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 $T_{2\text{m}}$ 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 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.

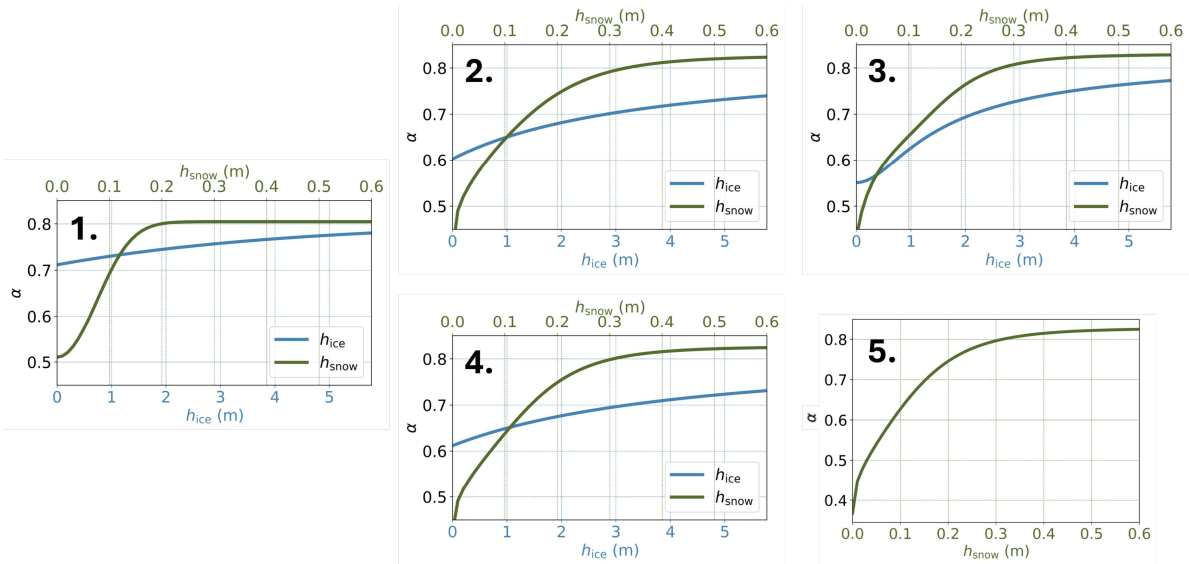
Comment 4

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

3. ...

4. ...

