Dear reviewer,

Thank you very much for your insightful comments on our article. We have taken care to revisit all the points you raised and to address them as precisely as possible. Please find our answers below.

Comments:

1. *My primary concern relates to the optimization strategy used to infer the preferred combinations of noise and mixing parameters. The model-observation agreement is quantified using an RMS score calculated from only two points in the power spectrum, namely the inflection point and the high-frequency limit. While this approach is computationally simple, it does not fully exploit the information contained in the observed spectra. Different parameter combinations may produce similar values at these two frequencies while exhibiting substantially different behaviour elsewhere in the spectrum. Because the inferred 8 cm mixing scale is one of the central conclusions of the manuscript, I encourage the authors to demonstrate that this result remains robust when alternative metrics based on the full spectrum, spectral slope, or integrated spectral differences are considered.*

Several reviewers noted this inconsistency for the RMS score calculation. We thus modified our method for calculating the RMS score. We now calculate the RMSD score across a wide continuous range of the spectrum, i.e. between the start of the frequency's PSD drop and the very high frequencies, stopping at twice the resolution of the widest samples to ensure that at least two points are included in the spectrum calculation.

Over this frequency range, we calculate the log-RMSD score using the following formula:

$$\log RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(OBS_i) - \log(MODEL_i))^2}$$

We also carried out around a hundred simulations for each combination, in order to derive the average spectrum and eliminate the random component in the noise implementation, as it became apparent to us that 10 cores were insufficient to eliminate random variability in the VFC spectrum due to the addition of noise.

Both improvements improved the robustness of the comparison between the model and the observations spectra. The new figures and results will be discussed in Section 4.2 of the revised version.

This new method reveals two main findings (Figure R1):

"(1) At least 7 cm of surface mixing is required to achieve RMS scores < 0.5 and thus bring the model closer to the observations. Above 6 cm, the RMSD score drops to reach its minimum values whatever the noise level. It appears that the mixing is the main parameter controlling the model convergence towards the observations.

(2) Above 50% noise, the peak observed around 18 cm in ICORDA-LDC is no longer visible. See Fig (N2) for a mixing scale of 7 cm."

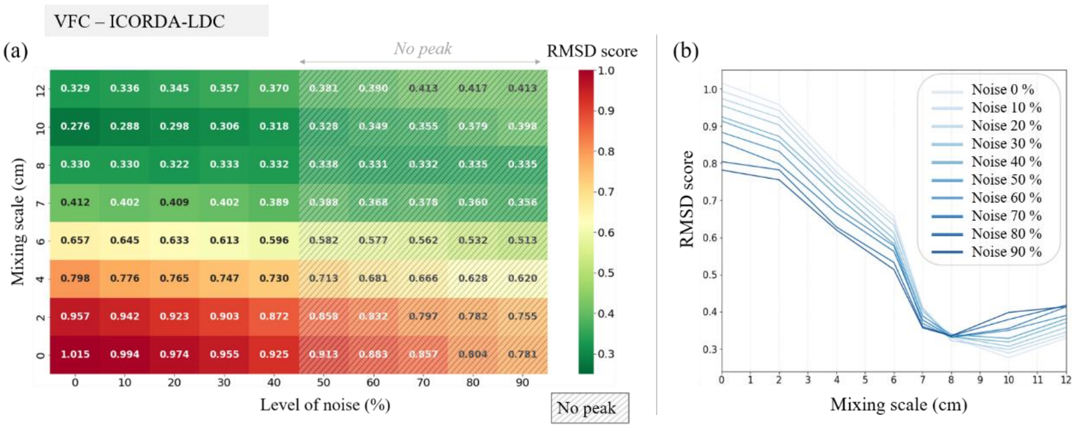

Figure R1 – (a) Model-observations RMSD score for the various sensitivity tests of the VFC model, with a noise level varying from 0 to 90 % and a mixing scale ranging from 0 to 12 cm. The grey hatched area corresponds to simulations where the frequency peak at 18-19 cm does not appear. (b) Evolution of the RMSD score along with the mixing scale for each level of noise. A low RMS score indicates a good match between the model and the observations, while a high RMS score indicates that the model-observation fit is poor.

