RESPONSE TO RC2

We thank the reviewer for their detailed assessment and constructive feedback, which has helped us clarify the scope and limits of our proposed methodology.
The answers to each comment are detailed below in blue.

Review of „A Novel Synergistic Approach Using Altimeter Backscatter for an Improved Radiometer Thin Sea Ice Thickness Retrieval" by Stefan Hendricks

In their paper the investigate way to improve the synergy between SMOS L-Band radiometer and CryoSat-2 Ku-Band altimeter data to estimate sea ice thickness in the Arctic. The current state of the art is to merge individually derived sea ice thickness estimates from radiometry and altimetry based on their uncertainties and spatial correlation using an optimal interpolation scheme (ESA CS2SMOS). The authors develop and test an alternative approach, where sea ice emissivity relevant for the L-Band radiometer retrieval are derived from CryoSat-2 observations. The authors then Compare with their own SMOS-only and official ESA SMOS sea ice thickness and find improvement compared to a set of reference data sets.

My Background

I am the current maintainer of the ESA CryoSat-2/SMOS dataset. My background is mostly on radar altimetry and other methods to determine sea ice thickness with geophysical sensors. My review will therefore mostly focus on the backscatter/emissivity relationship and the reference data and less on the emission modeling.

General Comments

The authors have developed a complex retrieval chain that includes many parametrizations and algorithm choices. I find that the derivation of L-Band emissivity from Ku-Band backscatter coefficient requires a set of simplifications and parametrizations that may not be justified for real-life sea ice conditions.

My main concern comes from the fact that the RF-synergy retrieval algorithms results in thin ice (< 1m) data in multi-year sea ice areas where there is certainly no thin ice (upper right panel in Figure 7). The authors correctly state that their algorithm should only to be considered valid for first year sea ice, but first it does not generate confidence in the RF-synergy algorithms maturity if it results in thin ice thicknesses where neither of the two source datasets would indicate any and second, some artefacts seem to be located in areas first-year seas. As far as I can see the thin ice is introduced mainly by the backscatter derived emissivity and I have detailed this concern in the specific comments section.

But I also had problems following some of the authors logic, because I found the manuscript lacking clarity at least in some sections. I have marked relevant passages in the attached document.

And as my last general comment, I would strongly recommend to compare the RF-synergy data not against the ESA SMOS, but against the ESA CS2SMOS, as this would be a fairer comparison.

Regarding the primary concern of anomalous thin-ice retrievals in known multi-year ice (MYI) regions (Figure 7 of the manuscript), and as detailed in the manuscript, at the current stage, our retrieval method is limited to first-year ice (FYI). The integrated modeling currently parameterizes the ice as a two-phase medium (ice and brine inclusions). Expanding this to a three-phase medium to account for the air inclusions of MYI significantly escalates the complexity, as the bulk permittivity must then be treated as a tensor. Because the algorithm is currently limited for FYI, applying it to MYI produces physical artifacts. While we understand the reviewer's concern regarding algorithm maturity, the main objective of this study is not to present a finalized pan-Arctic operational product. Rather, it is a proof-of-concept demonstrating that a multi-sensor synergistic approach can enhance retrieval accuracy specifically within the thin FYI domain. We acknowledge that the current limitation over MYI is a flaw for pan-Arctic mapping, but we view this as a step towards more advanced, physics-aware multi-sensor synergies.

Furthermore, we have thoroughly reviewed the manuscript to improve logical flow and overall clarity. Regarding the comparison with ESA's CS2SMOS instead of the SMOS-only product, we believe that it is fairer to compare with the latter, as we are proposing a retrieval method for thin SIT only, but not for the whole range of SIT. Therefore, even though we are using the same inputs as CS2SMOS, the objective is different, since the ESA's CS2SMOS operational product covers the whole thickness range, while the SMOS product has the same focus as the present method.

Specific Comment: Backscatter – Emissivity Relationship

The authors make the main argument that the Ku-Band backscatter is mainly a function of the dielectric contrast of the snow/ice interface as well as surface roughness. The authors then continue to use the term surface roughness, however there are two different roughness parameters to consider. There is the interface roughness and the large scale sea ice surface roughness, e.g. the surface height standard deviation. Both are relevant parameters CryoSat-2 waveform shape and most importantly, peak power.

The authors are not clear of their definition to as "surface roughness" (I assume it is interface roughness).

The authors also need to make the assumption that that Ku-Band backscatter is solely from snow/ice interface and there is plenty of evidence from field data that this is not the case. It is therefore questionable if dielectric properties of the sea ice surface have a reliable relationship to radar backscatter. Also one issue I did not see discussed, is whether the dielectic properties derived from surface backscatter are applicable for the whole sea ice column.

Clarification on roughness parameters: The reviewer is correct in their assumption; our initial use of "surface roughness" referred to the interface roughness. Large-scale roughness was initially omitted from the modeling because its effects can be mostly averaged out when regridding the high-resolution CryoSat-2 measurements into the coarse SMOS grid. Furthermore, FYI is generally smoother, making this large-scale deformation assumption more reasonable. While roughness uncertainties have been historically overlooked (Landy et al., 2020), we fully agree that their impact requires assessment. Consequently, we have now incorporated a dedicated sensitivity analysis for large-scale roughness into the revised manuscript (the results of which are showcased in our response to the subsequent comment).

Snow/ice interface assumption: Regarding the assumption that Ku-band backscatter originates from the snow/ice interface, we recognize the field evidence demonstrating that the scattering horizon can move vertically into the snowpack depending on the conditions. However, because it remains unfeasible to distinguish the exact scattering horizon from satellite altimetry on a pan-Arctic scale, assuming dominant scattering at the snow/ice interface remains a standard and necessary simplification in current retrieval algorithms (Landy et al., 2020).

