Review of "Daily melt pond and net ice surface fractions in the Arctic Ocean from MODIS visible imagery: 2000–2024"

This manuscript presents an updated and extended dataset of Arctic melt-pond fraction (MPF) derived from MODIS visible imagery, covering the period 2000–2024 at daily temporal resolution and at two spatial resolutions (500 m and 12.5 km). The dataset represents a substantial extension and methodological update of the earlier Rösel et al. product and is of clear relevance to the cryosphere and climate research communities.

The manuscript is generally well structured and includes extensive validation against independent datasets (high-resolution imagery, airborne observations, and other satellite products such as MERIS and OLCI). The dataset itself is valuable and suitable for publication in a data journal.

However, several major issues require clarification and strengthening before publication. In particular, the manuscript does not sufficiently address the continuity and comparability with the previous dataset, and (ii) the scientific justification for the 12.5 km product relative to the native 500 m dataset. Additionally, aspects of the methodology, uncertainty characterization, and interpretation require further elaboration.

I therefore recommend major revision.

1. Relationship to Previous Dataset (Continuity and Consistency)

The manuscript presents the dataset as an updated continuation of the Rösel et al. (2012, 2013, 2015) product. However, several methodological changes have been introduced, including:

- updated MODIS collection (v6.1),

> See our reply under your recommendations to your review section #1.

- daily instead of 8-day compositing,

> See our reply under your recommendations to your review section #1.

- modified spectral reflectance parameters,

> See our reply to section #3 of your review.

- removal of constraint terms in the optimization,

> We replaced the constraint by using an alternative optimization method with an inherent constraint term that keeps values between 0 and 1. This time, however, the method allows for exact values of 0 and 1, unlike the old constraint term, which only permitted values approaching 0 or 1.

- replacement of pixel-wise optimization by an ANN approximation.

> This approach was already used in the existing product by Rösel et al. (2012, 2013, 2015). We utilized the optimization results from every 50th pixel to create a training dataset for the artificial neural network. Otherwise, it would not have been feasible to compute a daily dataset within a few months.

These changes imply that the new dataset is not strictly consistent with the previous version, but rather constitutes a new product generation.

The manuscript does not quantify differences between the old and new datasets over the overlapping period (2000–2011).

It remains unclear whether the two datasets can be concatenated into a single time series without introducing artificial discontinuities.

> Thank you for pointing us towards this lack of clarity in the description of the dataset. Please see our reply to your recommendation #2 to your review section #1.

Recommendations:

- Include a direct comparison between the previous and current dataset for overlapping years (i.e. spatial comparison maps, time series comparison, bias and trend differences)

> Thank you. This is a very helpful comment. We will include two additional elements in this regard. We will i) provide a figure with panels illustrating how the most recent Rösel et al. (2015) product compares with the new product and ii) provide a figure with panels illustrating how the new product and the existing product compare to the Webster et al. melt-pond fraction data. We will summarize the results also in an additional table. We refrain from a comparison of the trends because the existing dataset is relatively short.

- Clearly state whether the dataset should replace the previous product, or be treated as a separate version.

> Thank you for this suggestion. This new dataset should be seen as a separate product in addition to the existing Rösel et al. (2012, 2013, 2015) data set. The aim was not to continue the dataset.
>
> Since the new dataset spans the period of year 2000 through year 2024, we thought that it is not necessary to state that one should not concatenate the existing Rösel et al. dataset, which ends in the year 2011, with the new dataset
>
> The new dataset uses reflectance values and filters of the MODIS Collection 6.1 product. The Rösel et al. dataset is based on MODIS data of collection 5. This is another reason for not concatenating both meltpond datasets.
>
> The new dataset uses daily reflectance data instead of 8-daily reflectance data as input. The 8-daily reflectance data are a composite of all MODIS tiles aimed towards maximizing clear-sky coverage of the surface reflectance values. However, in the 8-daily reflectance product it is (or at least was) not possible to assign a specific day (of the 8-day period used) to a surface reflectance value. This has been seen as a disadvantage of the 8-daily product because it obscures changes in the melt-pond fraction due to drainage events and because it is more difficult to use the dataset to evaluate daily sea-ice concentration products.
>
> We will state more clearly that i) the new dataset should be considered as a separate product - also why - and that ii) it is not recommended to concatenate both products for the reasons laid out above.

