Response to Reviewer 1 Comments

Anonymous Referee

for

Data-driven equation discovery of a sea ice albedo parametrisation

https://doi.org/10.5194/egusphere-2025-3556

Submitted to

The Cryosphere

by

Diajeng W. Atmojo[1,2], Katja Weigel[1,2], Arthur Grundner[2], Marika M. Holland[3], Dmitry Sidorenko[4], Veronika Eyring[2,1]

[1] University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

[2] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

[3] National Center for Atmospheric Research, Boulder CO, USA

[4] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI), Bremerhaven, Germany

Authors' Response to Reviewer 1

General Comments. This paper presents a novel and well-executed application of symbolic regression for data-driven discovery of a sea ice albedo parametrisation based on satellite and reanalysis data. The authors demonstrate how an interpretable machine-learning approach can be used to derive a physically constrained, low-complexity functional form that substantially improves upon the very simple albedo parametrisation currently used in FESIM. The methodological framework is sound, the ML methods are appropriate and carefully applied, and the Pareto-optimal perspective on error versus complexity is particularly convincing. The paper is well written, the figures are clear and convincing, and the presentation is very well structured. The appendices are comprehensive and helpful, and the authors do a good job at explaining and interpreting the discovered equation in physical terms. Overall, this is a very nice example of how observational and reanalysis datasets can be used in an innovative and effective way to inform parameterisations of physical processes in climate models.

My comments above are mostly in the spirit of model development and robustness, and aim at strengthening the physical interpretation, generality, and long-term applicability of the proposed approach.

General Comments. Main limitations and suggestions for future work

In its current form, the study remains an offline, data-driven evaluation: a proper assessment of the impact of the new parametrisation within FESOM itself is still missing. While I do not think this is a blocker for the present paper, it is an important next step if the results are to become truly impactful for climate modelling.

In addition, the benchmarking is currently limited to statistical and ML-based baselines. The study would be significantly strengthened by a comparison against a more physically based or numerically robust albedo scheme, for example radiation schemes used in models such as Icepack, or a parameterisation like that of Holland et al. (2012, 10.1175/JCLI-D-11-00078.1). Even if such schemes are not implemented in FESOM, they would provide a much more meaningful physical reference than polynomials or neural networks.

More generally, some of my comments raise concerns about generalisability, in particular:

- the reliance on predictors that may encode reanalysis-specific biases (e.g. the T0m–T2m difference),
- the question of transferability to different forcings (e.g. coupled configurations, future reanalyses),
- and the applicability to other regimes (e.g. Antarctic sea ice, future climates, palaeoclimate states).

Finally, given that the study does not yet demonstrate the impact of the scheme inside FESOM, it might be worth toning down the emphasis on FESOM/FESIM in the framing. At present, the work is best seen as a general, observation-driven albedo parameterisation study rather than a demonstrated FESOM model improvement.

Overall recommendation

Overall, I find this to be a strong, original, and well-executed study that convincingly demonstrates the potential of interpretable machine learning for parameterisation development. The paper is already of high quality, and addressing (or at least discussing) the points raised above would further strengthen its physical robustness, scope, and long-term relevance for climate modelling.

Response: We thank the reviewer for the insightful summary and positive feedback on our work. In the revised manuscript, we will address the reviewer's specific comments, clarifying the physical robustness, scope, and long-term relevance for climate modelling. The reviewer's concern about generalisability and robustness will be addressed in detail in the next comments. Here, we would like to respond to the suggestions regarding comparing against a more physically based or numerically robust albedo scheme and toning down the emphasis on FESOM/FESIM.

First, for an "apple-to-apple" comparison, we need data which come with the same spatial and temporal resolution like the data products we chose for our methodology. As pointed out in L496-498, there is no data available on a daily, pan-Arctic scale e.g. for snow grain size, black carbon or algae which are needed to compute albedo using more sophisticated albedo schemes including explicit melt pond treatment like Holland et al. (2012) or Flocco et al. (2010). We acknowledge that comparing with more sophisticated albedo schemes would significantly strengthen the study. On the other hand, the simple PW79 is still a well-established albedo scheme used in ESMs like in AWI-CM3 (Streffing et al. 2022).

Second, we acknowledge that our study remains an offline, data-driven evaluation. We will revise the manuscript in such a way that we town down the emphasis on FESIM/FESOM, including changes suggested by Comment 1 of reviewer #2.

> **Abstract**
> In many sea ice models, a single-category, zero-layer thermodynamic scheme is employed, in which sea ice albedo is prescribed based on surface types depending on snow cover, surface temperature, or sea ice thickness. The Parkinson and Washington parametrisation (PW79) is a commonly used one, which assigns four constant albedo values corresponding to distinct surface types. This parametrisation is too simple to capture the spatiotemporal variability of observed sea ice albedo.
> ...
> ..., showing that the equation excels in balancing error and complexity and reduces the mean squared error by about 51 % compared to PW79.

L39 ff.

Many sea ice models still employ simplified sea ice albedo parametrisations, trading accuracy or more complex physics for simplicity and lower computational cost. As an example, the Finite-Element Sea Ice Model (FESIM; Danilov et al., 2015), part of the Alfred Wegener Institute Climate Model (AWI-CM3; Streffing et al., 2022), employs a very simplified sea ice albedo parametrisation based on Parkinson and Washington (1979, hereafter PW79).

This study applies symbolic regression, a data-driven equation discovery approach, to discover an equation for sea ice albedo directly from observational data, targeting sea ice models which employ the zero-layer scheme with an implicit melt pond treatment (Parkinson and Washington, 1979).~~improving upon the simple PW79 sea ice albedo parametrisation within FESIM.~~

L68-69

2. Do we improve our physical understanding of the surface radiative budget of the Arctic Ocean with our data-driven equation and discover deficiencies in how sea ice thermodynamics are treated when using PW79?

Comment 1

Line 25: The statement that thinner and younger sea ice necessarily has a lower albedo than thicker, older ice seems too general. While thicker ice can indeed accumulate more snow, which increases surface albedo, the albedo of bare ice does not necessarily decrease with decreasing ice age or thickness. In particular, younger ice typically has higher salinity, which may increase scattering and thus lead to higher bare-ice albedo. The authors may wish to clarify the conditions under which this statement holds (e.g. snow-covered vs. bare ice) or provide a supporting reference.

Response:

We acknowledge that this statement holds for more idealised data (see Grenfell 1979) and will clarify the lines accordingly.

L24 ff.
Neglecting microstructural features such as salinity or atmospheric aerosols, thinner and younger sea ice, prevalent due to these changes, has a lower albedo (Grenfell, 1979), which leads to a higher absorption of the solar radiation by the sea surface, thereby promoting sea ice melting and the formation of melt ponds (Perovich et al., 2002; Light et al., 2022; Niehaus et al., 2024).

Comment 2

Line 115: It is unclear whether the reference here is still to PIOMAS. The sentence would benefit from greater clarity and, if appropriate, an explicit citation. In addition, the discussion would be strengthened by considering relevant recent work such as Cocetta et al. (2024, The Cryosphere), which may be pertinent to the points being made.

Response: The references are the Quality Information Document (`https://documentation.marine.copernicus.eu/QUID/CMEMS-ARC-QUID-002-003.pdf`) and the Synthesis Quality Overview (`https://documentation.marine.copernicus.eu/SQO/CMEMS-ARC-SQO-002-003.pdf`) of TOPAZ4b. We will add these references and also the reference you suggested to the respective lines, which also Reviwer #2 pointed out in Comment 6.

Comment 3

Line 132: Over the course of one week, sea ice can drift substantially. It is therefore unclear whether sea ice motion is accounted for in the weekly accumulation of snowfall and rainfall. Could the authors clarify whether these accumulations are computed in a Lagrangian (ice-following) or Eulerian framework? If not accounted for, this mismatch may introduce inconsistencies between surface forcing and ice properties. Relatedly, could sea ice velocity or the atmospheric wind beyond the day before provide additional predictive skill for albedo, particularly in regions with strong ice drift?

Response: The weekly accumulated snowfall and rainfall used in our analysis are computed in an Eularian framework and do not take into account sea ice advection. The reviewer has a good point that sea ice motion could produce a mismatch in these weekly accumulated fields and the albedo state. We now acknowledge this within the manuscript and that this could modify the influence of these precipitation fields for determining the albedo.

Regarding whether sea ice velocity or winds beyond the day before could provide additional predictive skill for albedo, we expect that they might but also expect that the general conclusions from our study that snow thickness and surface temperature are the most informative predictors would remain the case. Indeed, it is possible that other predictor variables at additional lead times could also be impactful but exploring this is beyond the scope of our study. We now include text within the manuscript to reflect this.

L131

Additionally, we adjust rain and snowfall using a cumulative sum from the preceding seven days to consider a weekly memory effect. This cumulative sum is computed using an Eulerian framework and so neglects the potential influence of sea ice advection. This could modify the influence of these fields on the albedo state. We acknowledge that some of our predictor fields could be considered at different time lags and that other variables not considered here could influence the surface albedo. However, we have retained the existing set of variables as a reasonable balance between completeness and feasibility.

Comment 4

Line 188 (Point 3): With respect to point 3, please see my earlier comment regarding the role of salinity. The assumption that thicker ice necessarily has a higher albedo under freezing conditions may not always hold for bare ice. An explicit dependence on ice age, which is closely linked to salinity and microstructure, might therefore be more appropriate or at least worth discussing.

Response: We thank the reviewer for their perspective. However, the sequential feature selection method consistently ruled out ice age from the most informative input features, suggesting that for the given dataset, ice age is not a key input feature that drives sea ice albedo. Therefore, it would be obsolete to formulate a physical constraint based on an input feature which is not used in the equation.

However, we acknowledge to clarify the physical constraints as approximations which do not take into account microstructural features and will revise the lines accordingly.

L190 ff.

The PCs are approximations which we assume for large-scale applications and for simplicity. We do not account for microstructural characteristics such as salinity and atmospheric aerosol deposition. For instance, younger, bare ice typically has higher salinity, which may increase scattering and therefore increase albedo comparable to multiyear ice (Light et al. 2015; Perovich and Grenfell 1981), while the deposition of atmospheric aerosols reduces albedo independent on snow or sea ice thickness (e.g. Warren and Wiscombe 1980; Hansen and Nazarenko 2004).