Representativeness of the ice vertical column: While the interface properties may not represent the whole vertical conditions, for L-band radiometry, the observed emission is heavily influenced by the near-interface conditions. Furthermore, operational thin ice products (such as the AWI/ESA operational SMOS SIT) also model the ice as a unique homogeneous sea ice layer characterized by bulk temperature, salinity, and dielectric properties. For thin FYI, after the initial rapid growth, the vertical profiles remain sufficiently homogeneous. Regarding the snow-ice interface, brine is usually trapped around it (Nakawo and Sinha, 1981), thus creating a high dielectric contrast that can dominate the L-band emission. Therefore, because the Ku-band radar signal is assumed to be reflected at this same interface, the derived dielectric properties can likely be representative of the primary emitting layer captured by the SMOS radiometer.

Specific Comments: Matching SMRT and CryoSat-2 sigma0

The authors derive the radar backscatter coefficient from the CryoSat-2 waveforms peak power, which is standard practice. But they then face the challenge that backscatter coefficient from SMRT and CryoSat-2 data have a different value range. The authors have developed a scheme to bridge SMRT and CryoSat-2 data and in this part of the manuscript I have the most comments:

Influencing factors on waveform peak power

The peak power of the waveform is determined by the type of surface (specular or diffuse) and their properties. For example, even for a simple homogeneous sea surface, the large scale surface roughness (here defined as surface height standard deviation) changes the waveforms peak power. As an example, I have used the SAMOSA+ waveform model to generate a set of waveforms with the same mean square slope value, but different significant wave height values (linear relationship with surface height standard deviation). Changes in surface roughness is not negligible for first year sea ice, especially in sea ice areas that undergo dynamic thickening due to convergent sea ice motion.

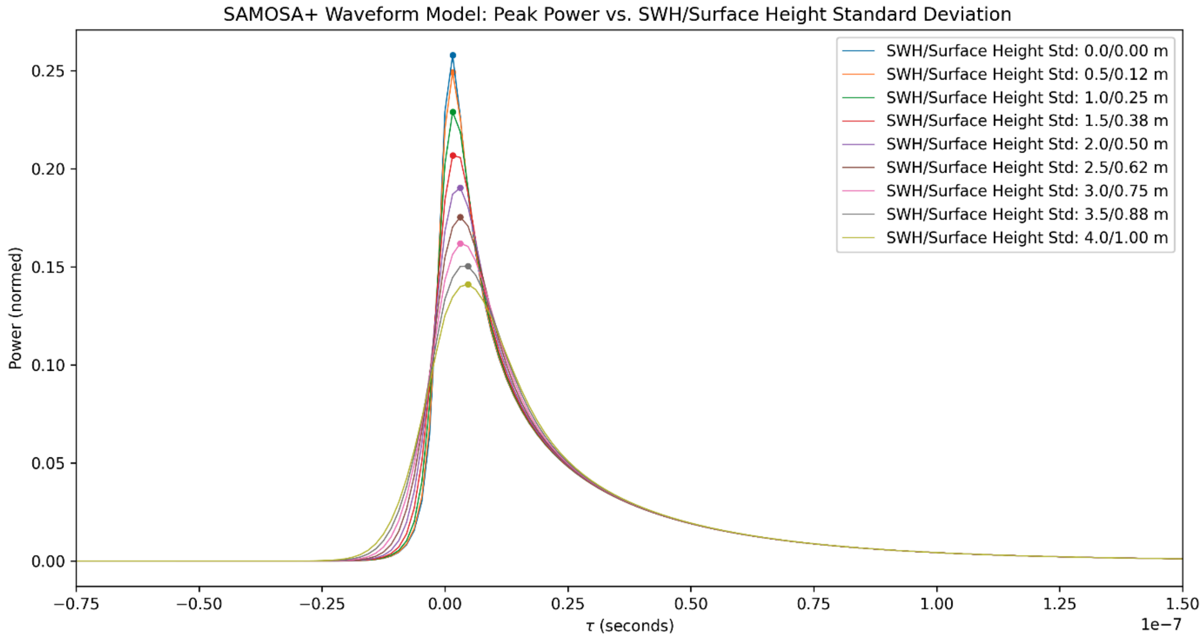

Figure 1:Results from SAMOSA+ waveform model (https://github.com/pysiral/samosa-waveform-model) for fixed interface roughness (nu = 1/mean square slope = 5.0e4) and different surface height standard deviations (implemented in SAMOSA+ via significant wave height). The peak power, and thus the sigma_0 estimate, depends not only on interface roughness but also on the surface height standard deviation.

Also, small patches of young sea ice or leads, can easily dominate a waveforms global maximum, even if located in the off-nadir. This effect may bias the results of the backscatter/emissivity relationship.
Is there any filter for the backscatter values, e.g. only sea ice waveforms? Or all values within a sea ice mask used?

We fully agree that large-scale roughness (as discussed above) and leads can be critical parameters that require further consideration. To address these concerns, we have conducted two new sensitivity analyses, which have now been incorporated into the revised manuscript.

Influence of large-scale roughness: We acknowledge that large-scale surface roughness modulates waveform peak power and, consequently, introduces variance into the simulated backscatter. To quantitatively assess this impact, large-scale roughness has now been integrated into our SMRT modeling via the surface slope (defined in the SMRT documentation as the RMS height of the surface topography). As depicted in Figure R1, an increase in deformation produces a smoothing of the modeled waveform, which results in reduced peak power and lower overall backscatter. This was also done to reproduce the behavior demonstrated in the reviewer's SAMOSA+ example. Although the absolute power values show different ranges due to normalization and the distinction between surface slope and surface height standard deviation, the physical response, an increase in roughness producing a reduction in waveform peak power, is consistent between both models.

In any case, the range between the maximum and minimum backscatter across the permittivity parameter space (i.e., the axis ratio dependency) remains constant, as further confirmed in Figures R2 and R3. This stability is similar to the effects observed with interface roughness, even less impactful, thus justifying its omission in the current state of the proposed method, as its effect is mitigated after the backscatter scaling process.