2. Justification of the 12.5 km Product

The manuscript provides both a 500 m and a 12.5 km dataset, where the latter is derived by averaging 25×25 pixels.

Concerns:

- The manuscript explains how the aggregation is performed but does not adequately justify why a 12.5 km product is needed.

> Thank you for voicing your concern in this regard. The main motivation to provide a gridded product at 12.5 km grid resolution is the aim to use this new dataset for the evaluation of sea-ice concentration datasets based on satellite microwave radiometry - as was already done with the existing Rösel et al. dataset (see Kern et al. 2020 in the references of our manuscript). Most of the currently available such products come at either 25 km or 12.5 km grid resolution.
>
> In addition, since many grid cells are missing in the 500 m product—primarily due to cloudy conditions—we (and our predecessor) introduced a 12.5 km product to improve coverage.

- The scientific trade-offs of aggregation are not sufficiently discussed. Please address points like the loss of spatial variability, smoothing of extremes, potential nonlinear aggregation effects, propagation of cloud masking biases.

> Thank you. We would like to emphasize that we published the 500 m product together with the aggregated gridded 12.5 km product. This allows any user to perform their own aggregation method.
>
> We have added the count of cloud-free cells as a parameter (number_of_valid_pixels) to the netCDF file, allowing users to estimate the uncertainty of the aggregated grid cell mean. Naturally, the more cloud-free cells there are, the more reliable the mean value becomes. We refrain from providing an additional data or flag layer with additional information about the reliability of the dataset. But a user can apply the variable mask_90percent_clearsky to ensure that at least 90% of the pixels in the 12.5 km product are clear-sky. We would like to emphasize in this context that neither the MERIS nor the OLCI product provides quantitative information about how many native resolution clear-sky grid-cells contribute to one meltpond fraction value in the 12.5 km gridded products.
>
> Additionally, we have included the standard deviation in the netCDF file (e.g., mpf_stddev), which serves as another indicator of the quality of the mean value.
>
> By averaging over 25×25 500 m pixels, we might smooth extreme values and lose some spatial variability. While one could additionally provide per-grid cell histograms of the melt-pond fraction distribution in the 12.5 km product, we are convinced that a user interested in the spatiotemporal distribution at sub-grid scale will use the 500 m product of the dataset instead. A user would also be able to compute other means of statistics if need be, such as the median and the robust standard deviation.

- The evaluation shows that the 12.5 km product performs slightly worse than the 500 m product, yet much of the analysis focuses on the coarser dataset.

> We think that our manuscript contains a credible amount of evaluation results of the 500 m product. While we could have tried to compare the 500 m product also to other satellite products, e.g. based on the Sentinel-2 MSI sensor, we would like to emphasize that these are also "only" a derived product, not necessarily adding value to our means of evaluation of the new dataset. We have deliberately chosen to use the independent very-high resolution satellite data (Fetterer et al and Webster et al) as well as high-resolution air-borne data from the OIB campaigns. We only chose one product (in the revised manuscript we will add the Rösel et al. 2015 product) to compare the 12.5 km gridded product with (the one by Istomina et al., 2025) because it covers two different satellites (Envisat MERIS and Sentinel-3 OLCI) that overlap in time with our new dataset. We are convinced that this is enough and is also not providing a view that is biased towards the coarser dataset. And of course we find it very important to intercompare these other gridded datasets with the independent datasets as well to have an informed statement about the quality of the three (in the revised manuscript four) datasets investigated. We would like to emphasize also that we provide the reader with examples of how the 500 m product looks like in comparison to the 12.5 km product. In our view this is sufficiently balanced.