Comment 5

Line 229: The statement that snow is "the most reflective medium in nature" appears too strong and would benefit from clarification and a supporting reference. While fresh, clean snow is among the most reflective natural surfaces, particularly in the visible spectrum, its albedo depends strongly on grain size, age, wetness, and impurities, and decreases substantially in the near-infrared. Moreover, other natural media (e.g. certain cloud types) can exhibit higher broadband reflectance. A more precise phrasing such as "snow is among the most reflective natural surfaces, especially when fresh and dry" would be more accurate.

Response: We agree with the reviewer and will revise the lines accordingly.

L299

h_{snow} being the most informative predictor is plausible since snow is among the most reflective natural surfaces, especially when fresh and dry. When present, snow represents the uppermost layer where solar radiation initially impacts.

Comment 6

Line 239: An alternative explanation could be that sea ice thickness is strongly correlated with snow depth, such that ice thickness partly serves as a proxy for snow information. Given that all features are standardised during training, this correlation may lead the model to attribute predictive skill to ice thickness that in fact originates from snow-related information. Clarifying the degree of correlation between these variables and its impact on feature selection would strengthen the interpretation.

Response: We thank the reviewer for this insightful observation and acknowledge further clarification of the correlation matrices shown in Appendices A1 and A2.

While the correlation matrices of the satellite and reanalysis data (A1) exhibit a moderate to strong Pearson correlation between snow and sea ice thickness (0.8 and 0.62, respectively), the merged, standardised dataset (A2) shows a significantly weaker Pearson correlation of 0.19. However, the difference in the correlations is not attributed to the standardisation, since standardising solely rescales variance which does not alter linear relationships. What actually happens is that the satellite and reanalysis data represent different seasons, namely from March until mid April and from mid April until September, respectively. The fact that both datasets show similar correlation reveals that they exhibit similar internal correlation structures, but the reduced correlation of the merged dataset implies seasonal heterogeneity, suggesting that the relationship between snow and sea ice thickness is regime-dependent. This is physically plausible since from March until April the dataset includes both multiyear (Central Arctic) and seasonal sea ice (e.g. Barents Sea, Kara Sea), while from April onwards, multiyear ice dominates the dataset due to the melting season. This is evident in Sec. 5.2, where we showed a reversed weighting of snow and sea ice thickness due to regime-dependent variance structure of snow and sea ice thickness when comparing pan-Arctic with Barents Sea data.

Furthermore, we ran SFS ten times where on average, sea ice thickness is ranked as third most predictive input feature. If sea ice thickness was only a proxy for snow thickness, sea ice thickness should add little additional predictive value after selecting snow thickness. However, evaluating the error-complexity plane (Fig. 6 in the manuscript), we see that after adding sea ice thickness (3-feature NN, third purple point from the right), the MSE is reduced by 0.0022 compared to the 2-feature NN, indicating that sea ice thickness contributes information beyond what is explained by snow thickness alone.

Overall, we agree with the reviewer to weaken the physical interpretation and instead interpret the ranking in a statistical sense. However, the differing correlations between snow and sea ice thickness in A1 and A2 stem from seasonal heterogeneity and not from the standardisation of the features, therefore not explaining the SFS ranking.

L239

As sea ice has a higher optical depth than snow, h_{ice} ranked below h_{snow} seems plausible, implying that h_{snow} provides larger marginal predictive improvement than h_{ice}.

Comment 7

Line 245: While the interpretation is plausible in principle, in practice it should be treated with caution. The 2 m air temperature (T2m) is taken from a reanalysis product that does not assimilate sea ice or snow thickness, nor near-surface Arctic observations, apart from surface pressure from stations and drifting buoys. In contrast, T0m is derived from satellite observations. The fact that the machine-learning model does not fully exploit the strong physical relationship between T0m and T2m may therefore reflect inconsistencies between these two datasets rather than genuine additional predictive skill of T2m for sea ice albedo. This distinction would be worth discussing more explicitly.

Response: We thank the reviewer for raising this point. We will incorporate it into the revised version, including changes suggested by Comment 3 of reviewer #2:

L244 ff.

Additionally, this may be due to the fact that T_{2m} can exceed the melting point, whereas T_{0m} cannot. Another consideration is that *ERA5* does not assimilate sea ice or snow thickness, nor near-surface Arctic observations, except for surface pressure from stations and drifting buoys. Previous studies have shown that this leads to warm temperature biases in *ERA5* over the Arctic, particularly during polar winter clear-sky events (Batrak and Müller 2019; Zampieri et al. 2023). Such biases could introduce inconsistencies between the satellite-derived $T_{0\mathrm{m}}$ and the *ERA5*-biased $T_{2\mathrm{m}}$, which might partly contribute to the predictive skill attributed to $T_{2\mathrm{m}}$. The documented warm bias is particularly large during polar winter stable boundary layer conditions. Our exclusive use of polar-day samples will thus help to mitigate the influence of this bias. However, this documented warm bias in *ERA5* and data inconsistencies between surface and 2 m air temperatures may play some role in our results, although we believe it is unlikely to fully explain the relationship identified here.

Comment 8

Subsection 3.2.2: The interpretation of the weighted difference between T0m and T2m as a physically meaningful seasonal proxy deserves further scrutiny. The seasonal variation of this difference largely reflects a well-documented bias in the ERA5 reanalysis, which has been shown to be larger in winter than in summer (e.g. Batrak and Müller, 2019; Zampieri et al., 2023). As such, this signal may primarily encode characteristics of the reanalysis system rather than an intrinsic physical control on sea ice albedo.

This raises the question of whether it is desirable to base a model parametrisation on a feature that is strongly influenced by a known, non-stationary reanalysis bias. In particular, the robustness of the proposed parametrisation is unclear if different forcing datasets are used to run FESOM (e.g. a future ERA6 reanalysis or a coupled model configuration), where the magnitude or structure of near-surface temperature biases may differ substantially.

Moreover, ERA5 temperature biases are not static in time and are themselves affected by climate change. It is therefore uncertain whether the derived parametrisation would generalise well under future climate conditions.

If the intention is to encode seasonal information, a predictor more directly linked to the physical drivers of seasonality, such as variables related to insulation, radiative forcing, or surface energy balance, might be more appropriate than relying on the difference between two temperature fields with differing observational constraints. A more explicit discussion of these limitations and their implications for generalisability would strengthen the manuscript.

Response: We thank the reviewer for highlighting this issue. Please see our response to Comment 7. We agree that the warm temperature bias in ERA5 documented in Batrak and Müller 2019 and Zampieri et al. 2023 should be taken into consideration when interpreting the temperature drivers of α. At the same time, our dataset exclusively contains polar-day data which help mitigate the influence of this bias. Moreover, the reviewer is right to point out that using other predictors more directly linked to the physical drivers of seasonality might be more robust to encode seasonal information. We will acknowledge these points in the manuscript, which could be explored in future work, including changes suggested by Comment 3 of reviewer #2.

L304 ff.
While we expect that ΔT^* is providing meaningful physical information, the seasonal cycle that is reflected in ΔT^* could be influenced by the aforementioned bias in *ERA5* $T_{2\mathrm{m}}$ which is largest during the cold season and not present during summer months. Nevertheless, it does suggest that information on the seasonal cycle is useful in providing a constraint on the surface albedo. Other possible predictors that encode information on the seasonal cycle, such as solar insolation or the surface energy balance, could also provide useful information and could be explored in future work. Considerations of training data biases and prioritisation of predictors that enable results to be generalised across regions and different climate states are important for possible ML-based parametrisations that could be developed based on this work.

L478 While the PW79 sea ice albedo parametrisation only uses the surface temperature as a proxy to define freezing and melting conditions, our equation shows that a weighted temperature difference between the surface and the air at 2 m better encodes information on the seasonal cycle. As our physical interpretation could be influenced by the warm 2 m air temperature biases in *ERA5*, other possible predictors that encode information on the seasonal cycle, such as solar insolation or the surface energy balance, could be explored in future work.

Concerning the robustness due to non-stationary ERA5 biases and generalisability of Eq. 4, please see our response to Comment 11.

Comment 9

Line 363: Even if these extreme albedo values are partly attributable to uncertainties or errors in the satellite retrieval and processing, it is still noteworthy that the model is unable to reproduce them. This may indicate limitations in the model's flexibility or in the chosen functional form, and could merit a brief discussion.

Response: We thank the reviewer for this remark. The reviewer is right to address the functional form of the proposed equation which includes two hyperbolic tangent functions with an asymptotic behavior. We already analysed the infimum and supremum of our proposed equation in L307-320, which is constrained by $\tilde{a}$ and $\tilde{b}$. So mathematically speaking, it is not the functional form of the equation but the coefficient values that limit the equation to predict extreme albedo values. We will add a reference to Sec 3.2.3 to clarify how the functional behavior of the equation plays a role in predicting extreme values, including changes based on Comment 4 of reviewer #2.

L367-368

In contrast, Light et al. (2022) reports highly localised albedo values for each surface type. Furthermore, as examined in Sec. 3.2.3, the lower and upper limits of Eq. 4 are determined by the coefficients $\tilde{a} = 0.84$ and $\tilde{b} = 2.19$, which have been optimised using the pan-Arctic dataset. With these coefficient values, Eq. 4 is unable to reproduce extreme albedo values. On the basis of these considerations, we infer that the upper albedo limits prescribed by Eq. 4 is physically plausible, particularly given that the reference observation is noisy and represents average albedo over a large area.

Comment 10

Section 5.1: It would be interesting to compare the regional and monthly optimisation strategies with a single global optimisation that explicitly encodes spatial (e.g. longitude/latitude) and temporal (e.g. day of year) information within an Arctic-wide model. In practice, such an approach might be more usable and could avoid issues related to sharp transitions between predefined regions and seasons, while still allowing the model to represent spatial and seasonal variability.

Response: We thank the reviewer for this suggestion. The motivation behind analysing regional and seasonal differences is to explore how well the functional form of the proposed equation transfers across different regions and seasons. Section 5.1 serves more of a diagnostic tool to uncover what drives reproducing accurate α, revealing that the coefficients are sensitive to the prevailing sea ice regimes.

