand the physical processes and feedbacks driving enhanced 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., 2016a).

Regional climate models (RCMs), including the *Modèle Atmosphérique Régional* (MAR, Fettweis et al., 2013), RACMO (Noël et al., 2018), and HIRHAM (Langen et al., 2017), have substantially advanced our ability to simulate surface melt and better capture the driving processes. Nevertheless, disentangling the complex, non-linear relationships among surface energy

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

balance (SEB) components—such as albedo, surface temperature, and shortwave and longwave radiation—and their influence on meltwater production remains computationally demanding and formally ill-posed (Mioduszewski et al. 2014 ;Pirk et al. 2022). 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 promoting attribution analyses (Reichstein et al., 2019; Doury et al. 2023). Tree-based algorithms such as XGBoost (Chen and Guestrin, 2016b), deep convolutional neural networks (LeCun et al., 2015), and ensemble approaches (Materia et al., 2024) have demonstrated strong predictive capabilities in Earth system applications. 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). 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.

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. Predictands are drawn from SEB components and near-surface meteorological fields. 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 variables available from widely used datasets such as ERA5 (Hersbach et al., 2020), enhancing applicability beyond MAR-specific outputs. 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). 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. 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.

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

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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). Topographic variables—elevation, latitude, and longitude—also modulate the local energy balance and atmospheric conditions (Tedesco et al., 2016a, 2016b; Fettweis et al. 2017). These variables are used to train our model, as explained in Section 2.3.

2.2 The XGBoost algorithm

Machine learning algorithms are currently widely used for many tasks (e.g., Bentéjac et al. 2021). Among these, the eXtreme Gradient Boosting (XGBoost) algorithm has become one of the most popular in view of its skills related to, accuracy, ability to handle a variety of data types, and to the capacity for processing large datasets (Chen and Guestrin 2016a).

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 predictions from multiple decision trees are combined within the XGBoost to obtain a higher predictive accuracy (Chen and Guestrin 2016a; Bentéjac et al. 2021). Two concepts are key to describing how XGBoost works: a) boosting and b) decision trees. Boosting iteratively improves models by adding new ones that correct previous errors (Mayr 2018). Here, each new model is trained on the misclassified samples by the previous model, with the final prediction being 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).

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

2.3 Attribution analysis and Shapley coefficients

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 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 study Greenland surface mass balance. Importance is summarized using mean absolute SHAP values, which capture the magnitude of contribution irrespective of direction; to convey a dominant directional tendency, assign is also computed from the Pearson correlation between feature values and their SHAP values across samples. To assess long-term changes in melt drivers, we have also computed annual signed SHAP attributions for each variable and fitted linear trends to the resulting time series.

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., 2016a; 2023).

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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 feature-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 distribution of the predictors (Fig. 1). 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.

We also studied the correlation among predictands (Fig. 2). 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 ($r > 0.9$. Specifically, we identified strong relationships between 2 m air and skin temperature ($r \sim 0.90$). 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. 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 skin 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. 3 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 3), 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 Bayesian optimization (Snoek et al., 2012), the latter using probabilistic models to explore the parameter space efficiently.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

For this study, we applied Bayesian optimization with five-fold cross-validation to tune a subset of key XGBoost parameters known to affect predictive accuracy (Chen and Guestrin, 2016a). 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 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. Starting from n_estimators = 100, performance was R^2 = 0.938, MSE = 1.17 mm w.e. day^{-1}, and bias = -1.0×10^{-4} mm w.e. day^{-1}. The global optimum occurred at n_estimators = 2500, achieving R^2 = 0.996, MSE = 0.062 (mm w.e. $day^{-1})^2$, 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 = 0.98, MSE = 1.82 (mm w.e. $day^{-1})^2$, and bias = 0.0006 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. We found the following optimal values: max_depth = 15, min_child_weight = 4, learning_rate = 0.05, and gamma = 0.

3. Results

3.1 Models' performances

All machine-learning (ML) models were trained on 80% of the whole dataset, with 20 % reserved 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 4 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. Performance metrics are summarised in Table 2 using the coefficient of determination (R^2), mean squared error (MSE), and prediction bias. R^2 quantifies the fraction of variance explained by the model (higher is better), MSE measures the average squared deviation between predictions and observations

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

(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.

As expected, the MAR-IA1 model (Fig. 4a), which includes the full predictor set, achieves the highest accuracy ($R^2 = 0.98$, RMSE = 1.38 mm w.e.day^{-1}), with an extremely low bias (-8×10^{-4} mm w.e. day^{-1}). Excluding skin temperature and upwelling longwave and shortwave radiation (MAR-IA2) has little effect on R^2 (0.97) but results in a higher RMSE of 1.887 mm w.e. day^{-1}, reflected in the greater spread for low melt values (< 20 mm w.e.; Fig. 4b). Using ERA-based predictors (MAR-IA2-ERA) further reduces performance ($R^2 = 0.83$, RMSE = 4.746 mm w.e. day^{-1}), with larger errors also concentrated at low melt rates (Fig. 4d). The small difference between MAR-IA2-ERA (Fig. 4d) and MAR-IA2-ERA-no_alb (Fig 4e, 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. 4f), confirming the deterioration of the ML emulator because of the lack of granular information on albedo within the ERA product.

The spatial distribution of mean annual meltwater differences between the ML emulator and MAR is shown in Fig. 5 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 Feature Importance and testing the model's explanatory nature

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.

In the case of MAR-IA1 (Fig. 6a), 2m air temperature shows the highest Shapley 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}). 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. 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 skin

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

temperature is relatively small and the relationship can be explained by the fact that the skin temperature saturates to 0ºC when melting occurs, hence providing no sensitivity to melting and showing negative values when no melting occurs.

The Shapley values for the MAR-IA2 (Fig. 6b) 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. 6c) show that, also in the case of an ML model without albedo, the model is still capable of identifying the main drivers, consistent with our knowledge of the physical 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 top three dominant drivers. 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. 6e), we observe that the Shapley values are only mildly affected by the removal of the albedo. Furthermore, Fig. 6f shows the Shapley coefficients 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 as a driver.

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. 7a) and Swiss Camp (Fig. 7b), 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 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}, 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.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

4. Discussion

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 spatially and temporally.

Figure 8 shows the summer (JJA) mean Shapley 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}, $p = 0.01$), 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}, $p = 0.03$), suggesting a rising role of moisture-related surface processes, while longwave radiation shows no meaningful directional trend ($p = 0.63$). 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. As a further example of the relationships between Shapley values and the surface energy balance quantities, Figure 9 shows the distribution of the averaged summed values for Shapley coefficients, 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 Shapley analysis for different elevation bands (Fig. 10). 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. 10a shows that for areas between 0 and 1000 m and those between 1000 m and 2000 m the main drivers are still albedo, air temperature and shortwave downward 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 Shapley 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. 10b), albedo becomes the dominant driver, though the Shapley values are comparable to those for air temperature. For areas above 2000m (Fig. 10c), downward longwave radiation becomes the third most important driver, still after albedo and surface temperature. This is consistent with recent studies (Nghiem et al. 2012) showing that the intrusion of warm, moist air to the top of the ice sheet was responsible for surface melting.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

Lastly, in Figures 11 and 12 we show the temporal evolution of daily Shapley values at the K-transect S6 (Figure 11) and Swiss Camp (Figure 12) stations for the year 2012 for the predictands used in MAR-IA2. The figures also shows the – on the left y-axis – the corresponding daily values of the predictands. Over the course of the season, the magnitude and sign of individual Shapley values fluctuate, reflecting changing physical conditions and the covariate dependency of melt drivers. Most Shapley quantities covary with the predictand value, and the relationship between Shapley 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 Shapley 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 Shapley coefficients 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 skin 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, with $R^2 = 0.99$ and mean squared error (MSE) = 1.60 $(\text{mm w.e. day}^{-1})^2$, confirming the emulator's fidelity. MAR-IA2 retains strong performance with $R^2 = 0.97$ and MSE = 2.87 $(\text{mm w.e. day}^{-1})^2$, despite using fewer predictors. In contrast, MAR-IA2-ERA shows reduced accuracy, with $R^2 = 0.86$ and MSE = 19.17 $(\text{mm w.e. day}^{-1})^2$, primarily due to ERA's lack of albedo representation. When MAR albedo replaces ERA albedo, performance improves markedly to $R^2 = 0.92$ and MSE = 10.84 $(\text{mm w.e. day}^{-1})^2$. 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.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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.