Influence of leads: Regarding the potential bias introduced by leads and patches of young sea ice, currently no filters are applied to discard these waveforms. However, we have conducted a statistical evaluation to quantify their impact. To do so, the estimated permittivity and the subsequent predicted SIT are collocated with an AMSR2 lead fraction product (Li et al., 2024). Figure R4 shows a daily (December 7, 2019) statistical distribution of the estimated permittivity and the predicted SIT, separated by the AMSR2-derived lead fraction. A qualitative comparison between the lead fraction map (Figure R4, left panel) and the CS2 radar backscatter shows no apparent spatial correlation. More importantly, the real component of the permittivity (computed via the RF-eps model) does not exhibit a significant dependence on the lead fraction. Over solid ice (0% lead fraction) and across varying regimes up to a 10% lead fraction, the median permittivity remains stable. While the real part increases marginally for lead fractions exceeding 10%, this deviation is not significant. A similar stability is observed in the imaginary component of the permittivity, which shows no correlation with the presence of open water. Finally, the rightmost panel confirms that the predicted SIT exhibits no bias as the lead fraction increases. Consequently, these results demonstrate that the omission of a lead-filtering process for the CS2 observations does not introduce significant errors into the final retrieved SIT.

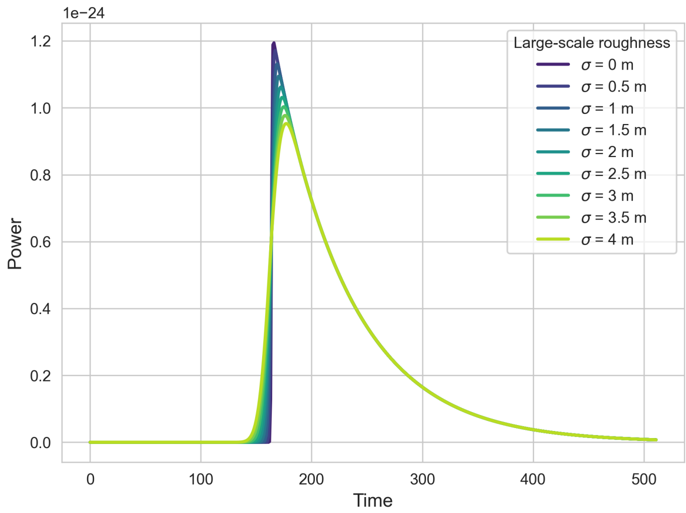

Figure R1. Simulated CryoSat-2 Ku-band waveform across a range of large-scale surface roughness configurations (parameterized via surface slope) using SMRT.

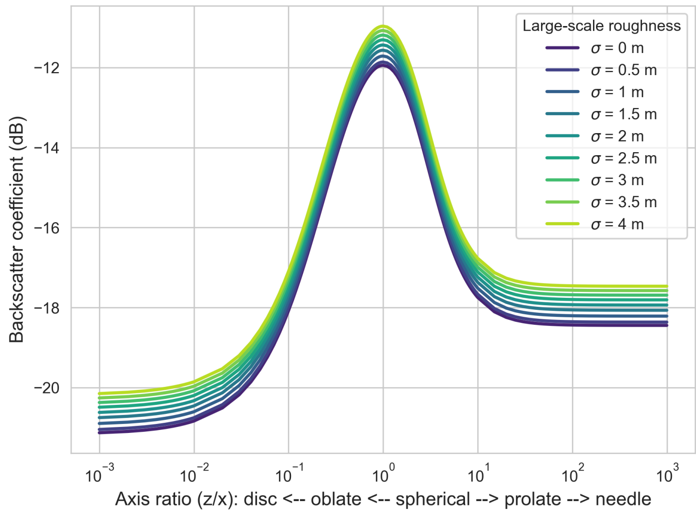

Figure R2. Simulated CryoSat-2 Ku-band backscatter as a function of the sea ice brine inclusion axis ratio across a range of large-scale surface roughness configurations (parameterized via surface slope).

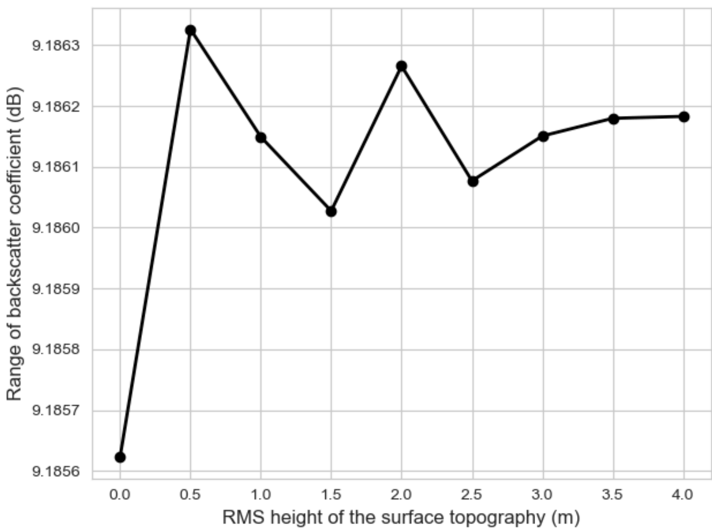

Figure R3. Range (difference between maximum and minimum) of the simulated backscatter as a function of the large-scale roughness.

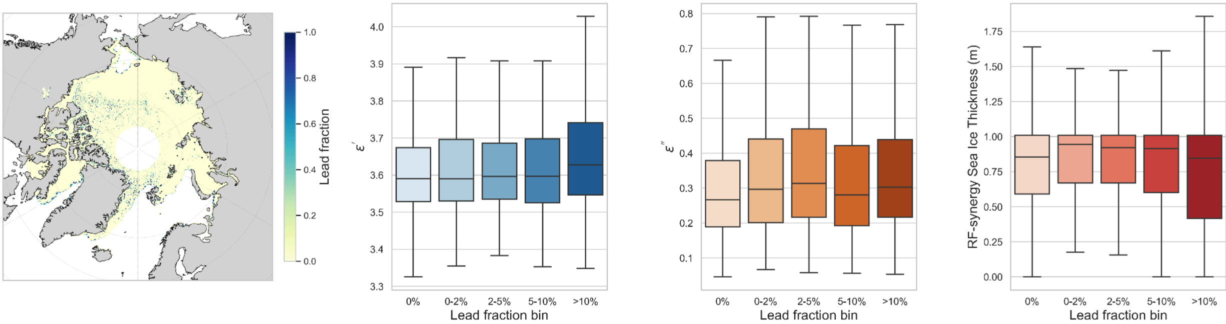

Figure R4. Statistical assessment of the RF-synergy algorithm performance as a function of sub-footprint open water leads for December 7, 2019. The left panel shows the lead fraction map, and then boxplots relating the estimated real and imaginary parts of the permittivity and the predicted sea ice thickness with RF-synergy with the lead fraction, from left to right.