We acknowledge that encoding spatial and temporal information could be useful to apply our proposed parametrisation to an Arctic-wide operational model, since in this way, we potentially capture these regime variations. The core of the idea would be to make the parametrisation regime-aware. Please see our response to Comment 11, where we elaborated on the study by Nath et al. 2026.

Comment 11

Section 6 (Conclusions): The conclusions would benefit from a more detailed discussion of several important aspects related to the applicability and implications of the proposed parametrisation.

First, FESOM is a global model. It would therefore be important to discuss how this parametrisation might be applied to the Antarctic sea ice, and whether the authors expect the same functional form and predictors to remain meaningful in the Southern Hemisphere.

Second, it would be valuable to comment on whether the proposed parametrisation is expected to generalise beyond the present-day climate, for example to future climate change conditions or even to palaeoclimate regimes. Given the emphasis placed on the physical interpretability of the discovered equation, the authors are in a good position to discuss this point.

Third, FESOM is used in a variety of configurations (e.g. coupled versus ERA5-forced). Introducing this parametrisation removes an important tuning knob that currently exists in the system. Is there a plan to retain some degree of tunability (e.g. through selected coefficients), and can the authors provide recommendations for how this scheme should be integrated and tuned in different model configurations?

Response: We agree with the reviewer to add a discussion on the generalisability of our proposed parametrisation and elaborate further on the applicability and implications when running online simulations, including consequences for tuning.

We will revise the following paragraphs, including changes suggested from Comment 3 of reviewer #2 and Comment 1 of reviewer #3.

L473 ff.

This case study gives a first insight on how Eq. 4 can be transferred to different ice regimes, specifically from a stable, pan-Arctic regime dominated by multiyear ice in the Central Arctic to a fragile ice regime characteristic of the Barents Sea. Although we acknowledge that the MSE in the Barents Sea remains relatively high compared to other regions after fine-tuning (see Sec. 5.1), the functional form of Eq. 4 remains to be physically reasonable. This enables a comparison between the coefficients obtained from global and regional optimisations demonstrating that the optimal coefficients are state-dependent.

L502

As this mirrors present-day conditions, other subregions are underrepresented in our dataset where the impact of climate change is more pronounced such as the Barents Sea.

L508 ff.

An important question concerns how well Eq. 4 generalises well beyond Arctic sea ice regimes represented in the training data. Ship-based field measurements in the Antarctic by Brandt et al. 2005 and Tersigni et al. 2025 demonstrate that already a thin snow layer of a few centimetres substantially increases sea ice albedo, emphasising that snow fractional coverage is more impactful than snow thickness. Here, snow redistribution is mainly driven by strong winds, particularly in the marginal ice zones. Since our dataset has a spatial resolution of 25 km, retrieved h_{snow} likely reflects a combination of snow thickness and fractional coverage. The strong sensitivity of Eq. 4 to small variations in thin snow therefore suggests that the influence of snow on albedo is captured in a physically meaningful way.
Nevertheless, Sec. 5 indicates that the optimal coefficients of Eq. 4 are state-dependent. While the functional form remains transferable, as demonstrated by the Barents Sea case study, the relatively high MSE in this region after fine-tuning suggests that additional processes, such as oceanic heat fluxes, may drive sea

ice albedo but are not explicitly represented in Eq. 4. Oceanic heat fluxes also drive sea ice melt in the Antarctic (Brandt et al. 2005), which may influence sea ice albedo, implying that further offline investigations are required to assess the robustness of the parametrisation outside the pan-Arctic region. This is presently limited due to the lack of data availability on a daily, Antarctic scale.

In practice, Eq. 4 naturally retains a degree of tunability that facilitates its implementation in ESMs. Integrating Eq. 4 online into an ESM or operational sea ice forecast model would not require substantial changes in existing tuning protocols, as the parameter space can simply be expanded by the seven coefficients of Eq. 4 to obtain physically plausible sea ice states. Under different atmospheric forcings, either in an ocean–sea-ice stand-alone configuration driven by an atmospheric reanalysis product or in a fully coupled configuration, we hypothesise that distinct optimal values of these coefficients will emerge, particularly those controlling ΔT^* where the sea ice model receives $T_{2\mathrm{m}}$ from the atmosphere, since biases in atmospheric temperature fields vary across forcing datasets (Batrak and Müller 2019). This highlights the potential value of regime-aware parametrisations, as suggested by Nath et al. 2026, in which the parameter space is dynamically adjusted in response to the prevailing climate state, allowing the scheme to remain applicable across Arctic, Antarctic, and potentially future or paleoclimate sea ice regimes.

Overall, our results suggest that the functional form of Eq. 4 provides a physically interpretable representation of sea ice albedo variability, while the optimal coefficients depend on the prevailing climate state. This state dependence implies that the globally optimised coefficient set may not remain optimal across different regions or climate states. Yet, the explicit formulation and limited number of coefficients make the parametrisation well suited for online implementation in ESMs, where the coefficients can be tuned to the model's specific climate regime. Further evaluation, particularly in Antarctic conditions and under future or paleoclimate conditions, will be necessary to assess the broader applicability of the approach.

Batrak, Yurii and Malte Müller (Sept. 2019). “On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice”. en. In: *Nature Communications* 10.1, p. 4170. DOI: 10.1038/s41467-019-11975-3.

Brandt, Richard E. et al. (Sept. 2005). “Surface Albedo of the Antarctic Sea Ice Zone”. en. In: *Journal of Climate* 18.17, pp. 3606–3622. DOI: 10.1175/JCLI3489.1.

Grenfell, Thomas C. (1979). “The Effects of Ice Thickness on the Exchange of Solar Radiation Over the Polar Oceans”. en. In: *Journal of Glaciology* 22.87, pp. 305–320. DOI: 10.3189/S0022143000014295.

Hansen, James and Larissa Nazarenko (2004). “Soot climate forcing via snow and ice albedos”. In: *Proceedings of the National Academy of Sciences* 101.2, pp. 423–428. DOI: 10.1073/pnas.2237157100. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2237157100.

Light, Bonnie et al. (Nov. 2015). “Optical properties of melting first-year A rctic sea ice”. en. In: *Journal of Geophysical Research: Oceans* 120.11, pp. 7657–7675. DOI: 10.1002/2015JC011163.

Nath, Pritthijit et al. (Jan. 2026). “Making Tunable Parameters State-Dependent in Weather and Climate Models with Reinforcement Learning”. In: arXiv:2601.04268 [cs]. DOI: 10.48550/arXiv.2601.04268.

Perovich, Donald K. and Thomas C. Grenfell (1981). “Laboratory Studies of the Optical Properties of Young Sea Ice”. en. In: *Journal of Glaciology* 27.96, pp. 331–346. DOI: 10.3189/S0022143000015410.

Streffing, Jan et al. (Aug. 2022). “AWI-CM3 coupled climate model: description and evaluation experiments for a prototype post-CMIP6 model”. en. In: *Geoscientific Model Development* 15.16, pp. 6399–6427. DOI: 10.5194/gmd-15-6399-2022.

Tersigni, Ippolita et al. (2025). *Seasonal Variability of Snow Cover and Impact on Albedo and Thermal Properties in the Antarctic Marginal Ice Zone*. Version Number: 1. DOI: 10.48550/ARXIV.2507.10916.

Warren, Stephen G. and Warren J. Wiscombe (Dec. 1980). “A Model for the Spectral Albedo of Snow. II: Snow Containing Atmospheric Aerosols”. en. In: *Journal of the Atmospheric Sciences* 37.12, pp. 2734–2745. DOI: 10.1175/1520-0469(1980)037<2734:AMFTSA>2.0.CO;2.

Zampieri, Lorenzo et al. (June 2023). “A Machine Learning Correction Model of the Winter Clear-Sky Temperature Bias over the Arctic Sea Ice in Atmospheric Reanalyses”. In: *Monthly Weather Review* 151.6, pp. 1443–1458. DOI: 10.1175/MWR-D-22-0130.1.

Response to Reviewer 2 Comments

Guillaume Boutin

for

Data-driven equation discovery of a sea ice albedo parametrisation

https://doi.org/10.5194/egusphere-2025-3556

Submitted to

The Cryosphere

by

Diajeng W. Atmojo[1,2], Katja Weigel[1,2], Arthur Grundner[2], Marika M. Holland[3], Dmitry Sidorenko[4], Veronika Eyring[2,1]

[1] University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

[2] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

[3] National Center for Atmospheric Research, Boulder CO, USA

[4] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI), Bremerhaven, Germany

Authors' Response to Reviewer 2

General Comments. This a review of the manuscript entitled "Data-driven equation discovery of a sea ice albedo parametrisation". The presents a method based on machine learning (ML) to derive an equation for sea ice albedo suited for use in ocean–sea-ice models, with a particular focus on FESIM/FESOM. The strength of this approach lies in its ability to retain the accuracy of established ML techniques—demonstrated through comparison with a neural network model—while yielding an equation of relatively low complexity. The authors compare the discovered equation with several existing parameterizations: the simple and historical scheme used in FESIM, a neural network model, and a polynomial fit. They convincingly show that the new equation substantially improves the representation of both the temporal evolution and spatial distribution of sea ice albedo, while remaining interpretable in physical terms. This interpretability, combined with strong performance, is a valuable advantage for model development. The manuscript is very well written. It is clear and provides an adequate level of detail to follow the method without requiring prior knowledge of ML. The demonstration of the method's relevance to the sea ice modelling community is well structured and compelling. Many aspects are thoughtfully discussed throughout the text. In fact, I briefly considered recommending acceptance without revision, as the manuscript presents a well-constructed and coherent narrative.

Yet, after further reflection, I would like to offer a few optional suggestions that the authors may consider. In particular, the impact of the paper could be strengthened by framing the approach in a slightly more general context—not only in relation to the targeted sea-ice model. A short discussion on the reproducibility of the method and its computational cost could broaden its impact and better illustrate its practical applicability for a wider modelling audience. I also think there may be some limitations in the method that could be discussed. As I mentioned before, however, the manuscript is clear and well balanced. Therefore, I do not require lengthy additions, and if the authors feel that these suggestions do not add much value to the study, they should feel free to disregard them.

Response: We thank the reviewer for the insightful summary and positive feedback on our work. In the revised manuscript, we will address the reviewer's specific comments, reframing our approach in a more general context, and discuss the reproducibility, applicability and limitations of our study.