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 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). The comparison of Shapley summer maps for low-melt (1992) and extreme-melt (2012) years further illustrates how the emulator and Shapley 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, 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. The results concerning albedo values and their Shapley 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 predictands or predictors such as snowpack properties, cloud microphysics, and atmospheric

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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 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.

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 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.

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.

Author contribution

MT conceived the study, prepared the datasets for the ML tool, co-created the ML tools and anaslyzed the results. XF provided MAR outputs. RM supported the development and analysis of ML results.

Competing interests

One of the authors is an editor at the Cryosphere.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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.

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. 2016a. “Xgboost: A Scalable Tree Boosting System.” 785–94.

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.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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*.

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.

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.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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.

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.

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.

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.

Pirk, Norbert, Kristoffer Aalstad, Sebastian Westermann, Astrid Vatne, Alouette van Hove, Lena Merete Tallaksen, Massimo Cassiani, and Gabriel Katul. "Inferring surface energy fluxes using drone data assimilation in large eddy simulations." *Atmospheric Measurement Techniques* 15, no. 24 (2022): 7293-7314.

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.

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.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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, 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.

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

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

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.

Watson, David, Joshua O'Hara, Niek Tax, Richard Mudd, and Ido Guy. "Explaining predictive uncertainty with information theoretic shapley values." *Advances in Neural Information Processing Systems* 36 (2023): 7330-7350.

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.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

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[mm w.e.]	RMSE 95% CI	Bias [mm w.e.]	Bias 95% CI
MAR-IA1	0.98	1.90/1.379	(1.375, 1.383)	-0.0008	(-0.0017, 9.5905e-05)
MAR-IA2	0.97	3.56/1.887	(1.883, 1.892)	0.0004	(-0.0008, 0.0016)
MAR-IA2_no_alb	0.91	11.91/3.451	(3.446, 3.456)	0.0014	(-0.0007, 0.0037)
MAR-IA2ERA	0.83	22.53/4.746	(4.740, 4.752)	0.0012	(-0.0018, 0.0043)
MAR-IA2ERA_no_alb	0.82	24.52/4.952	(4.946, 4.958)	0.0010	(-0.0021, 0.0043)
MAR-IA2ERA_mar_alb	0.91	12.41/3.522	(3.516, 3.529)	0.0011	(-0.0011, 0.0034)

Table 2: Metrics (R^2, MSE, RMSE, ias and 95% CI for RMSE and Bias) of the performances for each trained model obtained using the predictors configuration described in Table 1.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

527

Variable	Group	Slope (per year)	R^2	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

528 **Table 3: Trend (per year) of the mean summer Shapley coefficients reported in Fig. 9 together with the**
529 **corresponding R^2 value.**

530

531

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

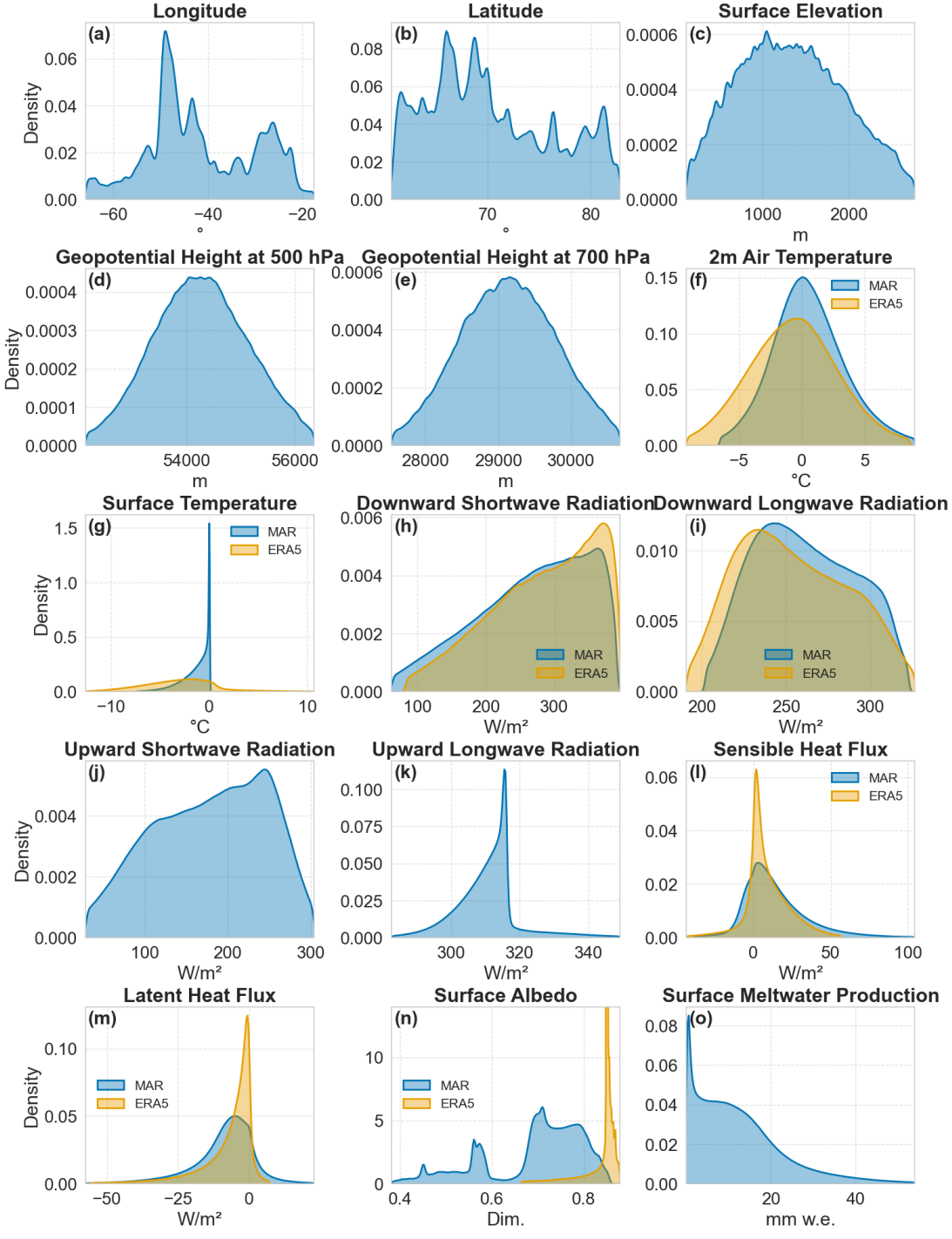
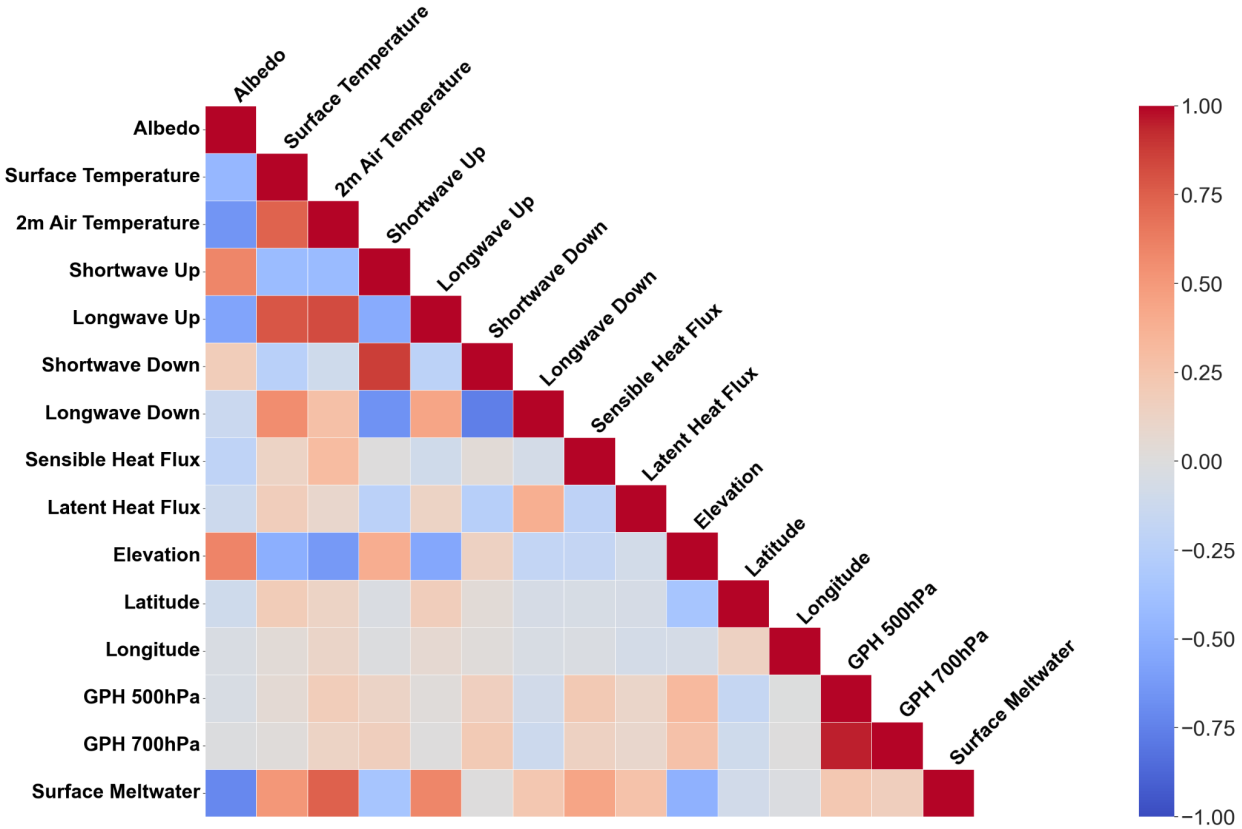


