RESPONSE TO RC2

We thank the reviewer for the detailed assessment and constructive feedback. The answers to each comment are detailed below in blue.

General Comments:

Dear Authors,
the manuscript "Observations of Coherent L-Band Emission from Snow-Covered Arctic Sea Ice" investigates whether a coherent or incoherent L-Band emission model is more consistent with in-situ observations from a 2024 measurement campaign in Cambridge Bay. The manuscript is clearly structured and well written. It finds that only the coherent model can reproduce the small polarization ratios (PR) found in some measurements. However, I am concerned about the relevance of this study. It is not new that coherent emission models are required for good predictions of TB emissions in idealized settings. This has also been shown for the effect of snow on sea ice (e.g. Huntemann 2015, Fig. 4.20). The fact that this is confirmed by field measurements is good to see. This might well be the first time such low PD is observed in L-Band radiometry in the field in conjunction with consistent snow thickness measurements (but I am not an expert on this). These low PD measurements seems to be completely driving the improved performance of the coherent model in the manuscript and all further interpretation (which is recognized in the manuscript).
I am skeptical regarding the motivation, that this is relevant for satellite algorithms. For this to be the case the snow thickness variability would, for my understanding, have to be below 5cm within satellite L-Band footprints. I am not convinced, that those conditions exist in a relevant number of cases (considering the large variability of snow thickness and very large satellite L-band footprints). The argument, that these conditions might be found for new ice or land-fast ice, would have to be backed up by literature or data (though I agree, that this would be the most likely places to find them, if they exist).
Of course, the results do not have to be relevant for satellite retrievals in order to be a valuable contribution. Though, in that case the study requires in my view some refocusing. Two ideas come to mind, but of course others are welcome.

(1): Focus on local L-Band measurements entirely, illustrating the potential of ARIEL (-type) sensors mounted on ships, sledges, maybe helicopters. To show its (ARIELs) potential to derive sea-ice and snow parameters it would require to adjust the study setup. Currently you show, how well (or not) the models can be optimized to fit a set of sea-ice and TB measurements. Here you would want to show that one/some of the parameters can be inferred from the others, e.g. by applying your optimization on all but one of the parameters (such as snow thickness) and validate whether the predicted (=optimized) left-out parameter matches the in situ observations. How sensitive would that retrieval be to the other input parameters?

We thank the reviewer for raising this point, which gives us the opportunity to clarify the methodology and its interpretation.

The OE framework we apply is, in fact, a retrieval: it inverts the observed L-band TBs to simultaneously estimate snow depth, snow density, ice thickness, ice temperature, ice salinity, and axis ratio. The retrieved parameter values are not prescribed, they are the outcome of the inversion, and their agreement with independent in situ measurements constitutes very similar to the validation the reviewer requests. The leave-one-out approach suggested by the reviewer was considered but found unsuitable for this work. The TB depends simultaneously on all optimized variables, which are physically coupled through the forward model. If one variable is fixed at its prior mean while the rest are optimized, the remaining parameters will partially compensate due to the nonlinearity of the forward model, potentially hiding the actual contribution of the fixed parameter. The OE framework explicitly accounts for these interdependencies by jointly adjusting all parameters for each measurement, so isolating one variable is not straightforward without breaking this physical consistency.

An important remark here. The OE framework used in this study, is not really used as a retrieval would be used on satellite data. In our case the priors on the geophysical data are tightly bound by the observations and their corresponding uncertainty. This ties the model much more down to a state very close to these tight priors in the state space. On the other hand, in the instrument measurement space, the brightness temperatures and their uncertainties determine how the model may diverge from the observation. Essentially we are only balancing the geophysical measurements and the brightness temperature measurement, related over the forward model by the cost function Eq. 2. We will clarify this important distinction from the classical OE usage in the revision. See also our detailed answer to point 4 of reviewer 1.

(2): Focus on validation and development of L-Band emission models. You show that one coherent model is more consistent with observations than one incoherent. This seems to be a good dataset to compare a number of available L-Band models to each other and see which ones are most consistent with real field measurements. Is the focus on thin snow layers in MEMLS a real limitation or is coherence lost in the data above 5 cm snow thickness anyways (as one could argue the equal performance of coherent and incoherent model for snow thicknesses near 15cm in your figure 5b suggests). If other coherent L-Band emission models are available, one would want to include them, too. The methodology presented here seems adequate for this investigation as it is. The description of the emission models and their differences would need to be expanded. It would be important to make sure that the comparison is not simply selecting the most flexible model.

