1 Response to Reviewer 2

1.1 Question & Suggestion

Reviewer comment: The coupled RTM is used to simulate TOA reflectance from various sea ice surface and atmospheric properties. The surface parameters are listed but the values were not mentioned as well as the sampling strategy. I cannot figure out how the authors determine the distribution and relevance among the parameters. Similar concern for the atmospheric parameters and the solar/view angles.
Response:
Physical parameters of ice, melt water on ice, and ocean water, physical parameters of snow cover, geometries, and atmospheric characteristics have been included in the revised version. As a reference, this section is included at the close of this response for your review (Tables 2∼5)) prior to our submission of the revised edition.

Reviewer comment: The machine learning method needs more detailed description about how it was used. How to deal with the invalid retrievals from the relationship? Is there a post-processing? It was mentioned there are two models trained. What are their difference and advantages?
Response to 'how post/pre-processing were used to avoid invalid retrievals':
A flowchart (Figure 1) illustrating the process of obtaining a final retrieval model was included in the revised version.

The training dataset only contains snow, ice and melt-pond surface types. As a result, a machine learning classification mask (MLCM) [1] is employed as a post-processing step to filter out invalid pixels.

In addition to 'post-processing', we use auto-associative neural network (AANN) technique as a 'feature selection' tool to avoid the machine learning model obtains invalid retrievals. AANN can effectively avoid 'covariate shift' and ensure that the data range of radiance we use in practice (input from satellite channels) is consistent with that in our training data (synthetic dataset derived from radiative transfer model).

The following paragraphs are included in the revised version to elaborate on the AANN.

---

[1]Chen, Nan, et al. "New neural network cloud mask algorithm based on radiative transfer simulations." Remote Sensing of Environment 219 (2018): 62-71.
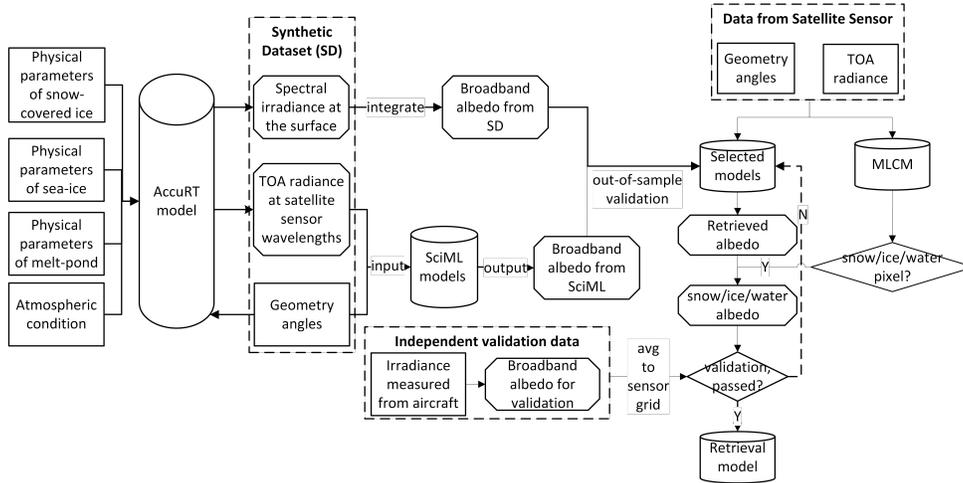
Figure 1: Flowchart of the proposed RTM/SciML framework for albedo retrieval.

With the assistance of the auto-associative neural network (AANN) technique, channels with a significant reconstruction error are deemed unsuitable for use as input to the retrieval model. More specifically, an AANN is trained using the synthetic data generated by radiative transfer model (RTM), which takes as input the three sun-satellite geometry angles as well as all radiance data that meets the aforementioned requirements and outputs all radiances. The trained AANN is believed to have picked up on the patterns in the RTM-generated dataset. Following that, the AANN is fed the same input features derived from satellite sensor. We calculate the absolute percentage error of the reconstruction output and prune channels with an error greater than 5%.

This method is intended to avoid 'covariate shift' — a phrase used in machine learning to refer to the difference between independent variables in training and real-world data. Covariate shift is due to either (a) the saturation of certain satellite channels, which results in a much narrower dynamic range of radiance data from the satellite sensor (real world) than that calculated using RTM (training data), or (b) the response function and wide wavelength range results in a non-negligible difference between the radiance derived from the central wavelength and that obtained from the sensor. It has been demonstrated that the AANN technique is effective in detecting mismatches between data acquired for the retrieval task and data utilized

for training. A recent paper[2] discusses how the AANN approach was used to identify optimal channels for retrieving ocean color products using a variety of sensors.