Applicability of SMRT

The authors state that the SMRT altimetry module was "validated in Larue et al., 2021". But this is a paper for land ice, which is a rather different physical environment from sea ice surfaces which are characterized by a strong backscatter heterogeneity. Therefore the term "validated" does is not warranted.

The waveforms simulated are also CryoSat-2 low resolution mode (LRM). For sea ice, CryoSat-2 operates in SAR and partly SARin mode, but there is the pseudo-LRM mode available. What do the authors use here, pseudo-LRM for consistency with SMRT?

Applicability of the SMRT altimetry module: We agree that the heterogeneous nature of sea ice presents relevant differences compared to land ice. While the SMRT altimetry module has not been explicitly validated for sea ice, Larue et al., 2021, demonstrated the reliability of the waveform-simulation module. The electromagnetic radiative transfer mechanisms governing the emission and scattering remain consistent. Therefore, we utilize SMRT not as a perfect sea ice emulator but as a framework to establish the relation between permittivity and backscatter, which allows us to simulate sea ice under many different conditions, including permittivity variations. We agree with the reviewer that the term "validated" might be an overstatement, and we have revised the text to state that the module was "demonstrated" for land ice by Larue et al. 2021, while acknowledging the need for future sea ice-specific model development and validation.

CryoSat-2 operational modes (LRM vs. SAR): Regarding the CryoSat-2 data, we utilize the SAR and SARin data rather than pseudo-LRM. We acknowledge that SAR processing differs in the along-track footprint and the trailing-edge shape of the waveform compared to the pulse-limited LRM waveforms simulated by SMRT. However, our methodology extracts and standardizes the backscatter coefficient derived from the peak power, rather than relying on other waveform features. Because the SAR measurements are subsequently aggregated over a 15-day temporal window and regridded into the coarse SMOS spatial resolution, the localized footprint-scale discrepancies between SAR and LRM modes are effectively averaged out.

Specific Comment: CryoSat-2 backscatter scaling

It is not clear to me how the scaling of CryoSat-2 backscatter works and this need to be better described. Do the authors implement the scaling by mean and standard deviation as in Formular 5? And if yes, does that not mean that the same CryoSat-2 sigma_0 value can be transformed to a range of scaled values, because the mean and standard deviation of the sample space is different? And what is the sample space, the 15-day window? Sigma_0 should have a mean trend as sea ice ages and get rougher. Scaling by mean also carries the risk of removing this trend?

Also the authors state in the conclusion “Furthermore, the influence of the roughness in the radar altimeter backscatter has been effectively diminished with the CS2 data treatment, i.e. the scaling and the re-projection into to lower-resolution radiometer grid”. Again it is not clear what the authors mean by “surface roughness”, but I cannot follow this logic either way. Both interface roughness and surface roughness are a function of sea ice stage of development and not just a random variations around an arbitrary mean state.

I could see that a seasonal correction factor between SMRT and CryoSat-2 data being a viable option. Also the absolute CryoSat-2 sigma_0 value could be used as filter for the validity of the backscatter/emissivity relationship. But I am extremely sceptical about the scaling approach, assuming that I have understood it correctly.

Scaling and temporal evolution: The scaling is implemented using the standard scaler defined in Equation 5. The sample space for this standardization is specifically bound to each 15-day window. Because the mean (μ) and standard deviation (σ) are updated for every temporal composite, the evolution of the ice pack (aging and roughening) is preserved in the shifting statistical parameters. The reviewer correctly notes that an identical absolute $\sigma 0$ value may yield different scaled values across different time periods. This is a deliberate design choice. The retrieval methodology is driven by the relative spatial contrast of the backscatter across the ice pack at any given time, rather than absolute thresholding. Because the absolute power ranges of the SMRT simulations and CS2 observations differ, standardizing the variance of both datasets is the most robust way to bridge the gap.

Mitigating the influence of roughness: Regarding our conclusion that the CS2 data treatment mitigates the influence of roughness, we fully agree with the reviewer that interface and large-scale roughness are physical parameters related to the sea ice development stage. However, the spatial variance associated with large-scale roughness is smoothed when the higher resolution altimetry footprints are averaged and re-projected into the coarse-resolution radiometer grid. More importantly, the scaling approach described by Equation 5 mitigates the impact of roughness. As demonstrated in Figure 3 (of the manuscript) and in our newly added SMRT sensitivity analysis (detailed in a previous response), a variation in the interface or in the large-scale roughness influences the backscatter, but the range between the maximum and minimum across the permittivity space remains constant enough. Therefore, because this processing normalizes the distribution based on these relative ranges, it is effectively mitigating the influence of roughness variation. Consequently, this scaling helps in isolating the underlying dielectric differences, which are ultimately used by the algorithm to estimate the sea ice permittivity.

Specific Comment: Spatial coverage between SMOS and CryoSat-2

How do the authors deal with gap-filling? 15 days of CryoSat-2 at a resolution of 12.5 km does not provide coverage.

To achieve pan-Arctic coverage and address gaps, the CryoSat-2 measurements aggregated over the 15-day window are regridded to the SMOS grid utilizing a nearest-neighbor interpolation. While we acknowledge that this spatial interpolation might introduce some uncertainty, the 15-day window represents a reasonable trade-off between maximizing spatial coverage and minimizing the effects of sea ice drift. To justify this choice and quantify the impact of both interpolation variance and uncorrected sea ice drift within this time window, we have incorporated a temporal mismatch analysis into the revised manuscript.