Figure 1: Distributions of MAR (blue) and ERA-5 (yellow) variables used for training the different models in this study.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

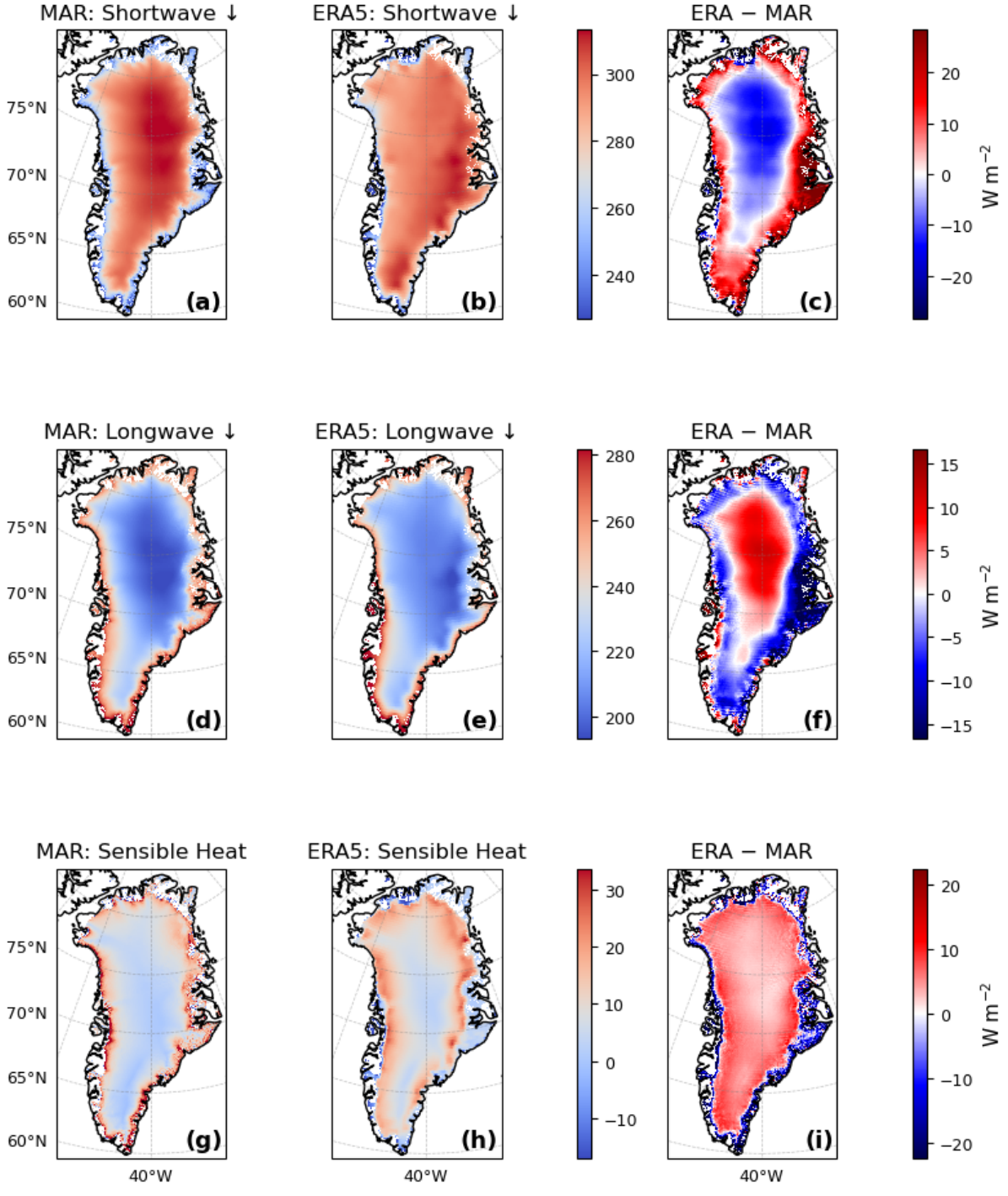


535

536 **Figure 2. Heatmap of correlation coefficient (r) between the predictors and the predictand for the datasets**
537 **used in the models' training.**