We appreciate this suggestion. However, the primary objective of this study is not to perform an intercomparison of different L-band emission models or to identify the model with the highest accuracy. Rather, our goal is to investigate whether coherent effects need to be accounted for in order to reproduce the observed in situ measurements. For this reason, we compare two modelling frameworks that differ in their treatment of wave propagation: one incoherent and one coherent variant of otherwise models that can be considered identical in practice. The intention is to assess whether the inclusion of coherent effects provides a more realistic representation of the observations, rather than to evaluate the relative performance of multiple models within the same framework. While additional coherent models could certainly be included in a model intercomparison study, we do not expect the conclusions regarding the importance of coherent effects to depend on the specific coherent implementation used. Therefore, we consider the comparison between coherent and incoherent approaches to be the most direct way to address the scientific question posed in this work.

To avoid the impression that the results are driven by differences in model flexibility rather than by the treatment of coherence itself, we have expanded the description of the modelling approaches and clarified the scope and objectives of the study in the revised manuscript.

Specific Comments:

As this will require a major revision, I will not go into detail regarding specific comments, but want to give some areas which will need attention:
-Description of the field sites: Ideally with a map of the measurements or at least a more general description. It is not clear what it means if measurements are described as 'close', it is not clear how far TB measurements are from each other, what it means if properties are described as 'homogeneous', etc.

The field sites descriptions have been enhanced in the manuscript.

-The naming of TB and sea-ice conditions ('measurements' and 'observations' respectively) is very confusing. The optimized x vector is repeatedly called observation even though \alpha is not measured at all and the other variables are optimized to match a self-consistend model state including the TB. TB and x are also called prior which is typically what we call expectations before they are updated by observations. I understand it refers here to the values before optimization, but since it is also the target (i.e. a small difference between optimized and its prior is considered a measure of model quality here) the naming is confusing.

The naming has been clarified in the manuscript.

-You never show simple forward simulations. What TBs do the models predict for x_a? This should be an intuitive first measure of how well the emission models perform to accompany by your optimization results.

We kindly appreciate this comment and we agree with it. The "without the OE" comparison is hard to evaluate. The metric for evaluating a mismatch of TBs to a given state with uncertainty, related over a model, is essentially the OE cost function in Eq. 2. However, we will perform the forward simulation for the given physical state and the corresponding evaluation of the brightness temperature mismatch below.

For assessing and showing how the models perform without OE, now Figure 4 has been split in two figures. The geophysical variables that are optimized will be shown in the same figure, but the original third row will be included as a new figure (shown below, Figure R1). This will show not only the ARIEL-measured and the optimized TBs, but also the TBs computed with the prior values of the geophysical variables. The optimized value generally lies between the observation and the prior simulation. This figure decoupling is also very convenient because now there is one plot for the geophysical priors (the x_a) and the measurements (y), thus having both parts of the cost function discussed separately.

To further assess this, both the coherent and incoherent forward models can be run using fixed prior parameters (the same prior mean values used in the optimization) together with the in situ measured snow depth and sea ice thickness, without any parameter optimization (see Table 2 in the manuscript). The results are presented in the new Figure R2 below, clearly showing that the coherent model better matches the measured TBs, particularly for TBH. In contrast, the incoherent model exhibits an almost flat response, failing to follow the 1:1 line.