Comment 1

There is a lot of emphasis on FESIM/FESOM, which the authors justify because it only has a simple representation of albedo (with an implicit treatment of key processes such as melt ponds). However, FESIM is not the only model in this situation. As implied in the second paragraph of the introduction, similar parameterizations are used in many ESMs, where simplicity and computational cost are often prioritised over accuracy or more complex physics. Operational sea-ice models also commonly rely on simple schemes, which tend to be robust and less sensitive to errors in atmospheric forecasts (I am not sure there is a reference that uses these exact words, but this is a common point raised in discussions with forecasters). This is also the case for TOPAZ4B; the technical report linked below shows that its albedo scheme is essentially a set of constant values based on surface temperature combined with a function depending on ice thickness.
`https://www.researchgate.net/profile/Knud-Simonsen/publication/349710197_Formulation_of_Air-Sea_Fluxes_in_the_ESOP2_Version_of_MICOM/links/603df213a6fdcc9c78082e46/Formulation-of-Air-Sea-Fluxes-in-the-ESOP2-Version-of-MICOM.pdf`
Therefore, I think the relevance of the study extends beyond FESIM, and the introduction/conclusion could emphasise this a bit more. The authors could also highlight the benefit of their method given the complexity of albedo parameterizations due to the many involved processes. They mention melt pond schemes, but they could also note that such schemes often rely on compensating errors to achieve reasonable albedo values (Light et al., 2022, already cited). In addition, a recent study by Smith et al. (2025) highlights the difficulty of parameterizing melt ponds because of their complex life cycle and drainage processes, which are hard to relate to large-scale model variables. I would argue (very subjectively) that the approach presented here is more promising than an explicit melt-pond representation in the near future, and the authors might phrase this as offering a "promising alternative" (while acknowledging that melt pond schemes remain valuable).

Response: We thank the reviewer for this insightful broader overview of existing sea ice albedo parametrisations which are commonly used both in ESMs and operational sea ice models. We agree to town down the emphasis on FESIM/FESOM and highlight the lower complexity, yet high predictive skill of our proposed parametrisation, offering a promising alternative to albedo schemes with explicit melt pond treatment. We will revise the manuscript in the following lines, including changes from the General Comment of reviewer #1:

Abstract
In many sea ice models, a single-category, zero-layer thermodynamic scheme is employed, in which sea ice albedo is prescribed based on surface types depending on snow cover, surface temperature, or sea ice thickness. The Parkinson and Washington parametrisation (PW79) is a commonly used one, which assigns four constant albedo values corresponding to distinct surface types. This parametrisation is too simple to capture the spatiotemporal variability of observed sea ice albedo.
…
…, showing that the equation excels in balancing error and complexity and reduces the mean squared error by about 51 % compared to PW79.

This study applies symbolic regression, a data-driven equation discovery approach, to discover an equation for sea ice albedo directly from observational data, targeting sea ice models which employ the zero-layer scheme with an implicit melt pond treatment (Parkinson and Washington, 1979).~~improving upon the simple PW79 sea ice albedo parametrisation within FESIM.~~

L68-69
2. Do we improve our physical understanding of the surface radiative budget of the Arctic Ocean with our data-driven equation and discover deficiencies in how sea ice thermodynamics are treated when using PW79?

L489 ff. One methodological constraint in our approach is the selection of features that are available on a daily basis across the entire pan-Arctic region and are also represented in sea ice models with implicit melt pond treatment~~FESIM and FESOM, our primary target ocean-sea ice models~~. This deliberate feature selection ensures compatibility with our modelling objectives, but overlooks other relevant factors that may substantially impact sea ice albedo. For example, snow grain size

(Perovich, 1996; Perovich et al., 2002) or black carbon (Hansen and Nazarenko 2004) substantially determine sea ice/snow albedo, but there is no data available on a daily, pan-Arctic scale. Furthermore, melt ponds are known to significantly reduce sea ice albedo, as demonstrated in numerous studies (e.g. Perovich et al., 2002; Webster et al., 2022; Niehaus et al., 2024). Yet, accurately representing melt ponds in ESMs remains challenging since melt pond evolution is highly sensitive to environmental conditions and small-scale processes such as ice topography and drainage (Popović, Silber, and Abbot 2020; Smith et al. 2025). These processes cannot be explicitly resolved and must be parametrised, potentially introducing biases (Smith et al. 2025). Although explicit melt pond schemes remain pivotal for robust polar climate projections, Eq. 4 provides a promising alternative for sea ice models that employ an implicit melt pond treatment. ~~Nevertheless,~~ Since the objective of this study is to capture large-scale patterns at $25 \times 25\,\mathrm{km}^2$ resolution, Eq. 4 sufficiently explains the variance in observed sea ice albedo, indicating that small-scale features may have a limited marginal effect at this scale.

Comment 2

The authors have chosen VIIRS as their "truth". Other daily datasets exist (for example Pohl et al. 2020, which I believe is from the same university as some of the authors). Would it make sense to test the robustness of the method with another "truth"? And if a new dataset with improved accuracy becomes available in the near future, how difficult would it be to reapply the method? Should we expect the function trained on VIIRS to still perform well? I am not necessarily expecting definitive answers, but these questions came to mind while reading and might be worth discussing. Similarly, when a new albedo dataset appears, or when ERA6 becomes available, how much of the workflow presented here would need to be redone? If a new function must be derived, would it be faster to do so because the methodology is already in place?

Response: We thank the reviewer for raising this point. We have looked into other surface albedo datasets like Pohl et al. 2020 which you suggested or Istomina, Niehaus, and Spreen 2025. There were two practical reasons for why we chose VIIRS: extensive temporal coverage and the "right" spatial resolution. Compared to other surface albedo products, the VIIRS/AVHRR product has an extensive temporal coverage from 1982 to present day without gaps, whereas Pohl et al. 2020 covers 2002 to 2011 and the improved version by Istomina, Niehaus, and Spreen 2025 additionally covers 2017 to present day. On the other hand, sea ice thickness data (CS2SMOS) are available from 2010 onwards and snow depth data (AMSR2) are available from 2013 onwards (`https://data.seaice.uni-bremen.de/SnowDepth/Arctic/n25000/`, there are some gaps in summer data between 2002 and 2012). Furthermore, AMSR2 and CS2SMOS provide the coarsest spatial resolution of all data products, which is 25 km, while Pohl et al. 2020 has a spatial resolution of 12.5 km. That would require additional preprocessing of the albedo data. We wanted to avoid preprocessing the target variable since the observations already come with uncertainty. Therefore, we found it appropriate to use VIIRS as reference observation which comes with a spatial resolution of 25 km.

Yes, it makes sense to test the robustness of the method to other datasets, but we need to decide on how much data we want to include for a valid robustness test to make "apple to apple" comparisons. To test the equation to new datasets, we recommend to follow our fine-tuning workflow: split the data into training and validation period, standardise the data and fine-tune the coefficients to the training set and test the validation mean squared error. If the equation does not perform well, that means that the underlying functional behaviour is different in this case we recommend to re-do the whole workflow and find other symbolic forms which describe the data better. Given data availability, it would be also interesting to test the equation for other time periods to test the validity of the functional behaviour discovered by our equation. That is beyond the scope of our study, like a regional analysis more detailed than conducted in Sec. 5, and we are excited to see future work addressing this question.

Comment 3

Linked to the previous point: while the manuscript generally discusses the advantages and drawbacks of the different methods well, it moves a bit quickly over the uncertainties linked to the training data. For instance, ERA5 shows a strong bias in temperature over sea ice (largely because the atmospheric model assumes bare ice with no snow). I am not sure how this affects $\Delta T*$ and the resulting equation, but I would not be surprised if it has some impact. This bias is also regionally dependent, which might contribute to the regional differences the authors observe in Fig. 8e. Could the authors discuss this potential influence and perhaps suggest ways to test it?

Response: We thank the reviewer for pointing out the biases in ERA5 and generally how to deal with potential biases in used data products when applying our proposed workflow. Including suggestions from Comments 7, 8 and 11 of reviewer #1 and Comment 1 of reviewer #3, we will incorporate this information in the revised version.

L244 ff.

Additionally, this may be due to the fact that $T_{2\mathrm{m}}$ can exceed the melting point, whereas $T_{0\mathrm{m}}$ cannot. Another consideration is that *ERA5* does not assimilate sea ice or snow thickness, nor near-surface Arctic observations, except for surface pressure from stations and drifting buoys. Previous studies have shown that this leads to warm temperature biases in *ERA5* over the Arctic, particularly during polar winter clear-sky events (Batrak and Müller 2019; Zampieri et al. 2023). Such biases could introduce inconsistencies between the satellite-derived $T_{0\mathrm{m}}$ and the *ERA5*-biased $T_{2\mathrm{m}}$, which might partly contribute to the predictive skill attributed to $T_{2\mathrm{m}}$. The documented warm bias is particularly large during polar winter stable boundary layer conditions. Our exclusive use of polar-day samples will thus help to mitigate the influence of this bias. However, this documented warm bias in *ERA5* and data inconsistencies between surface and 2 m air temperatures may play some role in our results, although we believe it is unlikely to fully explain the relationship identified here.

L304 ff.

While we expect that ΔT^* is providing meaningful physical information, the seasonal cycle that is reflected in ΔT^* could be influenced by the aforementioned bias in *ERA5* $T_{2\mathrm{m}}$ which is largest during the cold season and not present during summer months. Nevertheless, it does suggest that information on the seasonal cycle is useful in providing a constraint on the surface albedo. Other possible predictors that encode information on the seasonal cycle, such as solar insolation or the surface energy balance, could also provide useful information and could be explored in future work. Considerations of training data biases and prioritisation of predictors that enable results to be generalised across regions and different climate states are important for possible ML-based parametrisations that could be developed based on this work.

L478

While the PW79 sea ice albedo parametrisation only uses the surface temperature as a proxy to define freezing and melting conditions, our equation shows that a weighted temperature difference between the surface and the air at 2 m better encodes information on the seasonal cycle. As our physical interpretation could be influenced by the warm 2 m air temperature biases in *ERA5*, other possible predictors that encode information on the seasonal cycle, such as solar insolation or the surface energy balance, could be explored in future work.