2. *A second concern relates to the interpretation of the Subglacior-LDC record. The manuscript attributes much of the discrepancy between observations and simulations to storage diffusion occurring during the five years between drilling and analysis. While this explanation is plausible and supported by previous work, the additional storage-diffusion experiment still fails to reproduce the observed profile and spectral characteristics. The discussion subsequently suggests that even stronger storage diffusion may be required. At present, this argument remains somewhat speculative because storage diffusion is invoked to explain residual discrepancies that are not directly quantified. The authors should clarify which aspects of the mismatch can be explained by storage effects and which remain unexplained.*

Many thanks for your comment. We applied the new-RMSD-score method to Subglacior-LDC firn core, that better highlights how the storage diffusion improve VFC representation of Subglacior-LDC. We propose new results and figures (Figures R2, R3) and to modify the description of VFC/Subglacior-LDC results in Section 4.2 as:

"Taking storage diffusion into account clearly improves the RMS score (Figures [R2, R3]), with values below 0.5, although it does not perform as well as the best scores for the ICORDA-LDC reconstruction when mixing is included. It should be noted that, with the addition of this storage diffusion, the effect of mixing and noise is no longer visible in the RMS score. Moreover, the peak corresponding to a 19 cm periodicity disappears from the VFC spectrum, in agreement with the observations in Subglacior-LDC. When a firn core is stored for too long before being analysed, diffusion caused by storage erases the isotopic signal in the upper part whatever its origin (deposition or post-deposition effect)."

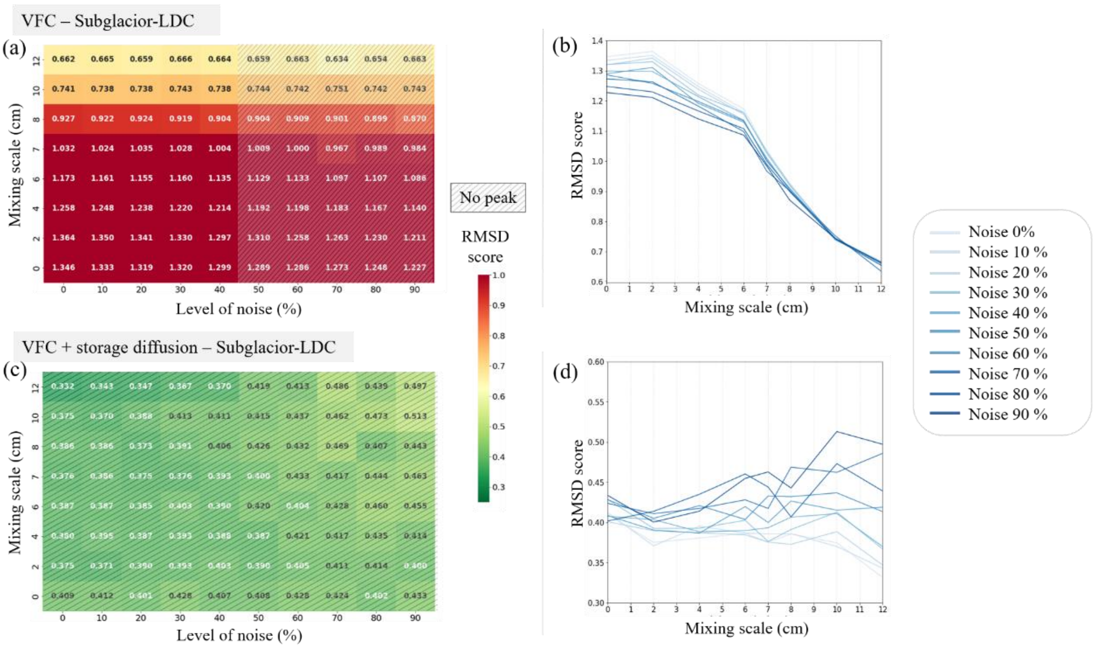

Figure R2 – Model-observations RMSD score for the various sensitivity tests of the VFC model, with a noise level varying from 0 to 90 % and a mixing scale ranging from 0 to 12 cm, (a) before and (c) after implementing the storage diffusion. The grey hatched area corresponds to simulations where the frequency peak at 18-19 cm does not appear. Evolution of the RMSD score along with the mixing scale for each level of noise, (b) before and (d) after adding the storage diffusion. A low RMS score indicates a good match between the model and the observations, while a high RMS score indicates that the model-observation fit is poor.