538

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026



539

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

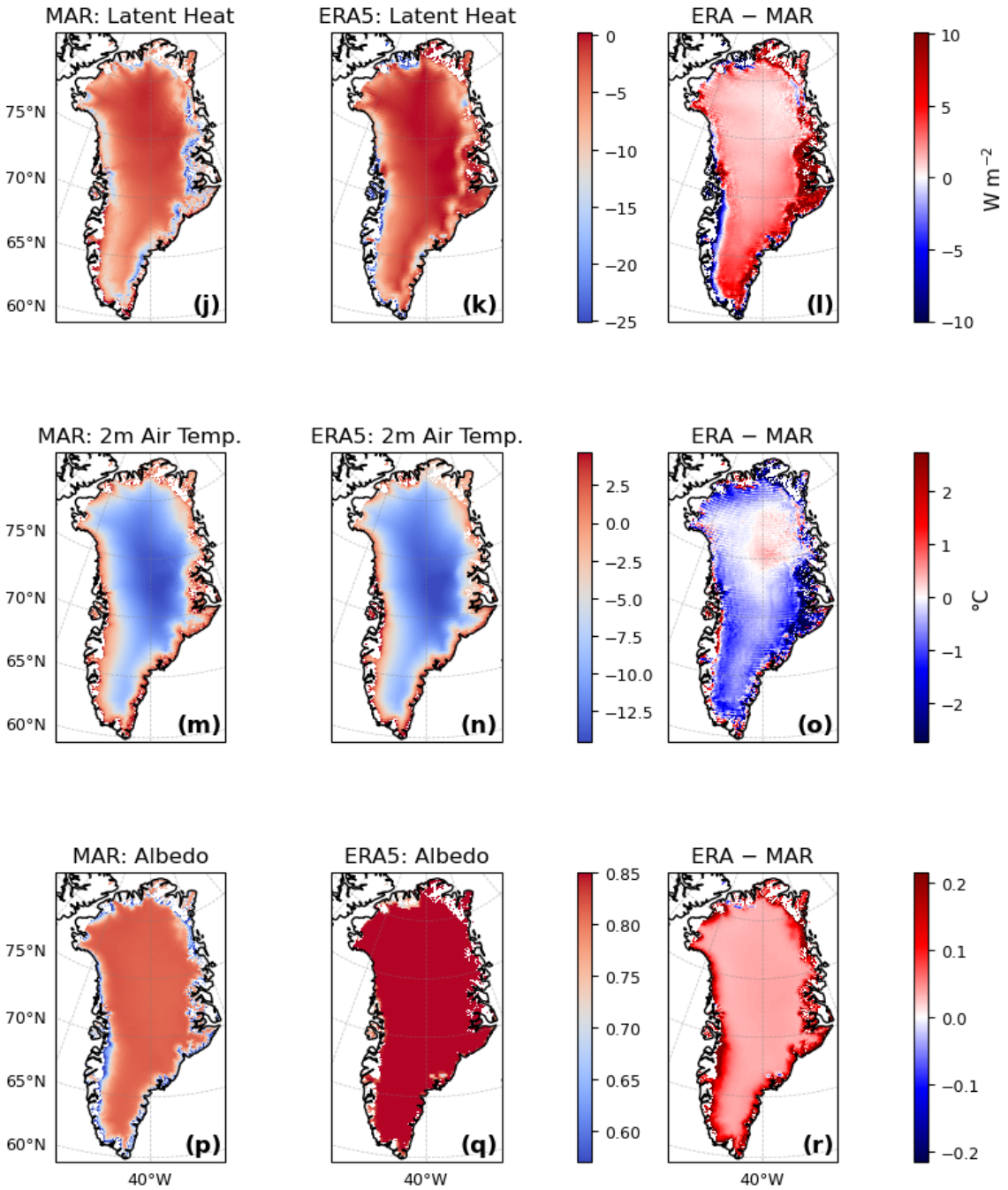


Figure 3: 1979 - 2024 summer (June - July - August, JJA) averaged values for the predictors used in the models' training (left and center columns) and differences between the MAR and ERA-5 values (right column).

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

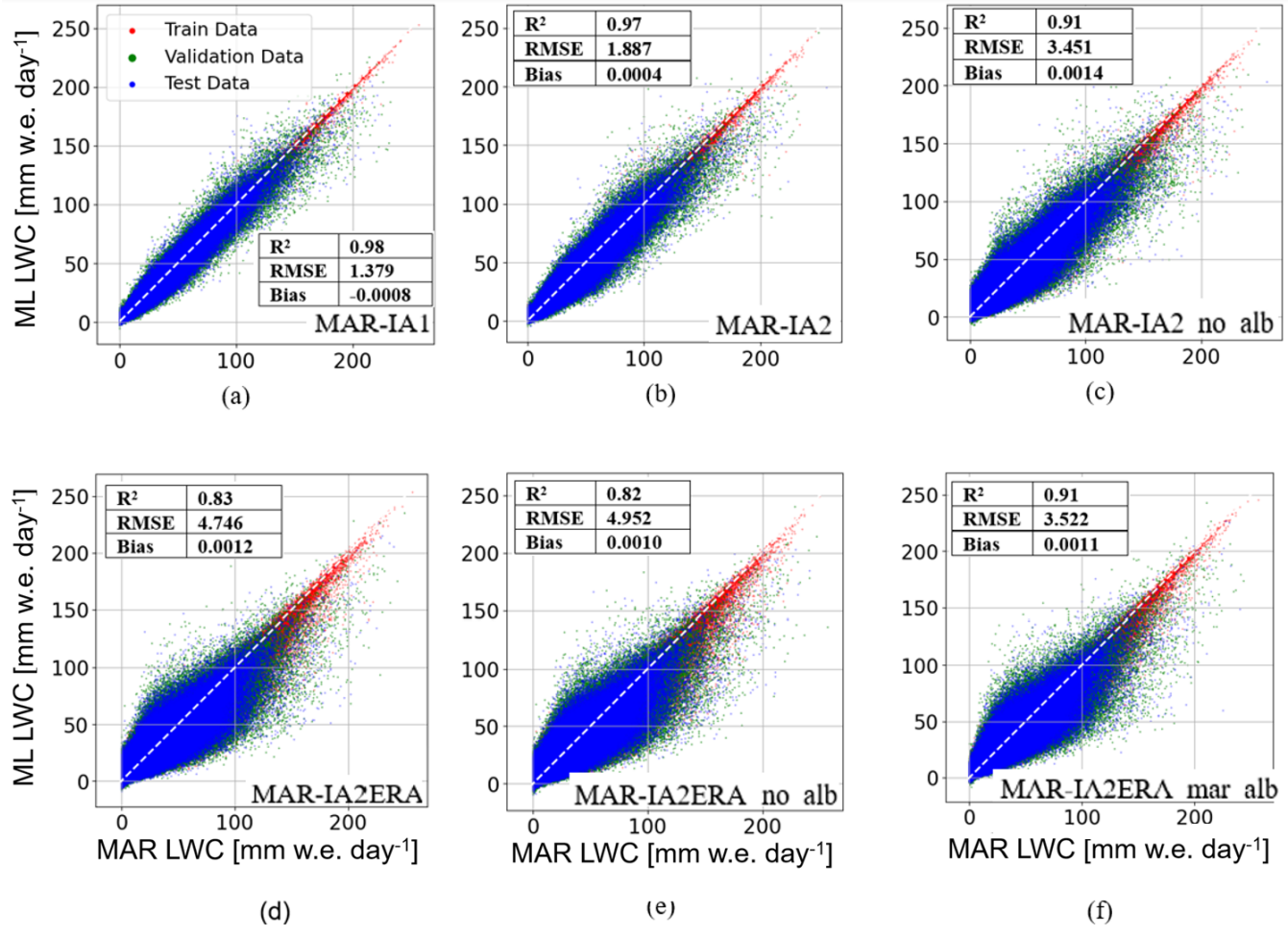


Figure 4. Scatterplot of the MAR (x-axis) and MAR-IA simulated (y-axis) surface meltwater production in the case of MAR-IA1 (a), MAR-IA2 (b), MAR-IA2_no_alb (c), MAR-IA2ERA (d), MAR-IA2ERA_no_alb (e), MAR-IA2ERA_mar_alb (f). Red dots represent training data, green dots validation data (used to monitor overfitting during training but not used for the training itself), and blue dots show the independent testing dataset.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