L505 ff.

To test the equation offline to new datasets such as regional subsets or other observational or reanalysis products, we recommend to follow our fine-tuning workflow as we did exemplarily in Sec. 5: split the data into training and validation period, standardise the data and fine-tune the coefficients to the training set and test the validation mean squared error. If Eq. 4 does not perform well, that means that the underlying functional behaviour is different. In this case, we recommend to re-do the whole workflow and find other symbolic forms which describe the data better.~~In future studies, focused on regional-scale modelling, we recommend to optimise Eq. 4 to the region of interest as we did exemplarily in Sec. 5.~~

L508 ff.

In practice, Eq. 4 naturally retains a degree of tunability that facilitates its implementation in ESMs. Integrating Eq. 4 online into an ESM or operational sea ice forecast model would not require substantial changes in existing tuning protocols, as the parameter space can simply be expanded by the seven coefficients of Eq. 4 to obtain physically plausible sea ice states. Under different atmospheric forcings, either in an ocean–sea-ice stand-alone configuration driven by an atmospheric reanalysis product or in a fully coupled configuration, we hypothesise that distinct optimal values of these coefficients will emerge, particularly those controlling ΔT^* where the sea ice model receives $T_{2\text{m}}$ from the atmosphere, since biases in atmospheric temperature fields vary across forcing datasets (Batrak and Müller 2019). This highlights the potential value of regime-aware parametrisations, as suggested by Nath et al. 2026, in which the parameter space is dynamically adjusted in response to the prevailing climate state, allowing the scheme to remain applicable across Arctic, Antarctic, and potentially paleoclimate sea ice regimes.

Comment 4

This may be due to my English not being fully proficient, but calling a satellite product the "ground truth" is a bit confusing to me. I understand what the authors mean—one dataset has to be treated as the truth—but perhaps it could be referred to differently, for example as a "reference observation".

Response: We agree that in this context, calling the satellite observations as "ground truth" is misleading. In Machine Learning, we refer to the dataset that the ML model was trained on as "ground truth". In the revised manuscript, we replace "ground truth" by "reference observation".

Comment 5

Line 34: I believe ESM has not been introduced yet

Response: We will introduce the term ESM when first mentioned in the manuscript.

Comment 6

L114-116 What's the reference that supports this?

Response: The references are the Quality Information Document (`https://documentation.marine.copernicus.eu/QUID/CMEMS-ARC-QUID-002-003.pdf`) and the Synthesis Quality Overview (`https://documentation.marine.copernicus.eu/SQO/CMEMS-ARC-SQO-002-003.pdf`). We will add these references to the respective including changes based on Comment 2 from reviewer #1.

Comment 7

L164: Introduce the SFS acronym

Response: We will introduce the term SFS when first mentioned in the manuscript.

Comment 8

L357: I would recommend adding the direction of the shift compared to observations, as the observation values are quite far above the text.

Response: We will add the direction of the shift in the respective line as follows:

L357

Instead, both 4-feature NN and Eq. 4 demonstrate a notable peak at higher albedo values (0.82 and 0.83, respectively), with Eq. 4 having an upper limit for sea ice albedo at 0.83. The reference observation shows a peak of 0.84. Hence, the models exhibit a slight shift to the left. At the lower end of the albedo scale, the 4-feature NN best captures the long tail, although all model peaks at lower albedo values are more shifted to the left compared to the reference observation with a peak at 0.46.: 0.42 for the 4-feature P3, 0.39 for the 4-feature NN, and 0.42 for Eq. 4.

Batrak, Yurii and Malte Müller (Sept. 2019). “On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice”. en. In: *Nature Communications* 10.1, p. 4170. DOI: 10.1038/s41467-019-11975-3.

Hansen, James and Larissa Nazarenko (2004). “Soot climate forcing via snow and ice albedos”. In: *Proceedings of the National Academy of Sciences* 101.2, pp. 423–428. DOI: 10.1073/pnas.2237157100. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2237157100.

Istomina, Larysa, Hannah Niehaus, and Gunnar Spreen (Jan. 2025). “Updated Arctic melt pond fraction dataset and trends 2002–2023 using ENVISAT and Sentinel-3 remote sensing data”. en. In: *The Cryosphere* 19.1, pp. 83–105. DOI: 10.5194/tc-19-83-2025.

Nath, Pritthijit et al. (Jan. 2026). “Making Tunable Parameters State-Dependent in Weather and Climate Models with Reinforcement Learning”. In: arXiv:2601.04268 [cs]. DOI: 10.48550/arXiv.2601.04268.

Pohl, Christine et al. (Jan. 2020). “Broadband albedo of Arctic sea ice from MERIS optical data”. en. In: *The Cryosphere* 14.1, pp. 165–182. DOI: 10.5194/tc-14-165-2020.

Popović, Predrag, Mary C. Silber, and Dorian S. Abbot (Aug. 2020). “Critical Percolation Threshold Restricts Late-Summer Arctic Sea Ice Melt Pond Coverage”. en. In: *Journal of Geophysical Research: Oceans* 125.8, e2019JC016029. DOI: 10.1029/2019JC016029.

Smith, Madison M. et al. (Feb. 2025). “Formation and fate of freshwater on an ice floe in the Central Arctic”. en. In: *The Cryosphere* 19.2, pp. 619–644. DOI: 10.5194/tc-19-619-2025.

Zampieri, Lorenzo et al. (June 2023). “A Machine Learning Correction Model of the Winter Clear-Sky Temperature Bias over the Arctic Sea Ice in Atmospheric Reanalyses”. In: *Monthly Weather Review* 151.6, pp. 1443–1458. DOI: 10.1175/MWR-D-22-0130.1.

Response to Reviewer 3 Comments

Anonymous Referee

for

Data-driven equation discovery of a sea ice albedo parametrisation

https://doi.org/10.5194/egusphere-2025-3556

Submitted to

The Cryosphere

by

Diajeng W. Atmojo[1,2], Katja Weigel[1,2], Arthur Grundner[2], Marika M. Holland[3], Dmitry Sidorenko[4], Veronika Eyring[2,1]

[1] University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

[2] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

[3] National Center for Atmospheric Research, Boulder CO, USA

[4] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (AWI), Bremerhaven, Germany

Authors' Response to Reviewer 3

General Comments. This is a reference for the submission entitled "Data-driven equation discovery of a sea ice albedo parametrisation". The authors use data-driven approaches to derive an albedo equation for use as a parametrisation in FESIM. This paper is a nice idea, and has novelties. It shows that AI has potential in learning and replacing parametrisations. The purpose is also clearly laid out (e.g. L65-69), well presented.
There are a few important points that I think should be addressed/clarified before recommending it for publication.

Response: We are grateful for the reviewer's careful reading of the manuscript which will improve the manuscript. In the following, we address the reviewer's concerns, mainly regarding the generalisability of our proposed parametrisation, and clarifying our methodology which leads to the functional form of the final equation.

Comment 1

L145 - by doing this you are essentially committing the marginal ice zone. I think this is ok in your analysis but perhaps you should phrase this as a parametrisation for sea ice in the interior ice pack/not so much the MIZ. That is quite important for the future performance of the emulator: we know that the future of the sea ice surface will increasingly resemble fragmented, MIZ-like conditions, with lower sea ice concentration, thus it is a good scientific first step, but there should also be substantial caution in using such an parametrisation in longer term projections because it is not learning to emulate albedo in these MIZ like surfaces/regions.
I think given the future of the sea ice will be very much like the MIZ, and this is approximately/overlapping with what you have ruled out, it should be noted in more than just this line e.g. in the title or abstract that this is an emulator for the central Arctic. By excluding lower sea ice concentrations, and approximately MIZ-like regions it has lesser value for future climate projections (as it is not trained on the regime that will dominate future summers). This needs to be noted more strongly, or all regions (including with low sea ice concentration should be kept in the training if the title/abstract is kept as is. Otherwise "A data-driven equation discovery of a sea ice albedo parametrisation ... for interior icepack/central Arctic ". Essentially its use in climate models for future climate change scenarios should either be addressed by updating the science a bit or just the text should make it clearer as to potential limitations.

Response: We appreciate the reviewer's insightful comment. The reviewer are right to question whether our proposed parametrisation does generalise well to MIZ regions, i.e. other sea ice states than pack ice. We already picked on this point in L148, L440-443 and L501-507 transparently, where we explained that our dataset is dominated by Central Arctic samples spatially and temporally (Fig. 2 in the manuscript) where we mostly find pack ice, and that the equation is likely optimised for the Central Arctic.

That is our motivation to show a regional analysis with the Barents Sea as a use case in Sec. 5.2, where the impact of climate change is greatest in the pan-Arctic region being ice-free in summer, which is projected to be the future state of the Arctic sea ice. In this section, we elaborated on the difference between the whole pan-Arctic, dominated by the Central Arctic, and the Barents Sea, where using our proposed equation, the sensitivity to thin sea ice explains sea ice albedo better than the presence of snow. Since the MIZ is characterised by more fragmented, thinner ice, Sec. 5.2 gives us a first insight on how the equation can be transferred to other sea ice regimes. Therefore, we do believe that our proposed parametrisation still holds value for long term projections, putting into context that some ESMs such as AWI-CM3 Streffing et al. 2022 and operational sea ice models like TOPAZ4b still use very simple sea ice albedo formulations with implicit melt pond treatment like PW79 or a simple sea ice thickness dependent sea ice albedo following Drange and Simonsen 1996, respectively. Our proposed parametrisation gives a better representation of the underlying physics driving sea ice albedo than the aforementioned ones.

To make our parametrisation more suitable for polar climate projections, taking into account sea ice concentration or broadly sea ice state is an appealing idea. Instead of our proposed regional analysis for regional modelling, which gives us a set of coefficients tuned on the subregions, we could follow Nath et al. 2026 which investigates state-dependent tunable parameters using Reinforcement Learning enabling regime-aware parametrisations. Specifically, the seven coefficients from our proposed equation can be used as tunable parameters, which then depend on the sea ice state characterised by a set of variables such as sea ice concentration and thickness. This goes beyond the scope of our study, but it gives a promising direction to more robust polar climate projections.