To quantitatively assess how this temporal mismatch impacts the retrieved SIT, a first perturbation sensitivity analysis is conducted. Synthetic Gaussian noise ($\sigma = 0.5$) is injected into the standardized CS2 backscatter array within a test dataset (composed of a subset of the training data). This perturbation aims to simulate the variance that ice drift and deformation can induce over a 15-day period. This approach was initially favored over a sea ice drift product given the coarse spatial resolution of such products (62.5 km). The perturbed backscatter is propagated through the RF-eps and RF-synergy models, and the absolute deviation in the predicted thickness (ΔSIT) is evaluated against the baseline prediction. As illustrated in Figure R5, the impact of the temporal mismatch under these synthetic random perturbations is dependent on the ice thickness regime. In the thin-ice domain (<0.25 m), where dynamics are expected to be highest, the deviation in the predicted SIT remains minimal. Conversely, the algorithm exhibits higher sensitivity to the perturbed CS2 data in the medium- and thick-ice regimes. The mean deviation rises, flattening at approximately 0.20 m for ice thicker than 1.0 m. In these domains, the L-band TB is saturated, so the algorithm relies more on the CS2-estimated permittivity. However, thicker pack ice is generally more rigid, exhibiting reduced drift speeds and less change over a two-week period compared to the MIZ. Therefore, while the temporal mismatch is an unavoidable constraint, its impact on the retrieval is minimized by the algorithm's nature.

However, to complement this theoretical assessment and ensure the robustness of the method, the multi-sensor EUMETSAT OSI SAF OSI-405-d sea ice drift product is also utilized. To assess the worst-case scenario, a target date (December 15, 2019) is selected at the end of a 15-day compositing window. The daily displacement vectors from December 1 to 15 are accumulated to calculate the drift per pixel over this period. This is subsequently utilized to backtrack the spatial coordinates of the satellite measurements, resulting in the extraction of the drift-corrected CS2 radar backscatter, and contrasting it against the “spatially-static” backscatter. The RF-synergy architecture is then evaluated utilizing both inputs to compute the error (ΔSIT) introduced by the ice drift. As demonstrated in Figure R6, the impact of real sea ice drift differs from the theoretical synthetic noise perturbation test. Even under a maximum 15-day temporal shift, the mean drift-induced deviation (dashed red line) remains below 0.005 m across all thickness categories. Furthermore, the interquartile range (IQR) across every regime is less than 0.01 m. This statistical stability is corroborated by the spatial distribution maps in Figure R7. The difference map (Panel c) illustrates that the discrepancies introduced by the temporal lag are minimal across the vast majority of the Arctic pack, with minor differences bounded within ±0.05 m.

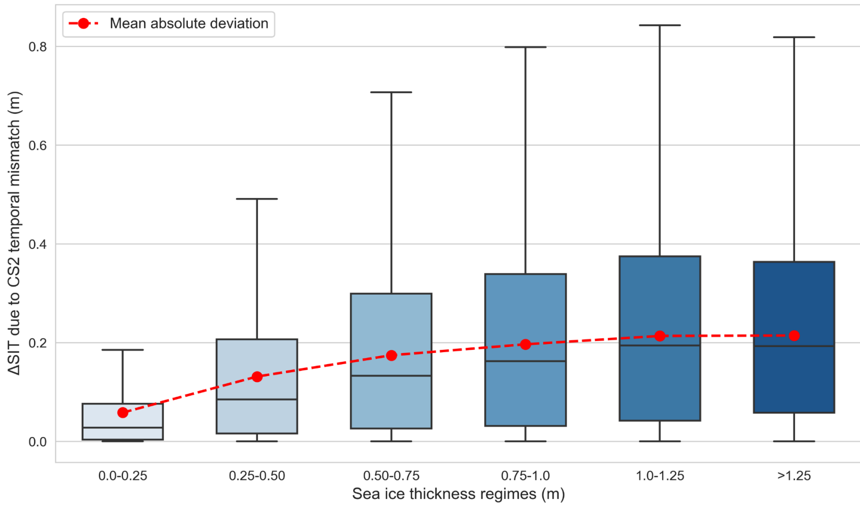

Figure R5. Absolute deviation in the predicted SIT resulting from a simulated 15-day temporal perturbation in the CryoSat-2 backscatter input. The boxplots and mean trendline (red) illustrate the algorithm's varying sensitivity to lagged backscatter data across different baseline thickness regimes.

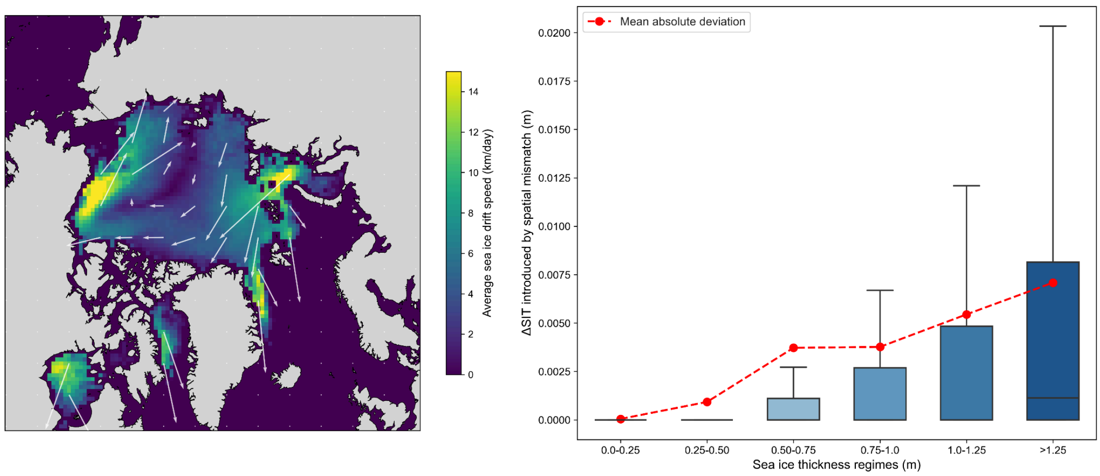

Figure R6. Quantitative assessment of the worst-case temporal mismatch driven by sea ice drift. Left: Full 15-day cumulative sea ice drift speed and direction (December 1-15, 2019) derived from the EUMETSAT OSI SAF OSI-405-d product. Right: Absolute deviation in the predicted SIT resulting from the non-corrected CS2 backscatter. Center lines denote the median, boxes represent the interquartile range (IQR), and the dashed red line indicates the mean drift-induced deviation across thickness regimes.