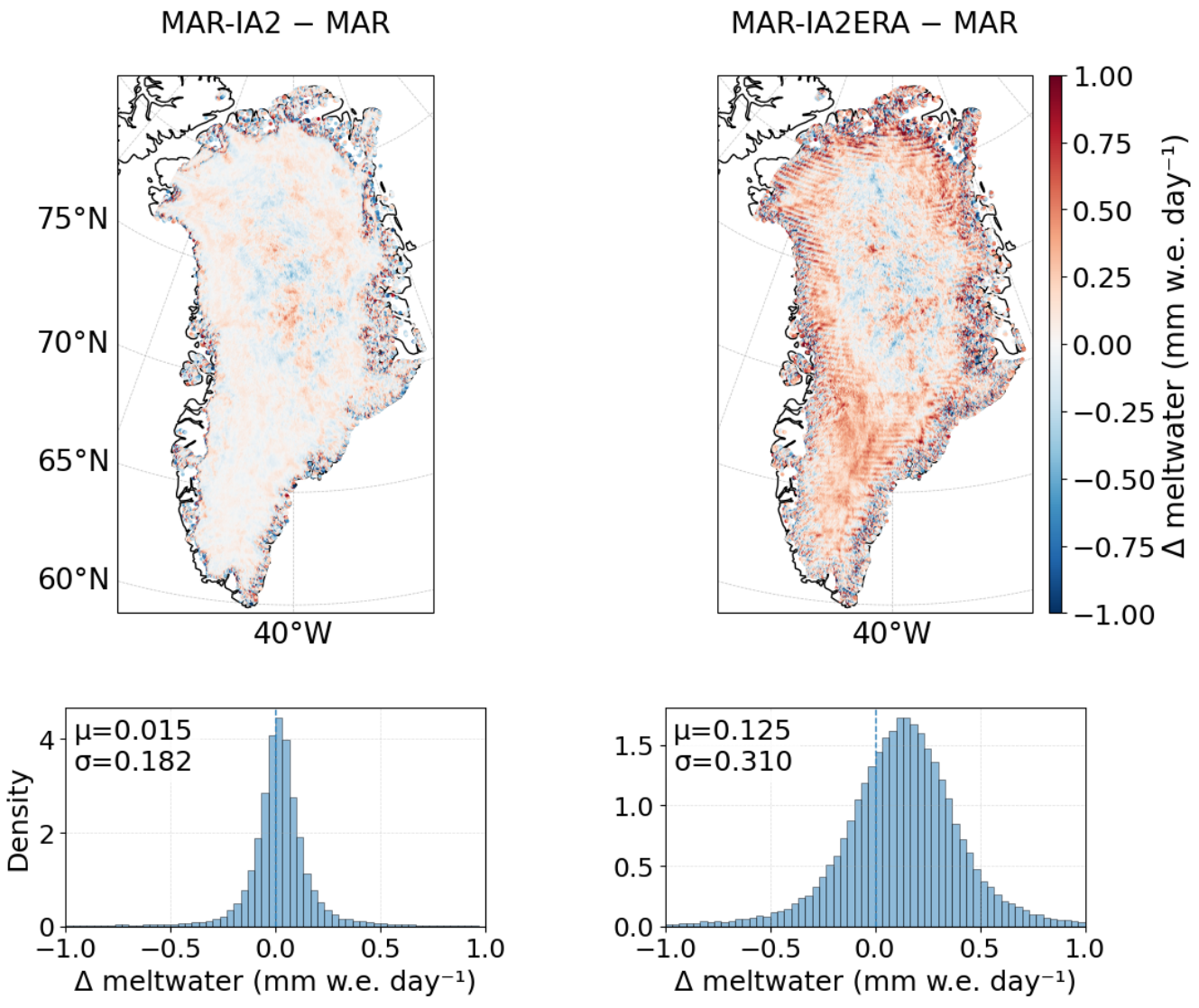
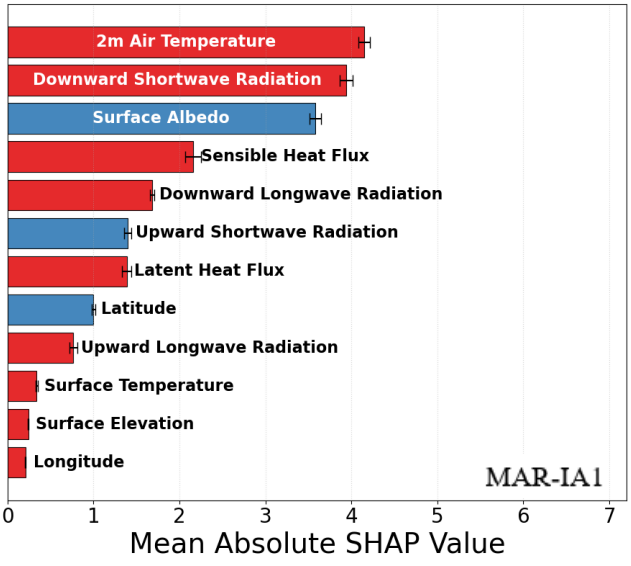


Figure 5: Mean difference in meltwater production between the machine−learning emulator and MAR for (left) MAR−IA2 and (right) MAR−IA2ERA configurations. Colors show the spatial distribution of the mean daily meltwater difference (ML − MAR) at each MAR grid point, obtained by averaging over all available days and years (1979−2024). The lower panels show the corresponding spatial distributions of the meltwater differences across the domain.

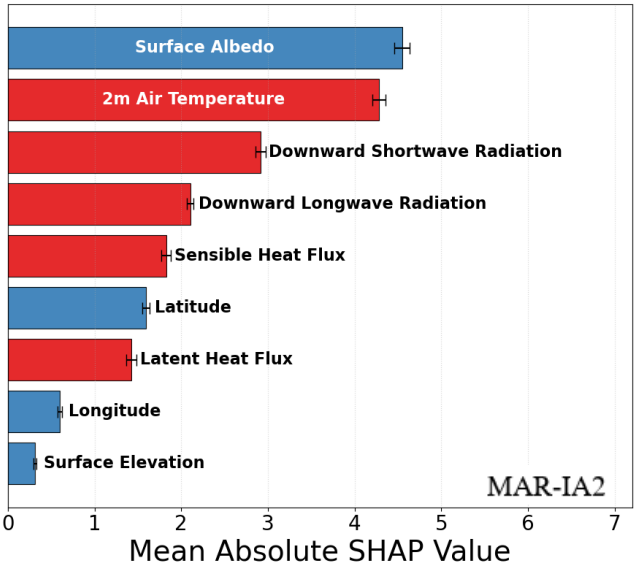
https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