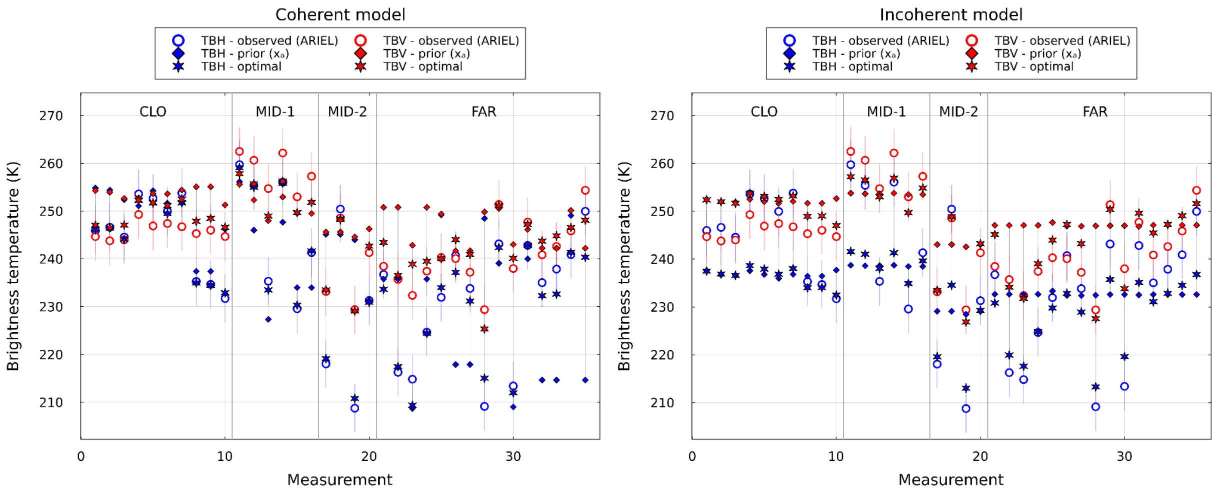

Figure R1. Brightness temperatures TBH and TBV for each measurement in each sampled site, for the coherent (left) and incoherent (right) forward models. The optimal values are shown as stars, the ARIEL measurements are shown as open circles with their uncertainties as error bars, and the brightness temperatures obtained by evaluating the forward model on the prior geophysical state x_a (i.e. without optimization) are shown as diamonds.