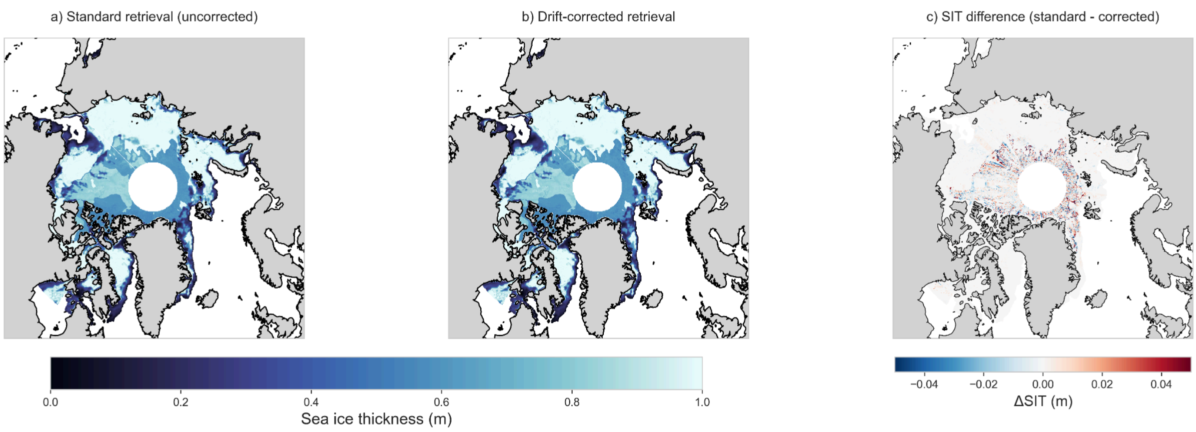

Figure R7. Spatial comparison of the retrieved sea ice thickness for the target date (December 15, 2019). (a) Standard retrieval utilizing the uncorrected, temporally lagged CS2 backscatter. (b) Drift-corrected retrieval utilizing the back-propagated radar data. (c) Spatial difference map.

Specific Comments: Maturity of validation exercise (especially Ship-EM data)

The authors state in the conclusions that "Regarding the theoretical framework, the statistical link between Ku-band backscatter and L-band permittivity derived from SMRT simulations is validated by the improved retrieval performance." Therefore, the validation exercise is of particular improvement for this paper.

The validation exercise however is based on limited set two parameters (mean absolute error and correlation coefficient), which in my experience do not necessarily tell the full story. For example, one can get rather good correlation factors even if the magnitude of the variation is not the same. And for the mean difference does not tell if thickness are systematically over- or underestimated. I therefore strongly adding at least the bias (mean difference).

The authors also state "Given its nature, it allows to continuously monitor the thickness while navigating through the ice, delivering highly valuable dataset for validating remote sensing products" when describing ship-EM data. I recommend caution here. First of all, the SIMS results include snow depth, thus a good match is not necessarily wanted, but the data may also be compromised by representation error, if the ship preferably choses thin ice areas for easier navigation or it speed may dependent on ice thickness. Was this the case? Also, the early cruise report which the authors cite, mentions substantial calibration issues with the SIMS. Was there any quality control and post-processing done on the data?

We agree that relying only on correlation and absolute error might provide an incomplete picture of the method validation. Therefore, we have updated the manuscript's validation to include the bias metric, as well as below in Tables 1 to 3. As shown, the RF-synergy algorithms presents a significantly reduced bias for all the datasets (except for the SEM subset), with majoritarily positive values, and generally around or lower than 10 cm.

Metric	AWI/ESA	RF-standard	RF-synergy
		HEM	
Bias (m)	-0.372	-0.346	0.096
		SEM	
Bias (m)	-0.001	0.005	0.195
		ALS	
Bias (m)	-0.307	-0.291	0.115

Table 1. Bias (mean difference) computed for the SMOSice dataset for the different algorithms.

Metric	AWI/ESA	RF-standard	RF-synergy
Bias (m)	-0.295	-0.214	-0.074

Table 2. Bias (mean difference) computed for the SIMS dataset for the different algorithms.

Metric	AWI/ESA	RF-standard	RF-synergy
		BGEP-A	
Bias (m)	0.352	0.341	0.028
		BGEP-B	
Bias (m)	0.432	0.422	0.069
		BGEP-D	
Bias (m)	0.324	0.336	0.030

Table 3. Bias (mean difference) computed for the BGEP mooring dataset for the different algorithms.

Regarding the specific concerns around the maturity and utilization of the SIMS dataset, we provide the following clarifications:

Instrument calibration and quality control: We acknowledge that the preliminary cruise report documented initial calibration issues. However, during the cruise, the SIMS instrument operated nominally and was calibrated at multiple stations. As shown in Figure R8, the SIMS measurements align with collocated ice drillings conducted directly where the instrument was measuring prior to stopping.

