Referee #1 :

You are absolutely right that just the optical depth, rather than the geometrical one, determines the reflectance. They are related by Eqs. (42). We consider all three parameters, *z*, *H*, and σ_t , as independent ones. We vary all of them independently when fitting spectra and don't make any additional assumptions about σ_t. (Except vertical variability). Of course, we don't have enough information to retrieve the vertical profile of σ_t , so we assume that we retrieve some constant effective value for a layer). Thus, all these three values are retrieved for every spectrum. In Table 2 we show only two of them just for comparison with the *in situ* measured values of *z* and *H*. This information will also be added to the manuscript. However, nobody measures σ_t , so we don't show its values. But we added the retrieved values of σ_t for the light and dark portions of the SHEBA pond (see the last paragraph of Sec. 4.3), where they are important for calculation of the scattering coefficient by bubbles.

Do I understand correctly (from your answers to the other referee) that you are best fitting the observed spectra with the model output spectra, based on triplets of H,z,σ_t values? If that is the case,

a) it should appear clearly in the abstract and in the methodology
b) retrieved σ_t should also appear in Table 2 (even if there is no measured equivalent) and something said about how these values compare to "sensible" values in the literature…

Also, captions for Figs 7, 8, 9, 10, 11 need to contain information about the general locations of each series.

We put in the text about the Polarstern cruise: "The stations, where ponds were observed, were located from 84°3N, 31°7E to 82°54N, 129°47E. For more information about the cruise, see Boetius et al. (2012) and Istomina et al. (2016)."
For Barrow and SHEBA the locations are given: Chukchi and Beaufort seas.

This is not enough. Captions should be self-explaining, with location referring to each pond shown

Fig 8 If these ponds were heterogeneous, then the exact location of the albedo measurement matters! Can this location be shown?

The exact point of the measurement can be seen in the photograph, where a person is taking observation from the light portion of the pond. Unfortunately, there is no photo for the dark one.

This is in Figure 7, not 8. I believe that what the referee would like to see, given the heterogeneity, and I agree, is where on the picture, the measurement was taken. If this info is not available, mention it in, the caption.

Referee #2:

data, which is hardly made by anyone for *in situ* measurements. As far as we're concerned, we think that all these points are stated in the Introduction. Also, according to your advice, we added these points to the Abstract.

Well, I believe this still remains cryptic for a non-specialist in the domain. I suggest that you add a short introductory paragraph at the beginning of section 2 (model description) with a tentative title such as: 2.1 New developments on melt pond reflections model. That paragraph would summarize your arguments presented here above.

o How can or should this model be used in future (the outlook at the very end is rather unspecific and too general)?
o What kind of scientific merit do the authors expect from this and following studies (applications of the model).

Of course, we cannot predict all possible merits. But some applications are obvious: such a model is absolutely necessary for satellite data processing in remote sensing of Arctic ice. Particularly, this model has served as a basis for the MPD (Melt Pond Detector) algorithm for melt pond fraction and sea ice albedo retrieval from MERIS data (Zege et al., 2015).

How did you change the conclusion accordingly?... This again is important for non-specialists to see the broader applications. From the detailed comments, nothing seems to have been changed in the conclusion.

- The comparison with in-situ observations show differences of under-pond ice thickness and water depth of 50% and some even significantly higher. I do not follow the argumentation that this is satisfactory, in particular since there is very little discussion about this (see comments below). I consider these differences as more significant than the discussion reveals. In particular with respect to the under-pond (substrate) thickness, which should be the most important parameter to determine pond albedo.

Actually, the most important parameter that determines the pond albedo is the transport optical thickness of under-pond ice τ_t that is a product of the transport scattering coefficient σ_t and ice thickness H: $\tau_t = \sigma_t H$. Partially this explains the retrieval error: τ_t is retrieved with much higher accuracy, however there is no way to compare it with a measured value. There could be also other different sources of error. First, the under-pond ice might not be flat, especially its lower boundary. In this case the optical retrieval gives some average value, while the *in situ* measurement gives a random value taken in some particular point. From this point of view the measurement makes a mistake, rather than the retrieval. The second source can be the presence of some impurities that affect the absorption spectrum. Additional absorption can affect the retrieval of the scattering coefficient and, consequently, of H. Besides, there could be other sources of uncertainties, like finite pond size, presence of snow in the receiver FOV, clouds in the sky etc. In view of that, the RMS error of 37% seems to us more than reasonable, especially given the fact that the microwave sounding methods fail absolutely in ice thickness retrieval, when ice is covered with a thin water layer.

a) I am sorry, but I don't see how you can judge "accuracy" if you don't know the real/measured value!?!..
b) I don't see these arguments developed clearly in the discussion. Where is this said precisely?