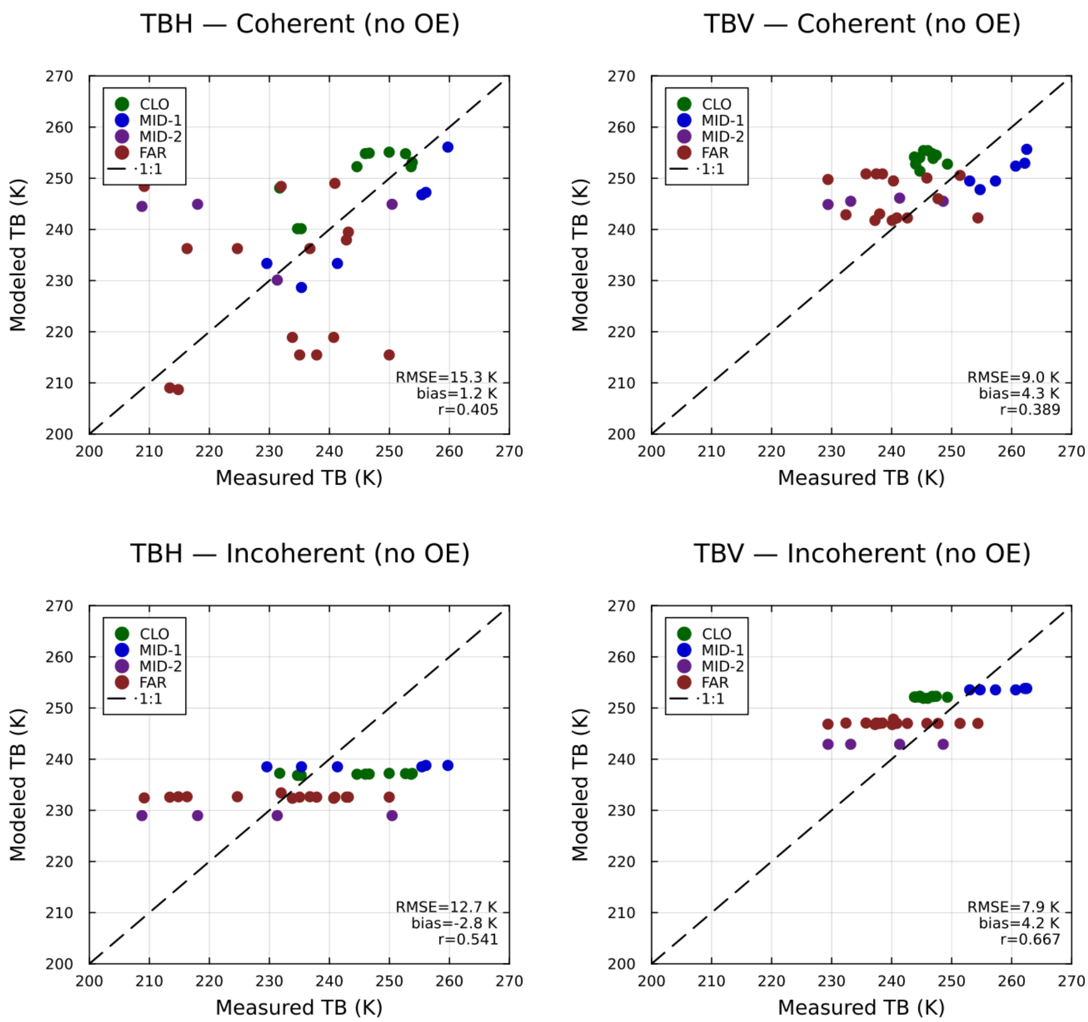

Figure R2. Scatter plots of measured versus modeled brightness temperatures for the coherent (top) and incoherent (bottom) forward models applied without Optimal Estimation, using fixed prior parameters and in situ measured snow depth and sea ice thickness. Left and right columns show H- and V-polarization, respectively. Colors denote the four measurement scenarios: CLO (green), MID-1 (blue), MID-2 (purple), and FAR (dark red). The dashed line is the 1:1 reference. Overall RMSE, bias, and Pearson correlation coefficient (r) computed across all scenarios are annotated in the bottom-right corner of each panel.

-Figure 3: For TBV the incoherent value is clearly above the average of the coherent values. This seems to directly contradicting your statement in line 148. Can this be explained by the differences in models (besides coherence?)

The statement in line 148 has been revised to avoid confusion. The intent was to highlight that the incoherent model tends to fall near the average of the coherent model's oscillations, but there is no physical constraint requiring this to hold exactly. This is because even though the used incoherent and coherent radiative transfer models can be considered here as identical in practice, they are fundamentally distinct modeling frameworks. In the coherent case, phase information is incorporated explicitly, producing interference-driven oscillations at layer boundaries. In the incoherent case, only energy is propagated, yielding a monotonic result. Any quantitative agreement between the two is therefore model-dependent, and the offset visible in TBV reflects this structural difference.

-The used uncertainties are:

> -Not always provided exactly, but sometimes as range. This is not reproducible and a reference to the code is not particular reader-friendly.
>
> We agree that reproducibility is important. The uncertainties are presented as ranges in Table 2 to improve readability and avoid an excessively large table containing individual values for every retrieval. The ranges are intended as a summary of the uncertainty estimates used throughout the study. For full reproducibility, the exact uncertainty values employed in each case are available in the publicly released code associated with the manuscript. We have clarified this point in the revised text and will consider providing additional information in the supplementary material if deemed necessary.
>
> -base on not exact criteria. What is considered 'homogeneous', what 'close', what 'few measurements'?
>
> The uncertainty assignment directly reflects the data availability and quality for each measurement. Sites and variables with more co-located samples received smaller uncertainties, while larger uncertainties were assigned where fewer or more spatially dispersed samples were available. For example, snow depth uncertainties range from 0.005 to 0.02 m depending on the number of nearby drillings, while snow density uncertainties are systematically larger (25–100 kg m^{-3}) reflecting the much sparser snow pit sampling. The full set of assigned uncertainties is documented in the publicly available code repository, allowing complete reproducibility and transparency of the optimization setup.
>
> -surprisingly small for snow thickness and ice thickness. At least for thicker snow it is common to have more than 2cm difference within measurements of the same snow-pit. Please double-check that uncertainties are consistent across parameters. For example, the difference snow-thickness uncertainties of 1 cm and 0.5 cm might seem small, but it is effectively doubling the weight of snow thickness in your optimization.
>
> The uncertainties are assigned individually for each measurement based on data quantity and quality, as described in Section 4. For snow depth, values of 0.5 - 1 cm are assigned

when multiple co-located drillings provided the local conditions. We acknowledge, however, that spatial variability within a site can exceed this range, particularly for thicker snow, and the assigned uncertainties are intended to capture the representativeness of the local sampling rather than the broader site variability. Regarding consistency across parameters, the different uncertainty levels reflect the unequal sampling across variables. Snow depth benefited from frequent drilling measurements, whereas snow density relied on the much sparser snow pit data (see Table 1), and was therefore assigned larger uncertainties. Sea ice temperature and salinity were less sampled, with only four ice cores collected within the entire campaign, resulting in fixed priors and constant uncertainties across sites. The uncertainties are thus not arbitrary, and they allow the optimization to weight each parameter according to how well it was characterized in the field.