Response to 'difference between the two neural networks'

This article discussed two neural network models, the only distinction being the wavelength range of the broadband albedo output. The pyranometer and the albedometer have distinct ranges of wavelengths. The albedometer measures spectral irradiance (up/down) in the range of 0.4∼2.1 μm, whereas the pyranometer measures broadband irradiance (up/down) in the range of 0.3∼3.6 μm. As a result, we trained two distinct models to make the most use of the data from the two types of equipment. The revised version includes a 'Data' section that separates the backgrounds of satellite and validation data sources from the discussion of the results. Table 1 is included to show the difference between the equipment types and the matching models.

	Model 1		Model 2	
	λ range (nm)	validation data	λ range (nm)	validation data
Visible	300-700	/	400-700	albedometer
Near Infrared	700-2500	/	700-2100	albedometer
Shortwave	300-2500	pyranometer	400-2100	albedometer

Table 1: Difference between the two models mentioned in the text. Figures 3, 6 and Table A2 show retrieval and validation results of the two models.

Reviewer comment: The author emphasized many times about the advantage of the proposed method than the previous MPD or direct-estimation methods. However, many descriptions needs to be clarified or discussed more. What are the advantage of the coupled RTM rather than the separate radiative transfer models? Is there any quantitative comparison about this? Is a classification within sea-ice surface needed in previous methods? The method is claimed as independent on sensor or spatial resolution, how that is realized without considering the spectral response function difference? Did the previous method restrict to a specific spatial resolution?
Response:

---

[2] *Fan, Yongzhen, et al. "OC-SMART: A machine learning based data analysis platform for satellite ocean color sensors." Remote Sensing of Environment 253 (2021): 112236.*

The following text has been added to the revised version to discuss the advantages of utilizing a coupled RTM over decoupled models.

Due to the fact that uncoupled RTMs make disparate assumptions about the atmosphere and surface when constructing the BRDF, albedo retrieval with uncoupled RTMs is difficult to scale.

While MPD-based approaches are capable of retrieving both the albedo and melt pond percentage for a surface composed of melt water and white ice, they are only effective during particular seasons (discussed in [3] [4] [5]). Additionally, because the spectral reflection coefficients for the melt-pond and thin ice boundaries, as well as the thick ice and snow-cover boundaries, are manually adjusted based on the surface condition, there are greater uncertainties in the retrieval during the transitional seasons of spring-summer and summer-autumn, as well as when the surface is highly heterogeneous (low sea ice concentration, discussed in Istomina2015Melt). Two more issues with the method are its omission of open-water conditions and restriction of sea ice type. The MPD algorithm models sea ice's BRDF exclusively for dry and white ice (based on the absorption of yellow pigments), ignoring the effects of air bubbles and brine pockets (discussed in Zege2015Algorithm).

In the direct-estimation method, linear relations between TOA reflectance and surface albedo are derived for different angular bins. Qu2015Mapping [6] and Peng2018VIIRS [7] generated datasets detailing the intervals of geometry angles in their models, which already have over 40,000 combinations. Multiplying the RTM-simulated or measured surface BRDFs (which range between 100,000 and 120,000 for their retrievals) by the possible atmospheric configurations and by the geometry-angle dataset, results in an extremely big value

---

[3]Istomina, L., et al."Melt pond fraction and spectral sea ice albedo retrieval from MERIS data–Part 2: Case studies and trends of sea ice albedo and melt ponds in the Arctic for years 2002–2011." The Cryosphere 9.4 (2015): 1567-1578.

[4]Istomina, L., et al. "Melt pond fraction and spectral sea ice albedo retrieval from MERIS data–Part 1: Validation against in situ, aerial, and ship cruise data." The Cryosphere 9.4 (2015): 1551-1566.

[5]Zege, E., et al. "Algorithm to retrieve the melt pond fraction and the spectral albedo of Arctic summer ice from satellite optical data." Remote Sensing of Environment 163 (2015): 153-164.

[6]Qu, Ying, et al. "Mapping surface broadband albedo from satellite observations: A review of literatures on algorithms and products." Remote Sensing 7.1 (2015): 990-1020.

[7]Peng, Jingjing, et al. "The VIIRS sea-ice albedo product generation and preliminary validation." Remote Sensing 10.11 (2018): 1826.

for the look-up table (LUT). However, note that only hundreds-thousand level surface conditions were characterized using the monstrous LUT, and because a LUT is essentially a linear regression model, the LUT does not learn the possible interactions between the input features (geometry angles and radiance/reflectance values from various channels), making the approach less efficient.