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

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

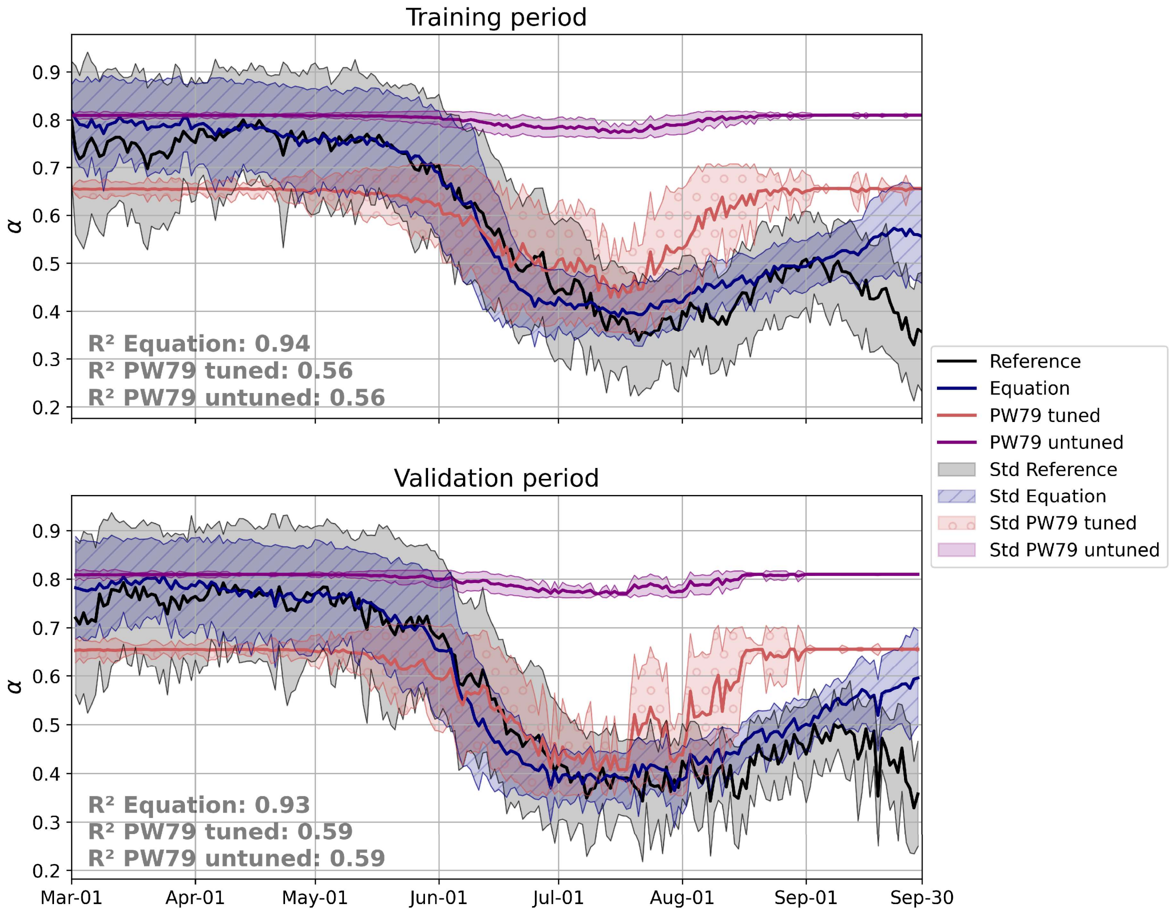


Figure 1: Seasonal cycle of sea ice albedo for different parametrisations during the training period (2013-2018, upper panel) and validation period (2019-2020, lower panel).

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.

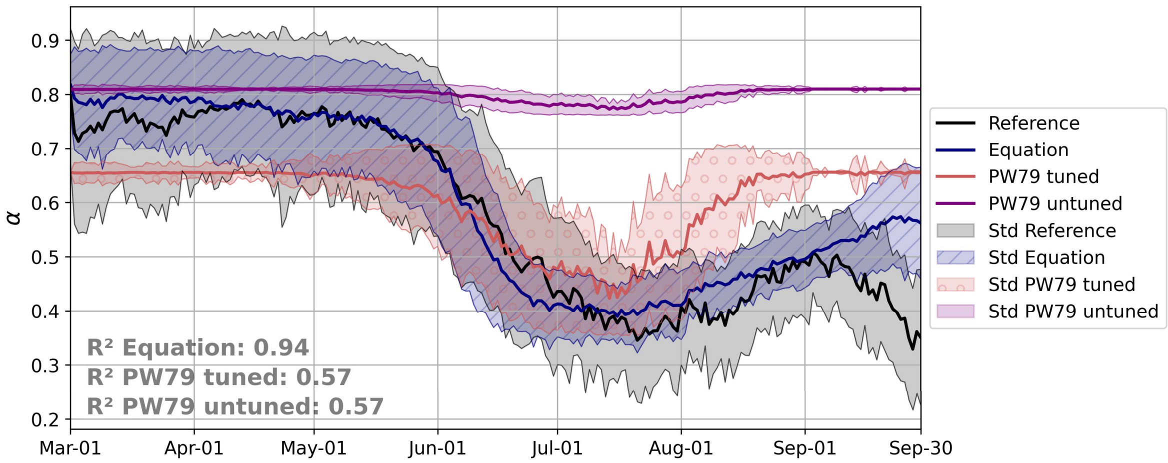


Figure 2: Figure 9 in the manuscript revised.

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.