Overall, we acknowledge the reviewer's point that we should better emphasise that our proposed parametrisation, with the tuned coefficients, is likely biased toward the Central Arctic, and we will add this in the Abstract and revise the Conclusions, including changes suggested by Comment 8 and 11 of reviewer #1 and Comment 3 by reviewer #2. However, we respectfully disagree to add "interior icepack/Central Arctic" to the title since in Sec. 5.2 we proofed that the equation is transferable to different sea ice regimes and that further investigation of tuning the parameters of the equation online, e.g. as suggested by Nath et al. 2026, is needed to make our proposed parametrisation robustly applicable for climate projections.

Abstract

...

Leveraging daily pan-Arctic satellite and reanalyses data from 2013 to 2020 - dominated by conditions representative of the Central Arctic - we apply sequential

feature selection which identifies snow depth, surface temperature, sea ice thickness and 2 m air temperature as the most informative features for sea ice albedo.
...
Unlike NNs, our discovered equation allows for further regional and seasonal analyses due to its inherent interpretability. When fine-tuning its coefficients to regional or seasonal subsets offline, we uncover differences in physical conditions that drive sea ice albedo. As a use case, we further assess the Barents Sea as a contrasting sea ice regime compared to the Central Arctic, showing that the functional form of the equation remains transferable across different sea ice regimes.
...

L473 ff.

This case study gives a first insight on how Eq. 4 can be transferred to different ice regimes, specifically from a stable, pan-Arctic regime dominated by multiyear ice in the Central Arctic to a fragile ice regime characteristic of the Barents Sea. Although we acknowledge that the MSE in the Barents Sea remains relatively large even after fine-tuning (see Sec. 5.1), the functional form of Eq. 4 continues to be physically reasonable. This enables a comparison between the coefficients obtained from global and regional optimisations demonstrating that the optimal coefficients are regime-dependent.

L476

Our best-performing data-driven equation (Eq. 4) combines two mechanisms that critically impact sea ice albedo, likely optimised for the Central Arctic since this region dominates our dataset (64%): high sensitivity to small changes in thin snow, and the temperature difference between the sea ice surface and 2 m air, weighted in a way such as to reflect the current season.

L508 ff.

An important question concerns how well Eq. 4 generalises well beyond Arctic sea ice regimes represented in the training data. Ship-based field measurements in the Antarctic by Brandt et al. 2005 and Tersigni et al. 2025 demonstrate that already a thin snow layer of a few centimetres substantially increases sea ice albedo, emphasising that snow fractional coverage is more impactful than snow thickness.

Here, snow redistribution is mainly driven by strong winds, particularly in the marginal ice zones. Since our dataset has a spatial resolution of 25 km, retrieved h_{snow} likely reflects a combination of snow thickness and fractional coverage. The strong sensitivity of Eq. 4 to small variations in thin snow therefore suggests that the influence of snow on albedo is captured in a physically meaningful way.

Nevertheless, Sec. 5 indicates that the optimal coefficients of Eq. 4 are state-dependent. While the functional form remains transferable, as demonstrated by the Barents Sea case study, the relatively high MSE in this region after fine-tuning suggests that additional processes, such as oceanic heat fluxes, may drive sea ice albedo but are not explicitly represented in Eq. 4. Oceanic heat fluxes also drive sea ice melt in the Antarctic (Brandt et al. 2005), which may influence sea ice albedo, implying that further offline investigations are required to assess the robustness of the parametrisation outside the pan-Arctic region. This is presently limited due to the lack of data availability on a daily, Antarctic scale.

In practice, Eq. 4 naturally retains a degree of tunability that facilitates its implementation in ESMs. Integrating Eq. 4 online into an ESM or operational sea ice forecast model would not require substantial changes in existing tuning protocols, as the parameter space can simply be expanded by the seven coefficients of Eq. 4 to obtain physically plausible sea ice states. Under different atmospheric forcings, either in an ocean–sea-ice stand-alone configuration driven by an atmospheric reanalysis product or in a fully coupled configuration, we hypothesise that distinct optimal values of these coefficients will emerge, particularly those controlling ΔT^* where the sea ice model receives $T_{2\mathrm{m}}$ from the atmosphere, since biases in atmospheric temperature fields vary across forcing datasets (Batrak and Müller 2019). This highlights the potential value of regime-aware parametrisations, as suggested by Nath et al. 2026, in which the parameter space is dynamically adjusted in response to the prevailing climate state, allowing the scheme to remain applicable across Arctic, Antarctic, and potentially future or paleoclimate sea ice regimes.

Overall, our results suggest that the functional form of Eq. 4 provides a physically interpretable representation of sea ice albedo variability, while the optimal coefficients depend on the prevailing climate state. This state dependence implies that the globally optimised coefficient set may not remain optimal across different regions or climate states. Yet, the explicit formulation and limited number of coefficients make the parametrisation well suited for online implementation in ESMs, where the coefficients can be tuned to the model's specific climate regime. Further evaluation, particularly in Antarctic conditions and under future or paleoclimate conditions, will be necessary to assess the broader applicability of the approach.

Comment 2

L181 - when downsampling how do you ensure that you have some sort of representative sample of the whole dataset? I would assume to go from a few million to a few thousand you would need to have some tests/assurance it is representative. I think something needs to be done to show this is somehow a meaningful/representative sample - especially in the context of section 5. Hopefully this could be a relatively simple test.

Response: We thank the reviewer for the comment. We did not clarify that we randomly downsampled the training set to ensure that the training set is representative for the whole dataset. Additionally, we fine-tuned the coefficients on a larger dataset of 10^5 samples, also randomly sampled, leading to a Pareto-optimal equation evaluated on the validation set (L201-202). To clarify, we will change the respective lines to:

L182
Consequently, we randomly downsample the training set to 10,000 data samples, ensuring that the training set is representative for the whole dataset and leveraging the efficiency of PySR in handling limited data.
L201-202
Keeping the physically consistent equations that satisfy all PCs, we perform a secondary optimisation on a randomly sampled subset of 10^5 from the training set.

Comment 3

Eqn 8 - maybe it is better to use something like $\hat{\alpha}$ or some symbol to suggest that this is the albedo according to equation discovery from this dataset, and a proposed relation rather than definitely fundamental for the real albedo. It may help people track it but is a small issue.

Response: We agree that there might be a confusion when using α to describe both sea ice albedo as a physical property and the parametrised, computed sea ice albedo. To avoid redundant sub- or superscripts, we reformulate the manuscript in such a way that when referring to the physical quantity, we write "sea ice albedo", and when referring to the parametrised, copmuted sea ice albedo, we use α.

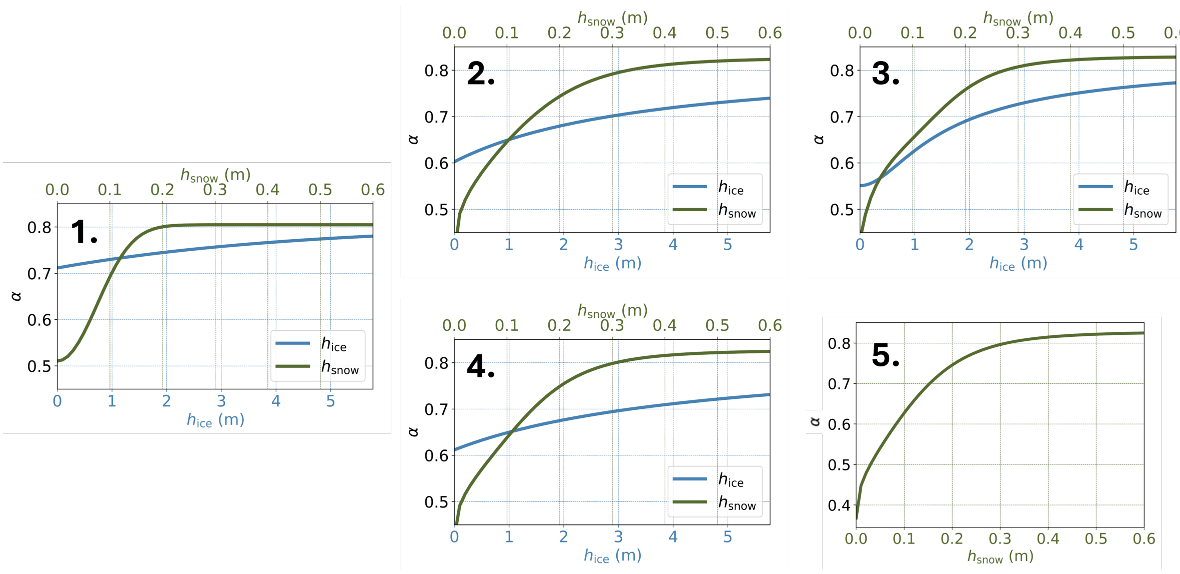
Comment 4

When you refer to Appendix C with the multiple equations it now casts doubt on the ability to infer "physical meaning from the first one". Whilst they have near-identical MSE and lie at the pareto front. Yet their representations/equation forms are markedly different (e.g. presence or absence of tanh, others are quadratic, coupling etc.). This indicates equation discovery problem isn't well-identified: multiple structurally different equations explain the data equally well. I think this should be noted - what you are showing is effectively that you cannot draw a single family of consistent equations using your approach, and it therefore weakens the validity in choosing any given one and then drawing physical meaning from that. Why not choose any another equation with nearly identical MSE and then gain a substantially different physical interpretation?

Response: We thank the reviewer for raising this important concern. We acknowledge that the method and usage of PySR need some further clarification. The reviewer's concern is directly related to L176-185, especially L183-185, where we do not elaborate further which configurations and hyperparameters we used that led to the different symbolic forms of equations shown in Appendix C.

First, PySR is based on genetic programming and randomised search, which is inherently stochastic since the starting candidate equations are drawn randomly. PySR's strength lies in the exploration of a large range of possible solutions, which is not the case for deterministic methods, where we potentially converge to suboptimal or overfit solutions. Furthermore, PySR as a stochastic method reveals uncertainty in the inferred equation, so instead of providing a single equation, PySR returns a set of candidate equations that may fit the data well. This leads to the main clarification that we further need to address in the manuscript: the different PySR configurations.