- The rationale appears to be mainly consistency with legacy datasets (Rösel et al., MERIS, OLCI), but this is not explicitly stated or critically discussed.

> Our aim was not to demonstrate consistency with legacy datasets but rather to demonstrate how well (or not well) these datasets compare to the new dataset and - more importantly - to the independent data (high-resolution satellite and OIB). We will critically assess our text in this regard and add the respective pieces of information to avoid the impression that we are after consistency. We would like to emphasize in this context that we would have loved to include the fine resolution versions of the MERIS and the OLCI meltpond products but - to our knowledge - these are not freely available.

You may add a dedicated subsection, addressing the consistency with legacy datasets, suitability for climate-scale applications, and the comparison with other satellite products. Please discuss explicitly the advantages and disadvantages of both resolutions, clarify when to use 500 m vs. 12.5 km, and explain the limitations of each dataset. Depending on your output, you may consider emphasizing the 500 m dataset as the primary product, with 12.5 km as a derived product.

> Thank you for this suggestion. As pointed out in our previous reply to your comment, we were not after addressing consistency with legacy datasets. We did, however, compare the new dataset with other satellite products (Envisat MERIS and S3-OLCI); so this is done. We agree with the reviewer to point out a bit more clearly the limitations and potential advantages of both products of our new dataset. Whether this will end up in a dedicated subsection remains to be seen.

3. **Choice of Spectral Reflectance Parameters**

The spectral reflectance values used for the unmixing (Table 1) are derived from visual interpretation of previous literature. In this manuscript, there is no sensitivity analysis provided to assess how these choices affect the results. In addition, this approach introduces subjectivity and may limit reproducibility. Also please give reasons.

Please provide a justification, comparison and appropriate validation of the chosen reflectance values.

I am also missing a discussion on associated uncertainties: Please include a sensitivity analysis showing how variations in these parameters affect MPF retrieval.

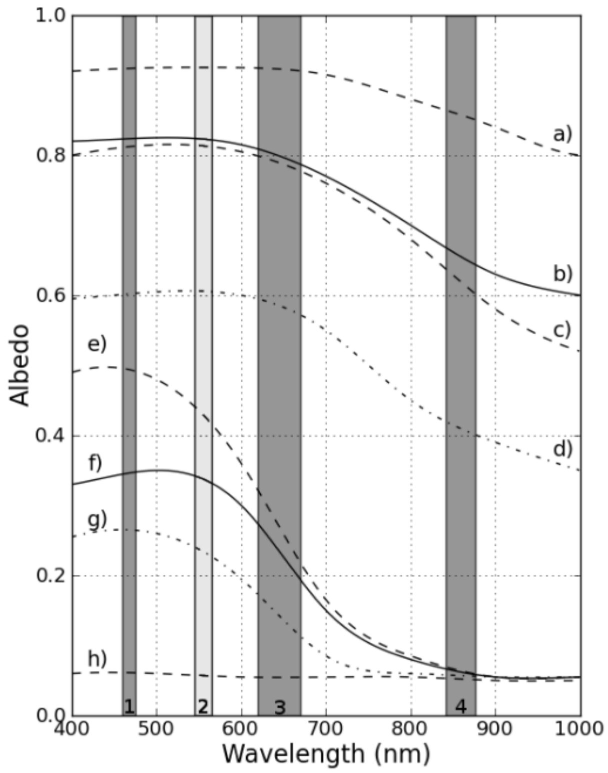

Fig. 2. Spectral albedo values for different surface types on Arctic sea ice: **(a)** Snow-covered ice (dry snow), **(b)** cold bare ice, **(c)** wet snow, **(d)** melting first year ice, **(e)** young melt pond, **(f)** and **(g)** two types of mature melt ponds, and **(h)** open water. The gray columns represent the range of the first four MODIS bands. For our study we use the spectral bands 1,3, and 4 (albedo values after Grenfell and Maykut, 1977).