Abstract: The abstract may be significantly improved by adding more results and a statement that explicitly names the additional benefit and further applications of the model:
- Page1/Line15 (P1/L15): … are examined: What is the result of the examination?

We added: "We find that atmospheric correction is necessary even for *in situ* measurements. Thus, an atmospheric correction procedure has been used in the model verification"

This is not enough, in my view…compared to what the reviewer asked

- P1/L17: "good performance" this is rather relative, good in what measure?

How can we measure the adequacy of a model or a theory? This is rather quality, than quantity measure.

Well, to me, by seeing how well it reproduces the observations… and you have the quantitative answer to that in your data…

To find the best fit solution we use the multidimensional Newton-Raphson method with the singular value decomposition of the pseudo-inverse matrix. We really think that the discussion of the method lies far beyond the paper scope, but the method name is added to the manuscript. Adding computational details will make the understanding of the work only harder. Also we are sure that the particular method of searching solution doesn't matter for model verification. It is enough that we find such values of the three pond parameters that give the best fit of spectra in the sense the least squares.

o How did the authors derive that these are the three main parameters. What other parameters were analyzed?

See above our answer about the role of these three parameters. Additionally we can note that refractive indices and absorption spectra of ice and water were not analyzed, because they are fixed, and sediment concentration was not analyzed, because we have no information about polluting substances. So, no more parameters can be analyzed from the point of view of albedo spectrum. Another question is that the transport scattering coefficient consists of the contributions of air bubbles and brine inclusions and thus is determined by their concentrations. Their relationships are considered in detail in Sec. 2.2c and 4.3.

o What about the transport coefficients? How were they studied/discussed?
o How are the thicknesses retrieved?

All three parameters are retrieved in the same manner. They comprise a 3d-vector, which is varied to provide the best fit of spectra. We added this phrase to Sec. 4.1.

a) I did not find the Newton-Raphson method mentioned in the paper

b) Retrieval of the parameters… I think this is something that was unclear to all readers… it is however a crucial output of the paper!... I think it would deserve a "method supplementary material" in which you show on an example how the three parameters are retrieved. Surely there must be a (graphic?) way to show that you chose the best combination of the three parameters by reducing RMS. How do we know this is a "univoqual" solution in each case? How do we know there are no multiple combinations of the three variables resulting in a similarly good fitting of the spectra?

- Section 4.4 should be the main discussion of the comparisons. This is too short and somewhat superficial.
o Where do these rather large differences of 50% come from? I do see various reasons in e.g. pond depth distributions, non-planar interfaces, footprint of sensors compared to pond properties. But this needs to be discussed in more detail.
o What precision may/can be expected in such models?

---

o What determines the uncertainties? Which of the given assumptions might not be ideal, but what would it mean to adapt this? It is most likely not realistic within this study, but some additional discussion would be useful and interesting for further studies.

Throughout the manuscript, making the derivations, we stated the assumptions we use in the model. Surely, every assumption is some approximation or idealization and any of them can limit applicability and accuracy of results. However, the perfect fit of the measured and modeled spectra is a proof that these assumptions were reasonable.

- With respect to those differences: As discussed, impurities are mostly low in the ponds,
so the result is mostly based on scattering (not absorption). In this case, the retrieved spectral shape may be expected to be in good agreement, while amplitude is the main aspect of evaluation. But if then the simulated differences are still around 50% for the under-ice thickness this is somewhat surprising to me. I agree that the RMSE match is quite good if not excellent, but may be not because of the right thicknesses, but other parameters in the model. This should be discussed more.

We think this question is answered in the section *'General comments'* (the 4th question). (Also note that the mean error for ice thickness is 37%, not 50%).

Here the reviewer makes explicit suggestions on the type of questions that should <u>be answered in this discussion section of the text</u>…not only as a response to him!...I don't think this has been done. It is not enough to answer that these matters have been "stated throughout the manuscript".

For all the comments on the Conclusion, the same applies… it is not enough to answer to the reviewer and not change things in the text, and state that "most of the facts are performed in the main text”. Please develop the conclusion as a summary of the main outputs and perspectives of your work.