It is worth noting that the 'RTM-simulated surface BRDFs' mentioned earlier for estimating the radiative properties of sea-ice and snow surface was derived from the IOP model, which is employed by the coupled-RTM that we are utilizing [8]. Because Qu's algorithm decouples the atmosphere from the ocean layer, it is unable to accurately simulate the 'snow-covered sea ice' situation. In a coupled-RTM, snow is simulated as a layer of snow floating on the surface above the interface (the upper slab of the coupled system), and sea ice is simulated as a layer of ice with brine pockets and air bubble inclusions floating on deep ocean water (the lower slab of the coupled system).

Although Qu used the same IOP model to calculate the optical properties of the two media, the 'snow surface' scenario refers to snow that has been placed on land. Additionally, snow possesses complex surface cover (that varies with density, impurity inclusions, thickness, and effective grain size), but the LUT only included a snow layer of a fixed depth. The same issue exists with regard to sea ice conditions. Due to the fact that the direct-estimation algorithm is very dependent on the quality of the LUT, Qu's model's snow and ice retrieval is rather crude in comparison to the more refined approach used in this study.

One last methodological difference is, when using the direct estimation method and building a BRDF table, the reflectance anisotropy (namely, the strong forward peak of snow and ice), which occurs when the solar zenith angle equals the sensor zenith angle, is manually corrected offline (see Qu2016Estimating). However, in our coupled-RTM, the forward peak can be adjusted inside the radiative transfer calculations (described in [9]).

Quantitative comparison between a coupled and decoupled

---

[8]Stamnes, Knut, et al. "Modeling of radiation transport in coupled atmosphere-snow-ice-ocean systems." Journal of Quantitative Spectroscopy and Radiative Transfer 112.4 (2011): 714-726. was cited in the relevant sections

[9]Jiang, Shigan, et al. "Enhanced solar irradiance across the atmosphere-sea ice interface: a quantitative numerical study." Applied optics 44.13 (2005): 2613-2625.

model?

Response:

Unfortunately, there currently is no literature comparing the performance between a 'coupled' and a 'decoupled' RTM. This topic is also out of the scope of the current paper.

Is a classification within sea-ice surface needed in previous methods?

To clarify, the classification in our framework is a 'post-processing' step; it is conducted independently and has no effect on the albedo retrieval results.

For both the direct-estimation method and the MPD-based method, other than cloud-screening, they did not have a post–processing classification algorithm to identify the surface type of each pixel after the surface albedo is retrieved. However, the logic of MPD algorithm requires the surface to be classified and obtains spectral albedo afterwards. In 2.3.2 of Zege's paper, it was mentioned that three channels are used to separate and discard open water pixels, two channels are used to separate the white surface (snow and ice are considered as the same category) from melt pond. The added contents (from the previous response) should help to clarify this issue.

Clarification on the spatial resolution

Other than the MCD43 product, all the other three methods discussed can all retrieve albedo based on single-angular observations. MCD43 requires data from multiple days and therefore the spatial and temporal resolution are lower.

1.2 Minor Comments

Reviewer comment in the context of 'but there is currently no reliable, operational albedo retrieval product capable of assessing the global sea-ice albedo with sufficient spatial-temporal resolution for studies of sea-ice dynamics and for use in global climate models:', **How about the GLASS sea ice albedo and VIIRS sea ice albedo?**
and the relevant comment in the context of 'indicates that the two data sources (measurements and retrieval) are similar, but does not provide statistical evidence for the albedo product's reliability'. **The issue is, the insufficient or unreliable validation data does not means the algorithm/product is not reliable.**

6

Response:

We appreciate that you have brought the GLASS albedo to our attention; it will be added to Table 1.

In the revised version, Sections 1 and 2 have been reorganized to emphasize on the algorithmic difference rather than the use of validation data.

Reviewer comment in the context of 'can be applied to *any* optical sensor that measures appropriate radiance data'. **The word *any* needs to be moderated.**

and the relevant comment in the context of 'The accurate RTM ensures that the 'forward problem' is solved *correctly*' **Reviewer comment:** I would suggest moderate the word. Any model has its limitations. We cannot get the truth by simulation.

Response:

Noted and agree. The sentences have been modified in the revised version.

Reviewer comment MCD43: That is very limited coverage for a cross-comparison

Response:

We agree that the spatial coverage from MCD43 is not ideal. However, MCD43 actually is used by GLASS to 'validate' the visible and near-IR albedo retrieval, as mentioned on page 329 of the following manuscript.