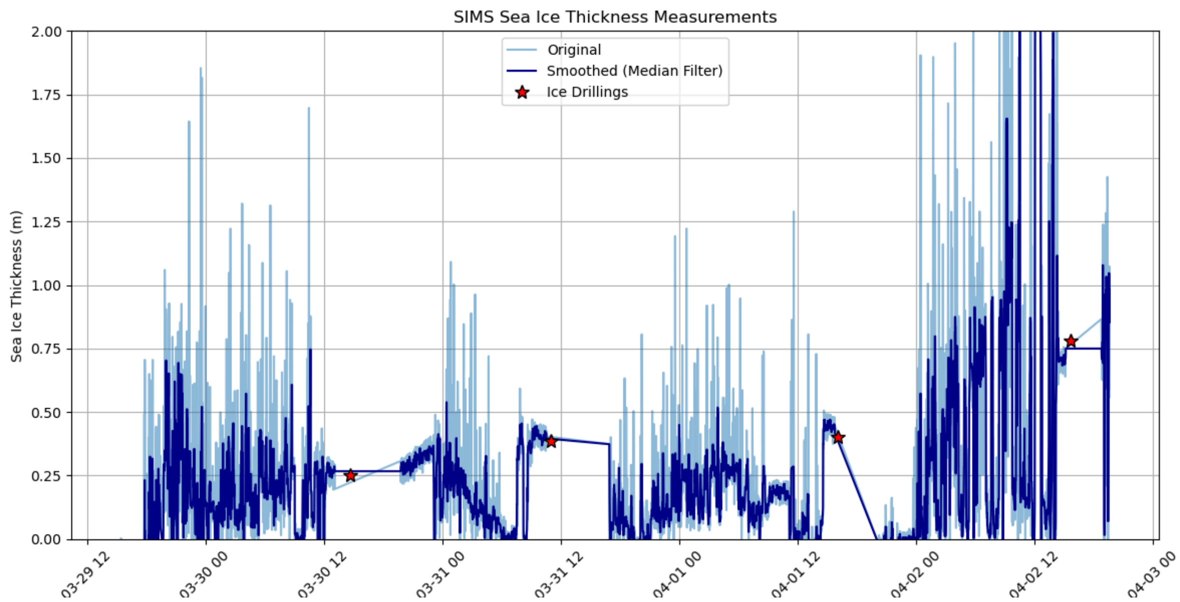

Figure R8. SIMS measurements collected during the campaign. The raw and smooth data is shown alongside the ice drillings collected for validation (red stars).

Representation error and bias: We fully acknowledge the reviewer's point that shipborne measurements are susceptible to spatial representation error, as these ships favor thinner ice to ease the navigation. However, a dataset like this remains highly valuable as it is necessary and convenient for the assessment of thin SIT products, and even more considering the scarcity of other in situ data specifically capturing thin ice. Nevertheless, to mitigate the impact of a possible bias on our validation, the BGEP moorings and the SMOSice datasets, which are independent and cover other areas and time periods, are also included.

Total thickness: The reviewer correctly highlights that the SIMS EM induction technique measures the total thickness (sea ice + snow depth). Because the proposed method and the baseline algorithms to which it is compared retrieve just SIT, a systematic bias can be expected. In any case, given that all the algorithms are treated equally, the same bias is to be expected in all the methods, not favoring one in particular, ensuring a fair validation. Furthermore, introducing an empirical coefficient for transforming total thickness to SIT would also introduce a bias, thus not resulting in a more robust validation.

Conclusions

In summary, I cannot recommend this paper for publication and instead advise major revisions. The author can resolve some of issue by clarifying details, but I also feel that the emissivity needs an uncertainty estimation giving the list of potential issues with backscatter/emissivity relationship and the unclear treatment of the CryoSat-2 backscatter values.

The uncertainty is now included in the manuscript, computed using a Monte Carlo (MC) approach. Particularly, the total predictive uncertainty per pixel is computed using a vectorized MC framework, consisting of K = 100 draws per pixel. However, the generation of the input ensembles differs between the baseline and synergistic approaches. For the baseline SMOS-only retrieval (RF-standard), the uncertainty is driven by the variance in the input feature space (TB, temperature, salinity, and snow presence). To accurately simulate this noise while preserving correlations between variables, the covariance matrix of the input features is computed. This matrix is subsequently used to generate K correlated multivariate normal draws per pixel, ensuring that the simulated variables preserve their statistical dependencies. The baseline RF thus predicts the SIT for each perturbed input, generating a statistical ensemble of predictions. In contrast, the uncertainty in the synergistic approach (RF-synergy) propagates hierarchically. The primary variance originates from the altimetry observations. The estimated permittivity distribution from the first stage (RF-eps), characterized by its ensemble mean and standard deviation, is sampled K times. These stochastic draws are concatenated with the deterministic inputs to generate the final ensemble of SIT predictions in the second stage (RF-sit). For both configurations, the total uncertainty per pixel is computed as the Euclidean norm of the aleatoric component (the standard deviation of the ensemble means) and the epistemic component (the mean of the ensemble standard deviations). Figure R9 presents the spatial reconstruction of the retrieved SIT and the corresponding total uncertainty for a given daily map.