In PySR, you can define which mathematical operators you would like to include in your hypothesis space. For Eq. 4 (Appendix C 1.), we included trigonometric functions in the hypothesis space, whereas for the other equations, we explicitly excluded them since we aimed to explore other symbolic forms that describe the data well. That is the reason why the candidate equations have different symbolic forms. Nevertheless, when including trigonometric functions, PySR consistently chooses tanh as the optimal solution, which gives us a saturating behaviour shown in Fig. 3 in the manuscript for snow thickness, with high sensitivity to small changes in thin snow. In fact, exploring the other candidate equations where we excluded trigonometric functions, all equations consistently model the same functional response: a saturating behaviour when increasing snow thickness, again with high sensitivity to small changes in thin snow (see Fig. 3).



It is also noted that with the first physical constraint (the sea ice albedo output should lie between 0 and 1), we are explicitly looking for a symbolic form that inherently has an asymptotic behaviour, which is only the case for the hyperbolic tangent function. In a changing climate with different forcings (out-of-sample), there is no guarantee that the other candidate equations retain this asymptotic behaviour. Furthermore, the fact that the MSEs are comparable to the chosen equation can be seen as a coincidence to some extent, one that would no longer hold on a different training set.

Therefore, across multiple runs and variations in the hypothesis space, PySR consistently discovers equations that approximate the same qualitative behaviour. This consistency suggests that PySR is a robust method for identifying the underlying physical mechanisms that are not dependent on the symbolic representation of the equation. However, only the chosen equation includes a hyperbolic tangent function which ensures the asymptotic behaviour that we are looking for as we formulated the first physical constraint. We suggest to reformulate the respective lines as follows to clarify PySR's functionalities and used training configurations:

Line 178-179

PySR is based on genetic programming and implements tree-based candidate solutions with tournament selection, local leaf search, and multiple populations, which is inherently stochastic. PySR's strength lies in the exploration of a large range of possible solutions, overcoming the potential issue to converge to suboptimal or overfit solutions as opposed to deterministic methods.

Line 183-184

We run PySR with varying hyperparameters to explore various symbolic forms that describe the data well, e.g. excluding trigonometric operators, exponents or logarithms. As there is no guarantee that the discovered equations are optimal for their complexity, we perform multiple runs, producing about 800 equations in total.

We suggest explaining the different PySR configurations in Appendix C.

Line 327-328

Figure 6 presents the five best-performing equations in terms of MSE discovered by PySR (see Appendix C for the equations ranked second to fifth with the respective PySR configurations), and baseline models, including polynomials and NNs, on an error-complexity plane (see Sec. 2.2.5).

We noticed that there are some typos in the equations that we will revise in the manuscript:

Line 536-540

1. ...
2. [0.0159/8]: $\alpha(h_{\text{ice}}, h_{\text{snow}}, T_{\text{0m}}) = (-0.92 + \frac{\sqrt{0.32+(0.83T_{\text{0m}}+0.35)^2}}{1.05h_{\text{ice}}\sqrt{h_{\text{snow}}}+1.52(1.04h_{\text{snow}}^2T_{\text{0m}}^2+1.18)^2})^2$

3. ...

4. ...

5. **[0.0180/5]**: $\alpha(h_{\text{ice}}, h_{\text{snow}}, T_{0\text{m}}) = 0.84\left(1 + \frac{-0.79}{2.04\sqrt{h_{\text{snow}}} + (1.24 h_{\text{snow}}^2 T_{0\text{m}} - 1.55)^2}\right)^2$

Line 541 Add this to Appendix C

Equations 1 to 5 are retrieved using different PySR configurations, which are shown in Tab. C1. The hypothesis space refers to the symbolic operators that PySR has access to within a run.

Table C1 (here, Table 1)

...

Exploring the dependency of sea ice albedo on snow and sea ice thickness in Fig. C2, all PySR equations without a hyperbolic tangent function approximate a saturating behaviour, therefore the physical interpretation demonstrated in Sec. 3.2.1 also holds for the other PySR equations.

Figure C2 (here, Figure 1)

...

Table 1: Configurations of PySR runs.

Equation	**Hypothesis space**
1	"div", "mult", "plus", "sub", "pow", "exp", "square", "cube", "sin", "tan", "sinh, "tanh"
2	"div", "mult", "plus", "sub", "square", "cube"
3	"div", "mult", "plus", "sub", "pow", "exp", "square", "cube"
4	"div", "mult", "plus", "sub", "square", "cube"
5	"div", "mult", "plus", "sub", "square", "cube"

Comment 5

Also I think the difference in the given equations should be highlighted/stated strongly here, rather than relegating it to the appendix. That equation discovery cannot give a similar family of equations seems to cast doubt on the approach, drawing in meaning from any given one. I do strongly appreciate these are novel first steps though. But seeing this reduces confidence in drawing physical interpretation from any equation.

Response: We would like to refer to our response to Comment 6, where we explained that different PySR configurations led to different symbolic forms of the equations. These candidate equations approximate the asymptotic behaviour that we seek from the hyperbolic tangent function due to the first physical constraint we formulated and exploring the dependency of sea ice albedo on snow and sea ice thickness, the functional behaviours are similar and therefore, our physical interpretation demonstrated in Sec. 3.2.1 is also valid for the other PySR equations. Hence, we argue that PySR is a robust method and we prefer to explain the details of the other PySR equations in Appendix C as suggested.

Comment 6

Fig 7 - this is quite a substantial difference it looks like both peaks/locations and particularly the magnitude/frequency are substantially different - in fact a frequency of over 4 times as high for the high albedo peak around .8 indicating that the models are not necessarily learning things associated with climate change (and more the central pack). I think it should be more heavily noted.

Response: We would like to refer to our response to Comment 1 including our suggestions. While we pointed out that the training data is dominated by the Central Arctic (also shown in Fig. 2), in Sec. 5.2 we showed that our proposed parametrisation is transferable to other sea ice regimes, namely to the Barents Sea characterised by seasonal sea ice, suggesting that innovative parameter tuning strategies need to be applied for robust climate projections.

Comment 7

L365 - I think it is slightly inconsistent to suggest making a pan-Arctic dataset for the value of including pan-Arctic observations (instead of one e.g. refined on MOSAiC only, and specifically because it includes more than MOSAiC which may be 'the best' for accuracy, but not pan-Arctic) and then when your emulator disagrees with the pan-Arctic data you cite MOSAiC as to why pan-Arctic data can be ignored. I favour your approach and a pan-Arctic design, but I do not think it works to then say that when your data disagrees with your model you can choose to ignore it as it has a wider variation than MOSAiC. You are precisely using pan-Arctic data as it represents more than a narrowed campaign and a narrow window in time, MOSAiC will not capture the full range of processes and values, and thus this is why you choose pan-Arctic data. I think it's better to rewording that there is just a disagreement, and that neither the observational data nor the emulator is necessarily perfect.

A comment after reading section (5): here it focuses on the regional variation and different processes e.g. in the Barents Sea, and this would supports that you can't use a single campaign in a single region to override pan-Arctic conclusions. Thus again I think it is safest to say your emulator simply disagrees with these observations, whilst admitting neither is necessarily perfect (but not that the pan-Arctic observation ranges can be ignored as they don't fit within MOSAiC, in the case that these observations disagree with your model).

Response: We thank the reviewer for raising the inconsistency of reasoning. We acknowledge that saying "We trust pan-Arctic data except when it disagrees with us" is inconsistent. However, we do not believe that it is a contradiction of reasoning to say that we use VIIRS to capture sea ice albedo with an extensive temporal and spatial coverage and at the same time using MOSAiC to further constrain our physical interpretation. Essentially, what we already pointed out is that the measured albedo by VIIRS is a combination of spatially averaged albedo and noise arising from measurement or retrieval uncertainty, while MOSAiC gives highly localised measurements which is better constrained. Put differently, we know that the satellite observations are noisy, while MOSAiC data can be used to constrain the satellite data, so we use them complementary for our reasoning to interpret our proposed parametrisation.

Comment 8

L383 - do you mean you are achieving an R2 score that is good because it is compared to the dataset it is trained on? (It covers 2013 onwards) Have you considered testing on a separate dataset to verify both against? Or likely testing on different times at very least? (Not e.g. 2013 onwards covering the training period) It seems over-enhancing the performance this way.

Response: We thank the reviewer for this helpful observation. Indeed, we use the whole dataset for creating Fig. 9 in the manuscript, including both the training period (2013-2018) and validation period (2019-2020), so PySR did not see the validation period during training. We decided to include the whole time period to include more years, but the reviewer is right to raise their concern since a fraction of the data was already used for training. However, we only used 10^4 data samples for PySR training, randomly sampled from the whole dataset (10^6), and for fine-tuning the coefficients we used 10^5 data samples, also randomly sampled. Therefore, the equation did not see the whole dataset during training. Moreover, looking at Fig. 1, we do not see considerable differences between the two periods. We prefer to include all years to show more years, which does not affect the interpretation, and we will add Fig. 1 in the Appendix for clarity about the different periods.

Comment 9

The PW79 is quite old compared to other potential schemes - why is this one chosen? Surely there are better models and better albedo representations? If AI is to be competitive it should be compared to the more state of the art/most modern approaches/models?