Figure from Rösel et al 2012

At first, the choice of parameters in the former product was unclear to us. Secondly, the former dataset had an 8% MPF bias, which we aimed to reduce by selecting different, more comprehensible parameters. For this reason, we calculated mean values from the figure in Rösel et al. 2012.

In detail, we averaged all ice values (a, b, c, d) and all melt pond values (e, f, g) to calculate the spectral reflectance parameters r_I and r_M for bands 1, 3, and 4. We used the open water values (h) to define r_W for the same bands.

We then adjusted this initial set of parameters for specific scenarios and compared the corresponding MPF results with the evaluation data. This adjustment process included a test with lower ice values (0.7, 0.69, 0.52 for bands 1, 3, and 4) and lower melt pond values (0.3, 0.19, 0.07), which led us to the following reflectance parameter settings:

MODIS band λ_i	Band width in nm	Meltpond r_M	Snow / Ice r_I	Open Water r_W
1	459-479	**0.3**	**0.7**	**0.05**
3	620-670	**0.19**	**0.69**	**0.05**
4	841-876	0.07	**0.52**	**0.05**

We published accidentally the wrong parameter table, here you can find the right one, which we will put into the revised version.

4. **ANN-Based Approximation**

The use of an artificial neural network (ANN) to approximate the optimization solution is a

pragmatic approach.

One concern is, that the training dataset is limited to selected years (2000, 2011, 2020), which may not capture the full variability of conditions and may thus not be representative.

> Since the optimization method is independent of time and space and depends solely on reflectance parameters, the choice of time periods should not matter. Nevertheless, to ensure we covered most possible reflectance combinations, we selected three different years (and melt-season months within those years). This approach provides a wide range of coverage for the training dataset.

The reported accuracy (99.85%) refers to agreement with the optimization output, not with ground truth.

> This accuracy is reported as part of the methodology section. We think that because of this it is clear enough that this accuracy value does not refer to the accuracy of the meltpond product - especially since the manuscript contains an extensive section about the evaluation of both (500 m and 12.5 km) products of the new dataset.
>
> Unfortunately, we cannot provide the difference between the optimization results and the (unknown) ground truth. However, we can provide the error introduced by the artificial neural network, which is the purpose of the reported accuracy.

Potential temporal biases (e.g., due to sensor changes or climate variability) are not discussed.

> We are not sure we fully understand what the reviewer is referring to here. Using reflectance data of one (the same) MODIS sensor aboard TERRA and of the latest reprocessing of the reflectance data ensures temporal consistency and that there is no need to inter-calibrate data - unlike for time series of passive-microwave based sea-ice concentration datasets. While ageing of the MODIS sensor might have influenced the quality of our dataset we see this as something to be investigated at a later stage. Temporal variability of the reflectance signals as a result of climate variability is taken into account by choosing training data across two decades and across all months of the melting season.

5. Interpretation of Results

The manuscript includes interpretation of regional and temporal MPF variations.

Some interpretations (e.g., trends or regional differences) are speculative given data limitations.

The influence of sampling biases (e.g., cloud cover) is not consistently considered.

> Thank you for pointing these issues out to us. Yes, the time series shown represent a relatively rough first interpretation of the results. We will change the text describing these results towards emphasizing more clearly the limitations of the results shown. We will comment on both, changes in sea-ice concentration and changes in cloud cover, influencing the number of valid data points per region (as indicated by the size of the symbols in the respective figure) and hence the representativity of the region mean and median melt-pond fraction values.

6. Missing a comparison or discussion on S2 MPF fractions (Niehaus et al.): they could be

useful to improve the MODIS MPF, i.e. by using super-resolution approaches

> We agree with the reviewer that we could open the floor and discuss a whole suite of methods that could be used to enhance the product. Using S2 is one of these. However, to our opinion, this does not add clarity to the main aim of this publication which is to present this new dataset. Also, we are convinced that we provided enough discussion about the limitations of the mixed-pixel approach and mentioned the other existing data sets so that another researcher can take it from there, and work on further improvements.