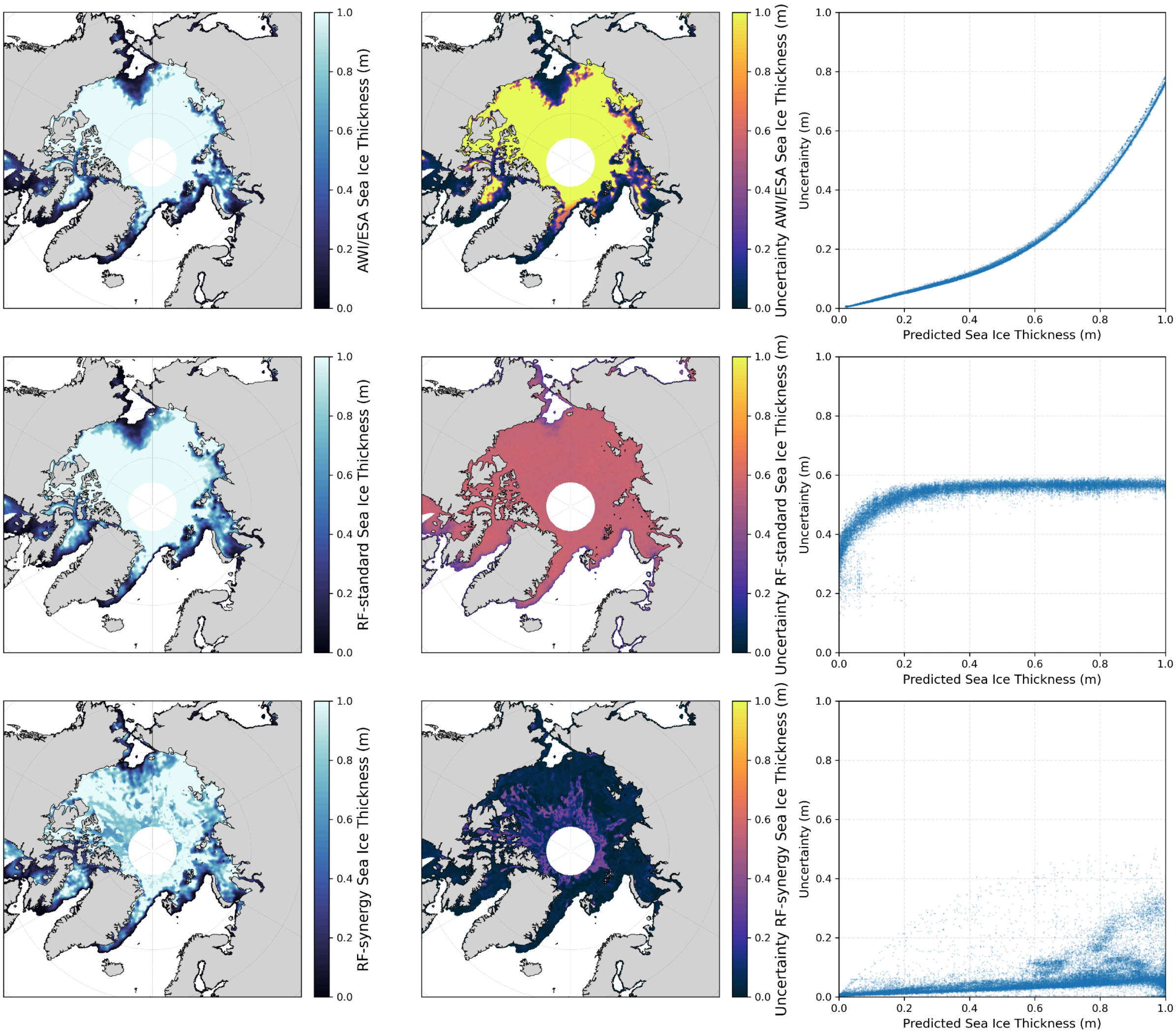

Figure R9. Spatial comparison of the AWI/ESA, RF-standard, and RF-synergy retrievals for December 7, 2019. The left column displays the retrieved sea ice thickness, while the center column maps the total uncertainty computed via the vectorized Monte Carlo framework (for RF-standard and RF-synergy), which is also plotted as a function of the predicted sea ice thickness in the right column for the three algorithms.

The spatial distribution of the retrieved thickness and its uncertainty reflects the physical constraints of L-band emission. In the MIZ and newly frozen areas where the ice is thin (< 0.3 m), the uncertainty for all the models remains low (< 0.05 m). As the ice thickens towards the central Arctic, the signal attenuation increases, and the uncertainty bounds widen. In fact, there is a clear distinction between the uncertainty behaviour of AWI/ESA and that of RF-standard. While AWI/ESA shows a reduced uncertainty for thin ice, which grows exponentially as ice thickens, the RF-standard has a minimum uncertainty of 0.20 m, which grows asymptotically towards 0.60 m, stabilizing there for any thicker ice. As described in Tian-Kunze et al., 2014, the uncertainty in the AWI/ESA retrieval is due to inaccuracy in the SIC, sea ice salinity, sea ice temperature, snow thickness, and also on the statistical thickness distribution function which is applied to correct their estimates, and so the uncertainty is computed with the standard deviations of these variables, thus differing from the approach used for RF-standard. However, the synergistic model's total uncertainty stabilizes (typically remaining below 0.35 m) and presents a different behaviour. Even in the thinner ice areas, especially in the MIZ, the uncertainty shown by the RF-synergy method is reduced compared not only to RF-standard, which is kept around and below 0.10 m compared to the 0.20/0.30 m from RF-standard, but also compared to AWI/ESA even in the lower end of the thickness range.

References

[Landy et al., 2020] Landy, J. C., Petty, A. A., Tsamados, M., & Stroeve, J. C. (2020). Sea ice roughness overlooked as a key source of uncertainty in CryoSat-2 ice freeboard retrievals. *Journal of Geophysical Research: Oceans, 125*, e2019JC015820. https://doi.org/10.1029/2019JC015820

[Larue et al. 2021] Larue, F., Picard, G., Arnaud, L., Ollivier, A., Delwart, S., … Féménias, P. (2021). Snow Altimetry Radar from ground-based to satellite measurements: Validation of the Snow Microwave Radiative Transfer model (SMRT) in the Alps. *Remote Sensing of Environment, 264*, 112615. https://doi.org/10.1016/j.rse.2021.112534

[Li et al., 2024] Li, M., & Liang, X. (2024). *Sea ice lead fraction for Arctic from AMSR2 data, November 2012 to April 30 2020*[Dataset]. PANGAEA. https://doi.org/10.1594/PANGAEA.967508

[Nakawo and Sinha, 1981] Nakawo, M., & Sinha, N. K. (1981). Growth rate and salinity profile of first-year sea ice in the High Arctic. *Journal of Glaciology, 27*(96), 315–330. https://doi.org/10.3189/S0022143000015409

[Tian-Kunze et al., 2014] Tian-Kunze, X., Kaleschke, L., Maaß, N., Mäkynen, M., Serra, N., Drusch, M., and Krumpen, T.: SMOS-derived thin sea ice thickness: algorithm baseline, product specifications and initial verification, The Cryosphere, 8, 997–1018, https://doi.org/10.5194/tc-8-997-2014, 2014