560

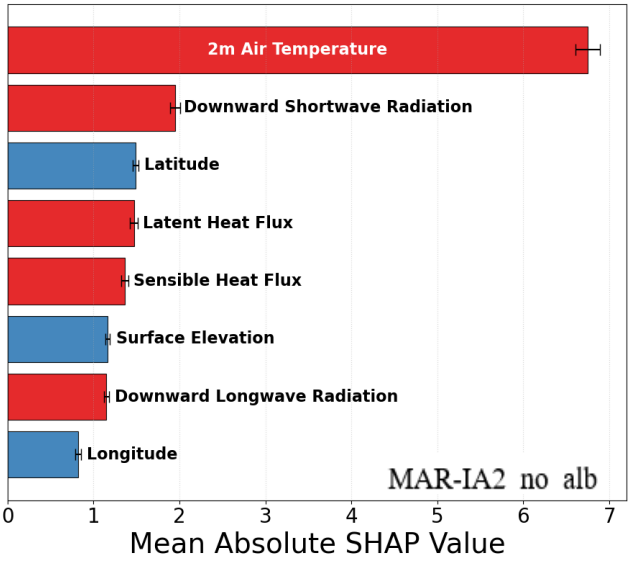
561



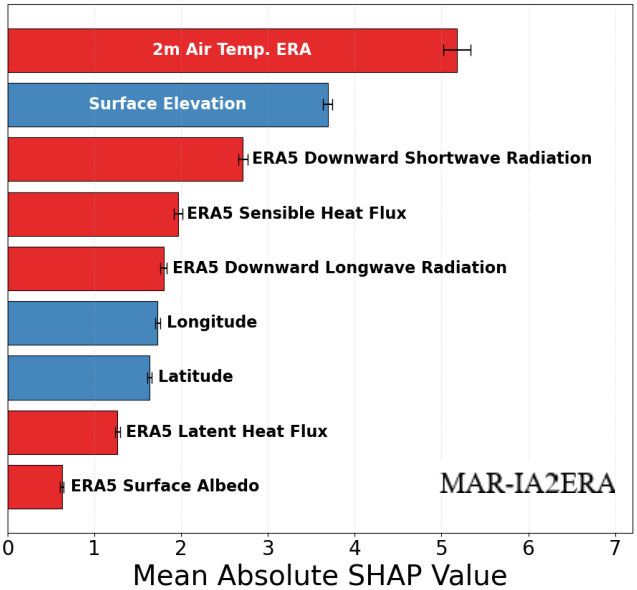
(a)



(b)

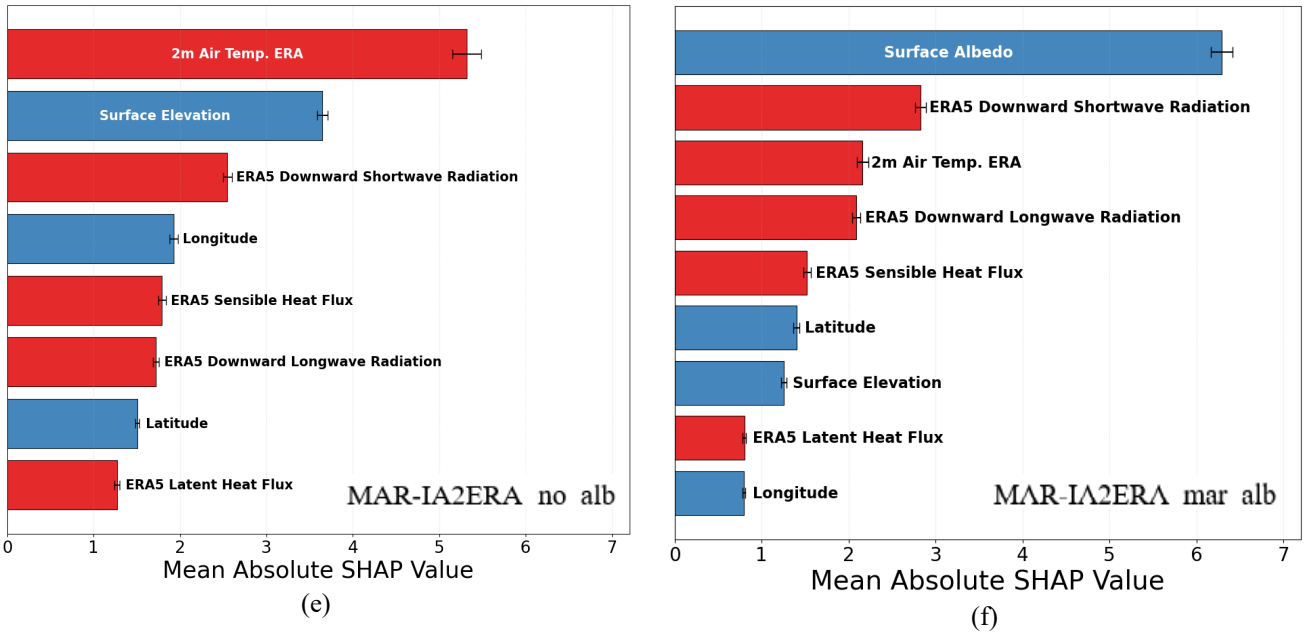


(c)



(d)

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026



562 **Figure 6. Mean absolute Shapley values obtained for the different models using the SHAP algorithm. Predictors that**
563 **are negatively (positively) correlated with surface meltwater production are plotted as blue (red) bars.**