Response: The reviewer is right to point out that more sophisticated sea ice albedo schemes have been developed (see L30-38). However, for an "apple-to-apple" comparison, we need data which come with the same spatial and temporal resolution like the data products we chose for our methodology. As pointed out in L496-498, there is no data available on a daily, pan-Arctic scale e.g. for snow grain size, black carbon or algae which are needed to compute albedo using more sophisticated albedo schemes including explicit melt pond treatment like Holland et al. (2012) or Flocco et al. (2010). We acknowledge

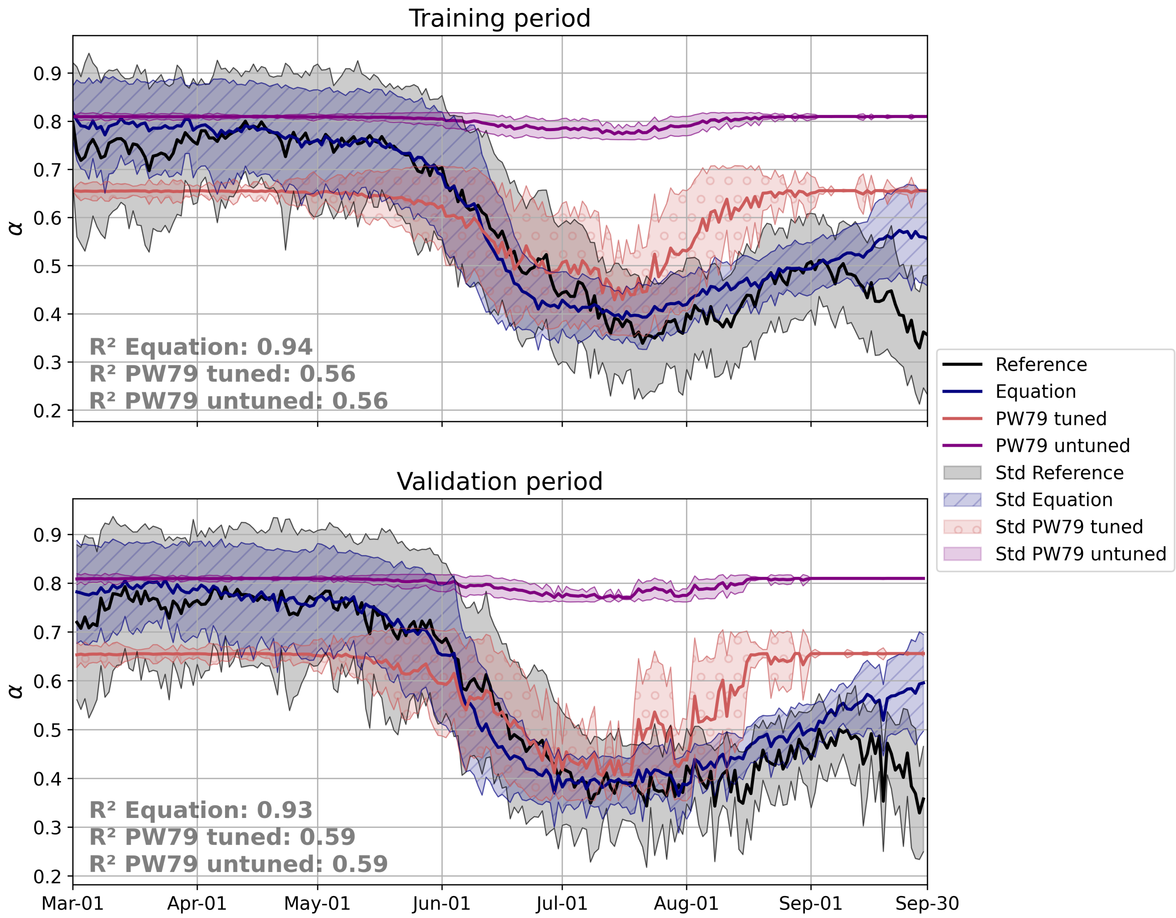


Figure 1: Seasonal cycle of sea ice albedo for different parametrisations during the training period (2013-2018, upper panel) and validation period (2019-2020, lower panel).

that comparing with more sophisticated albedo schemes would significantly strengthen the study. On the other hand, the simple PW79 is still a well-established albedo scheme used in ESMs like in AWI-CM3(Streffing et al., 2022)

Comment 10

Fig 9 I find 4 very similar shadings all overlapping really quite hard to see what is going on.

Response: In the revised version (Fig. 2), we use different shading styles for the equation and the tuned PW79, which hopefully better distinguish the different data.

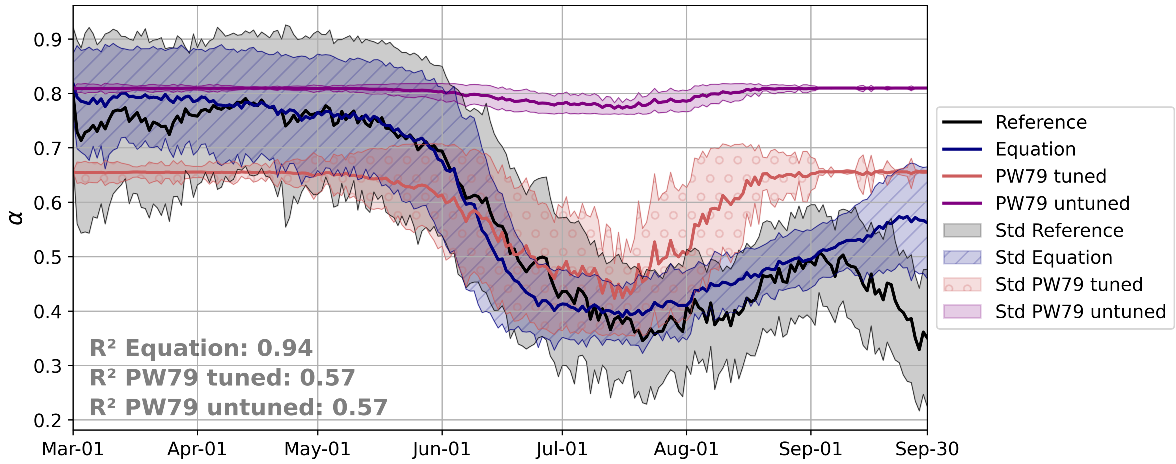


Figure 2: Figure 9 in the manuscript revised.

Comment 11

L474 - as regards again a physically "consistent" equation being used in the conclusions - you derived multiple equations that have substantially different terms i.e. they are inconsistent, yet all have nearly identical MSE. Does this not show that any given one of these is not truly valid for interpretation by definition because all would lead to different interpretations? Again if they were all very similar equations I think this is fairer to say it is consistent, and fair to then draw in meaning. I think you are demonstrating novel approaches, but at the same time currently not demonstrating the case that equation discovery can be used to come up with a single consistent or interpretable form.

Response: Please see our response to Comment 6, where we explained that different PySR configurations led to different symbolic forms of the equations. Nevertheless, they consistently approximate the same saturating behaviour, suggesting that PySR robustly identifies the underlying physical mechanisms that are not dependent on the symbolic representation of the equation.

Comment 12

Is it possible to do this process in whatever way such that you have always multiple models of low MSE (e.g. top 5 models) and that are all very similar equation forms? This to me would be consistent. I find it hard to suggest one equation should be valued above another when their MSEs are so very close. Currently any equation could be chosen pretty reasonably. (Which is almost the opposite of the point of this as such)

Response: We would like to refer to our response to Comment 6. Although the symbolic forms of the equations differ due to different PySR configurations, they consistently approximate the same saturating behaviour. Yet, only our chosen equation including a hyperbolic tangent function ensures the asymptotic behaviour that we seek due to the first physical constraint, keeping the sea ice albedo output between 0 and 1.

Comment 13

In fact the more I think about it, to use equation discovery I think you probably need to show either that you can do this such that your equations are all very similar in the Appendix and therefore you have identified the right data and methods such that you have found the genuine underlying physics, or if not maybe add in all 5 equation terms into the main text analyse them all for physical interpretations/drivers (similar to 3.2) and then explain why their differences in physical interpretations can be overlooked, why one should/could prefer 'the' one equation, and that it is the one you choose. Otherwise it seems like this approach is actually instead just showing that a single equation/specific derived physical meaning in fact cannot be found this way. I would anticipate under future forcings these could lead to notably different projections. Albeit it is indeed novel to try and I like the appeal.

Response: Please see our response to Comment 6, where we showed that all equations exhibit similar physical behaviour. Only our chosen equation explicitly ensures an asymptotic behaviour that we seek as we formulate the first physical constraint - the other equations will potentially fail in an out-of-sample application (e.g. when the forcings are different in a climate change scenario). Furthermore, the fact that the MSEs are comparable to the chosen equation can be seen as a coincidence to some extent, one that would no longer hold on a different training set. Nevertheless, we agree with the reviewer that the equation we identify should not be seen as a natural law. We do not expect there to be a natural law for parametrising sea ice albedo on our resolution. It is merely a good approximation that is also physically sensible.

Comment 14

L495 - I do not fully understand. Do you mean that the AWI will be proceeding with FESIM (and not Icepack) thus it is a pragmatic choice to try to work with this model?

Response: We did not mean to suggest that AWI will proceed exclusively with FESIM instead of Icepack. A coupled FESOM–Icepack configuration has been developed (Zampieri et al., 2021; Hunke et al., 2023), including explicit melt pond physics. However, this configuration is still undergoing validation and testing before it can become the default option. For the upcoming CMIP7 simulations with AWI-CM3 (Streffing et al., 2022), we rely on the more robustly verified FESIM sea ice thermodynamics with the PW79 parameterisation. In this framework, melt pond effects are represented implicitly through the albedo formulation rather than explicitly resolved as in Icepack. Icepack remains available and under active development, and its implementation in fully coupled simulations will be progressing. For now, however, the well-tested FESIM thermodynamic scheme remains the default configuration, which is why we focus on improving this framework in the present study.

Batrak, Yurii and Malte Müller (Sept. 2019). "On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice". en. In: *Nature Communications* 10.1, p. 4170. DOI: 10.1038/s41467-019-11975-3.

Brandt, Richard E. et al. (Sept. 2005). "Surface Albedo of the Antarctic Sea Ice Zone". en. In: *Journal of Climate* 18.17, pp. 3606–3622. DOI: 10.1175/JCLI3489.1.

Drange, Helge and Knud Simonsen (1996). *Formulation of Air–Sea Fluxes in the ESOP2 Version of MICOM*. Technical Report 125. Bergen, Norway: Nansen Environmental and Remote Sensing Center.

Nath, Pritthijit et al. (Jan. 2026). "Making Tunable Parameters State-Dependent in Weather and Climate Models with Reinforcement Learning". In: arXiv:2601.04268 [cs]. DOI: 10.48550/arXiv.2601.04268.

Streffing, Jan et al. (Aug. 2022). "AWI-CM3 coupled climate model: description and evaluation experiments for a prototype post-CMIP6 model". en. In: *Geoscientific Model Development* 15.16, pp. 6399–6427. DOI: 10.5194/gmd-15-6399-2022.

Tersigni, Ippolita et al. (2025). *Seasonal Variability of Snow Cover and Impact on Albedo and Thermal Properties in the Antarctic Marginal Ice Zone*. Version Number: 1. DOI: 10.48550/ARXIV.2507.10916.