Liang, Shunlin, et al. "The global land surface satellite (GLASS) product suite." Bulletin of the American Meteorological Society 102.2 (2021): E323-E337.

Reviewer comment in the context of "SciML models can be used to solve the 'inverse problem' ": **There might be ill-posed problems.**

Response:

We try to avoid the 'over-fitting' induced by ill-posed problems in two ways:
(1) Generate a large synthetic dataset to provide a large size of training data.
(2) Halt the training when the metrics (mean absolute error) of out-of-sample data does not improve.

Reviewer comment: Looks like a section 1 content. **on the 2.1 section, Existing albedo retrieval procedures and their constraints.**

Response:

In the revised version, Sections 1 and 2 have been reorganized

Reviewer comment: What are the value range of these properties?
Table 2 only includes the IOPs of the surface condition. How about the at-

Response:

Addressed and replied in the previous section.

Response:

In the revised version, this sentence is removed.

We hope the added flowchart in the revised version helps to explain why the framework is 'sensor agnostic'; for other sensors, retrieval can be obtained with the same framework as shown in in Fig. 1.

Adequate treatment of spectral response function is important, as discussed and shown in Figure 1 of the following paper: *Chen, Nan, et al. "Fast yet accurate computation of radiances in shortwave infrared satellite remote sensing channels." Optics express 25.16 (2017): A649-A664.*

Response:

As addressed in the previous section, the MPD-based approach implicitly classifies pixels.

Response:

(1) The values of the hyper-parameters influence model performance, but not to a large extent if the training is successful (i.e., gradient explosion or overfitting are avoided). Take the number of neurons in each hidden layer as an example. In our exploration, we found that a neural network model with two hidden layers and more neurons in each layer (256×512) has comparable results to a model with three hidden layers but much fewer neurons ($16 \times 10 \times 8$). When analyzed with the Shapley value, the two models have learned the same relations from the training data.

(2) Table 1 is included in the revised version to explain why there were two MLANN models.

(3) Table A2 in the Appendix shows the statistics of bias, RMSE, etc.

Reviewer comment: What is the differences between (c) and (d), albedometer vs. pyranometer?

Response: Addressed in the previous section.

Reviewer comment: It would be more clear to list a table to show the statistics, compared to other products.

Response:

Statistics will be added in the revised version.

Appendix

Parameter	Sym.	Unit	Value
Sea-ice thickness	h	m	$0 \sim 3$
Brine pocket volume fraction	V_{br}	–	$(-0.067 \cdot \log(h) + 0.1147) \cdot (1 + 0.2 \cdot r_{\text{bu}})$
Brine pocket radius	r_{br}	μm	$300 \sim 700$
Air bubble volume fraction	V_{bu}	–	$0.0214 \cdot h + 0.0068$
Air bubble radius	r_{bu}	μm	$-18.3 \cdot h^2 + 222.7 \cdot h + 96.5$

Table 2: Physical parameters of ice. In generating the sea-ice thickness, a truncated-normal distribution with $\mu = 0.03$, $\sigma = 1.5$ was used to ensure an adequate amount of thin ice in the SD. The brine pocket radius conforms to a Tukey-Lamdba distribution with λ=0.5.

Parameter	Units	Value
Melt water thickness	m	$0 \sim 1.5$
Chlorophyll concentrations	mg/m^3	$0.5 \sim 10$
CDOM at 443 nm	/m	$0.01 \sim 0.1$

Table 3: Physical parameters of melt water on ice and ocean water. Melt water thickness and CDOM values follow randomly-distributed uniform distributions in the specified ranges. For the chl-a concentration, a reciprocal continuous distribution (long tail extending to high values) was used.

Parameter	Symbol	Units	Value
Snow grain size	r_{e}	μm	$50 \sim 150$
Snow density	ρ_s	kg/m^3	200
Impurity fractions	$\mathrm{f_{imp}}$	-	$10^{-7} \sim 10^{-6}$
Snow thickness	h_{snow}	m	$0.01 \sim 0.2$

Table 4: Physical parameters of snow cover. The snow grain size and snow thickness were generated with a randomly uniform distribution in the specified ranges.

Parameters	Value
Solar zenith angle	20~80 degrees
Sensor angle	0.01~50 degrees
Azimuth angle	0.01~180 degrees
AOD at 500 nm	$0.01 \sim 0.3$
Relative humidity	0.5
Fine mode fraction	0.9

Table 5: Geometries and atmospheric parameters. All parameters conform to random-uniform distributions in the specified ranges.