564

565

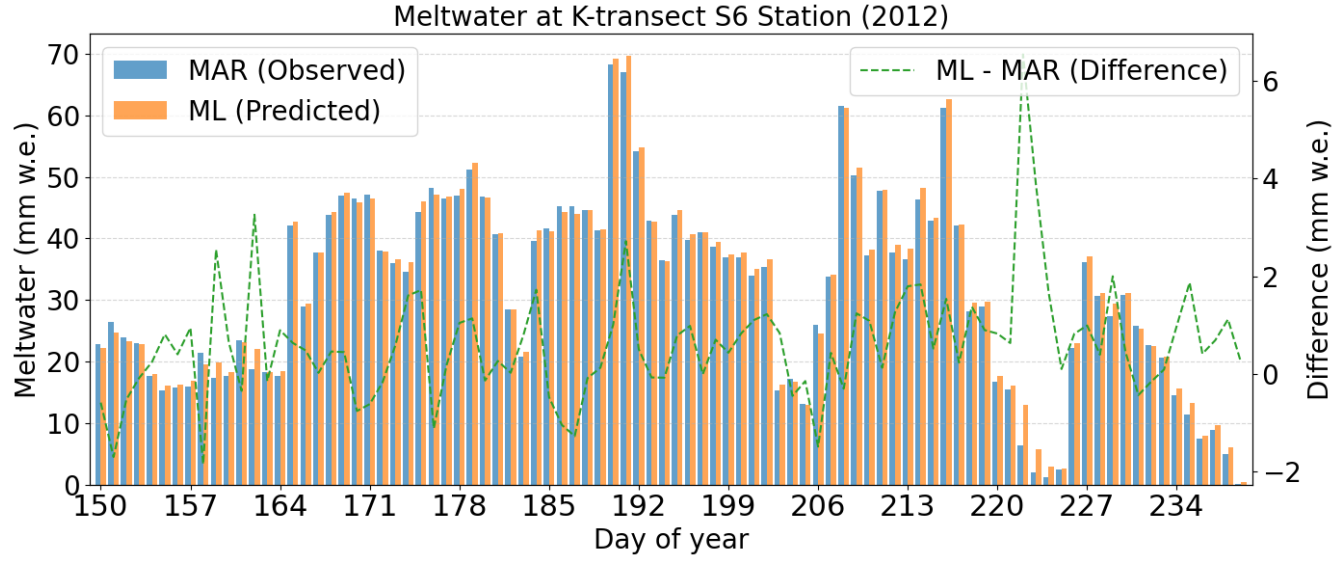
566

567

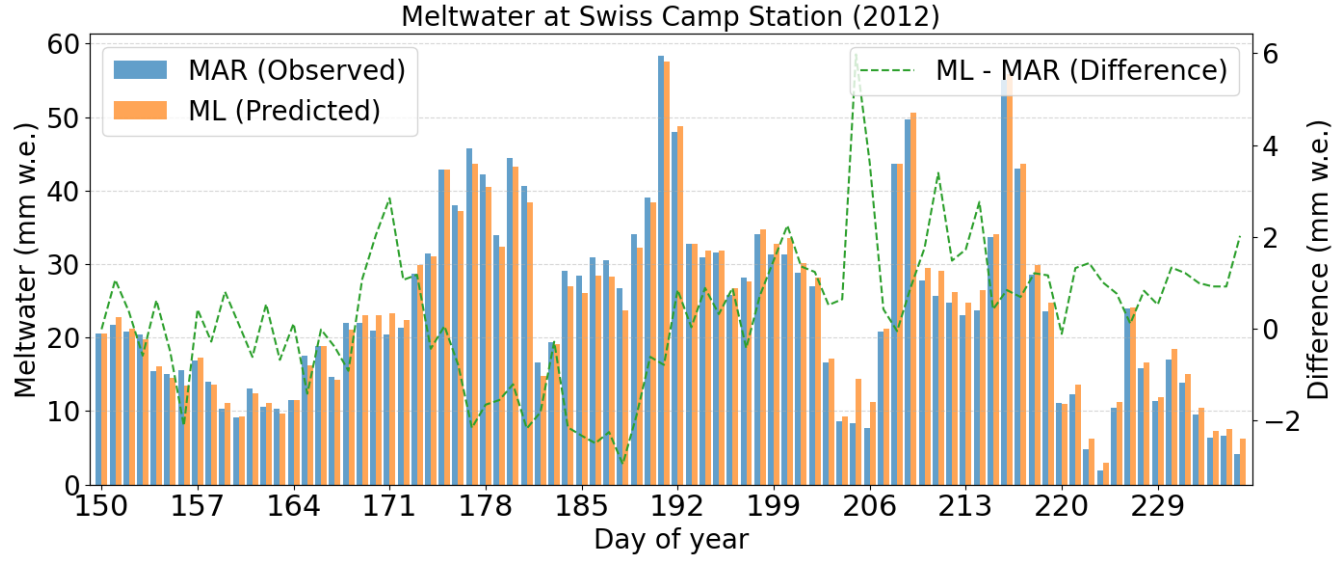
568

569

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026



(a)



(b)

Figure 7 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.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

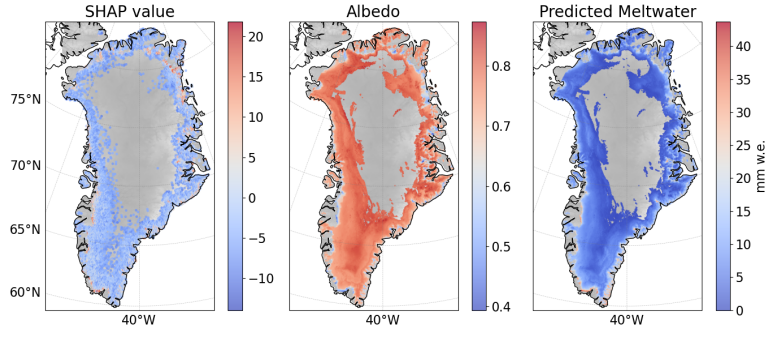
578

579 **Figure 8 Summer (JJA) mean Shapley values for the (a) non-flux and (b) flux variables for the period 1979 - 2024.**
580 **Linear trends are reported also as dashed lines with the same color as the corresponding variables. Note this is for**
581 **MAR-IA2.**

582

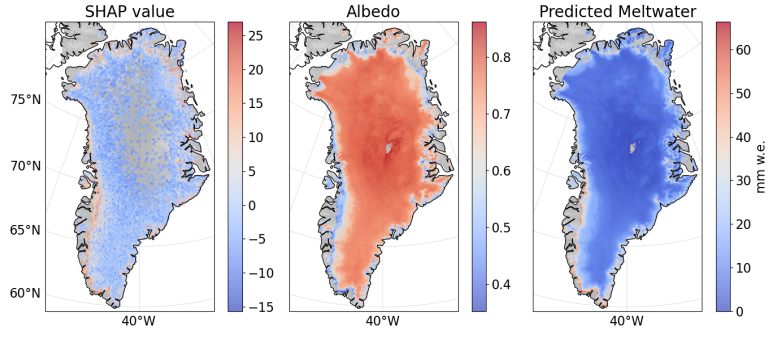
https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

583



584

585 (a) (b) (c)



586

587 (d) (e) (f)

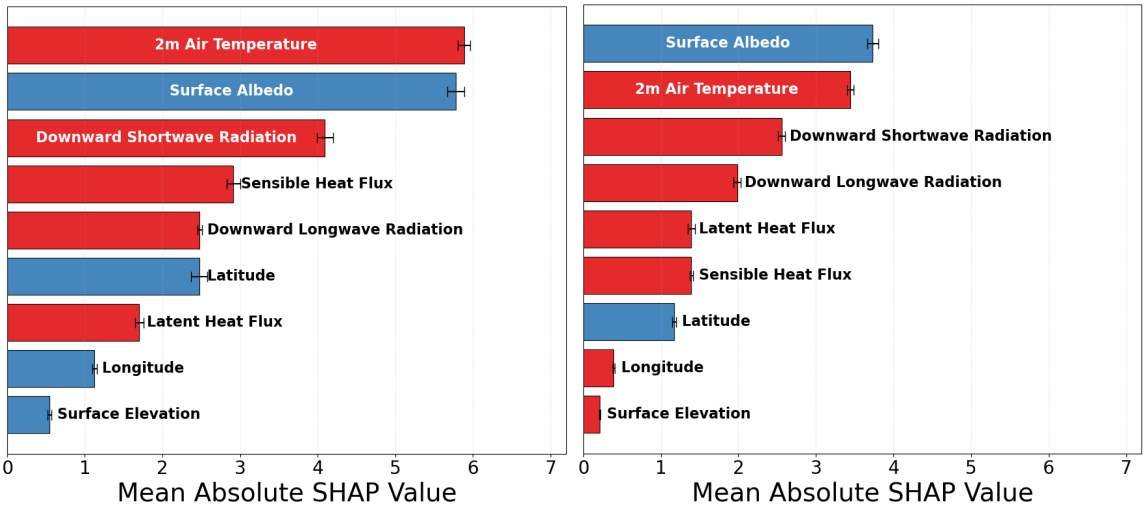
588 **Figure 9 Summer-averaged spatial distribution of Shapley (a,d), albedo (b,e) and ML-estimated meltwater (c,f) for**
589 **the 1992 (a,b,c) and 2012 (d, e, f) for the MAR-IA2 model.**

590

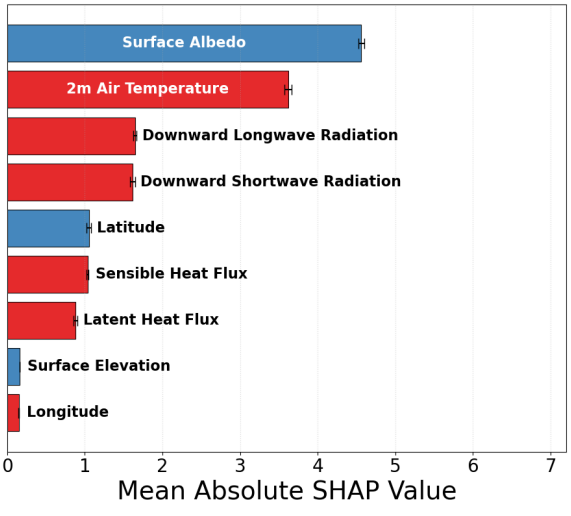
591

592

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026



(a) (b)



(c)

Figure 10 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.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

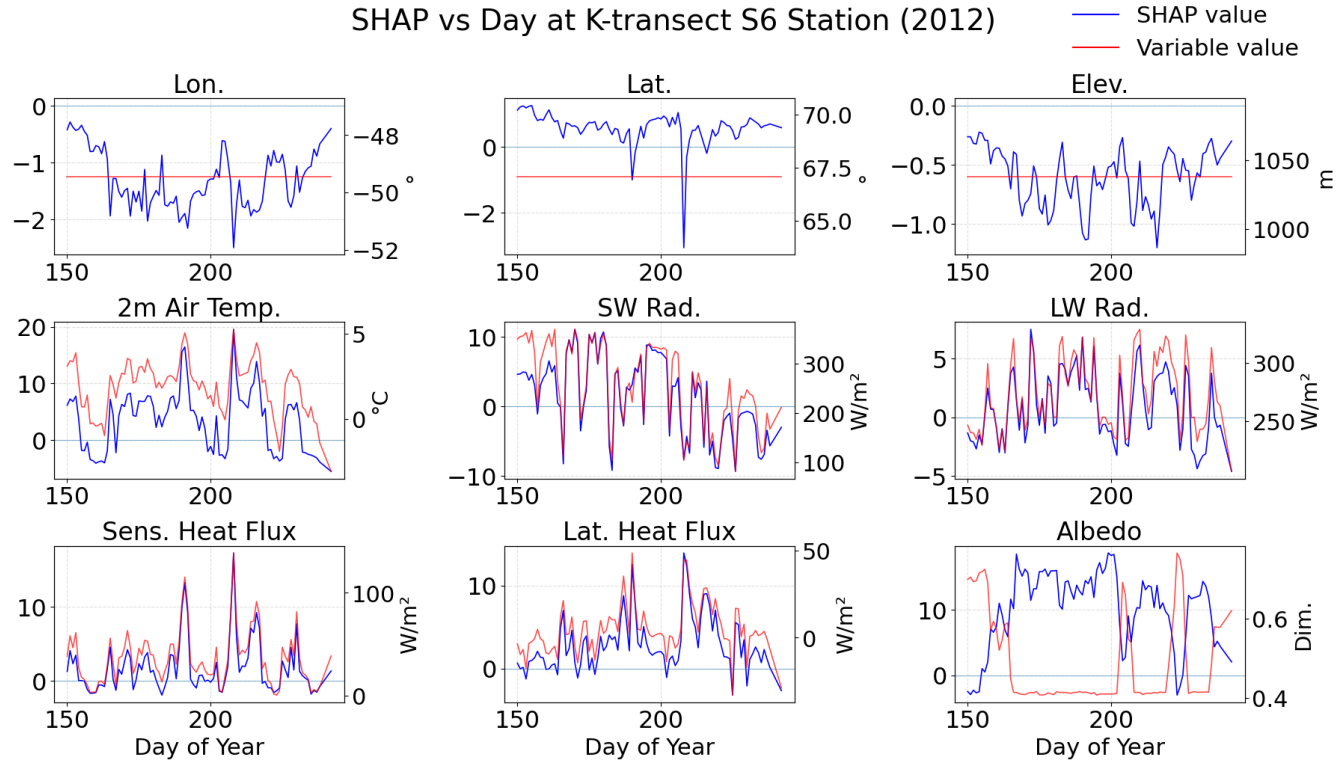


Fig.11Daily Shapley values for the different predictors obtained from the model MAR-IA1 and associated trends (dashed red line) for the pixel containing the K-transect S6 station.

https://doi.org/10.5194/egusphere-2026-490
Preprint. Discussion started: 10 February 2026

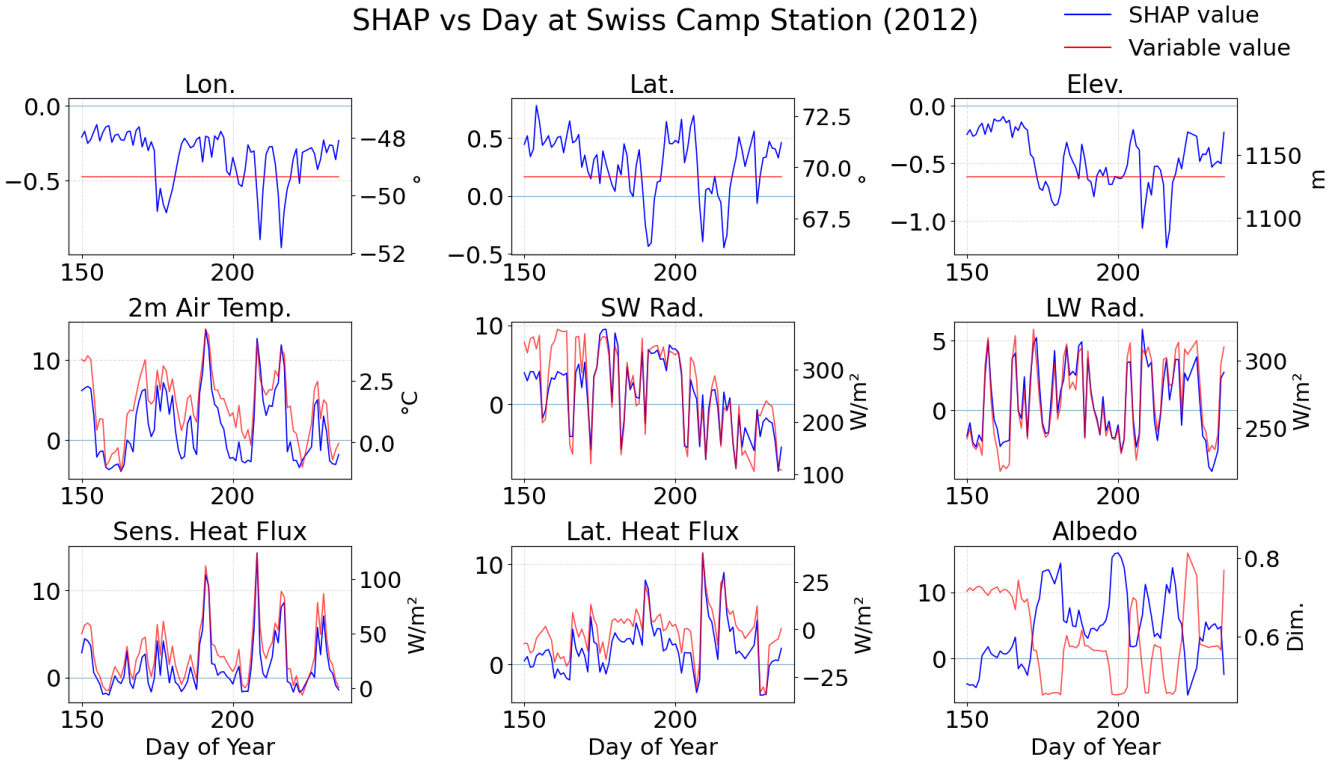


Fig. 12 Same as Fig. 8 but for the Swiss Camp location.