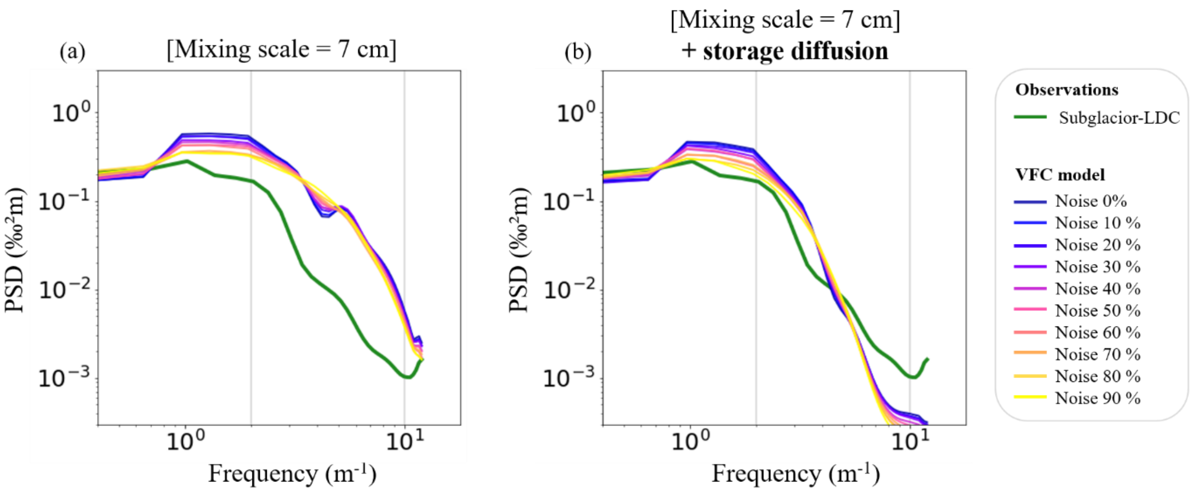

Figure R3 – Comparison of ICORDA-LDC spectrum and VFC spectra for a range of noise levels between 0 and 90%, and a mixing scale of 7 cm, (a) before and (b) after implementing the storage diffusion. The dashed line highlights the frequency peak in the observations and the grey solid lines delimit the spectral domain to evaluate the model.

"Adding the storage diffusion brings the spectrum of the model closer to that of the observations, but cuts off the high frequencies too sharply (Fig. 14). The storage effect as modelled in Dallmayr et al. (2025) cannot fully reconcile the model with the observations, but still clearly improve the RMSD score and the VFC spectrum, as observation match over the top part. We note that Dallmayr et al. (2025) estimated $\sigma_{storage}$ at a depth between 2.40 and 3.40 m, i.e., shifted slightly deeper than our 0-3 m depth interval. Because open porosity is larger close to the surface, we probably underestimate the diffusion due to storage effect in the Subglacior-LDC core.

This storage effect is minimized for the ICORDA-LDC core due to the short storage time (a few months) prior to isotopic measurements."

3. *The interpretation of the inferred surface mixing process would also benefit from further clarification. The manuscript demonstrates that introducing a mixing parameter improves agreement between simulations and observations, but the physical meaning of this parameter remains somewhat ambiguous. Throughout the discussion, the mixing term is interpreted as representing barometric pumping, firn ventilation, wind-driven snow reworking, and surface redistribution. These processes operate on different spatial and temporal scales and may not necessarily be represented by a single effective parameter. The limitations of this interpretation should therefore be discussed more explicitly.*

We agree with the reviewer that we used a simple representation for complex processes. This follows the general strategy when using the VFC approach in which post-deposition effects such as sublimation, wind redistribution and surface relief variations in local accumulation rate is classically represented by stratigraphic noise. We show here that the noise component is not enough to represent all the post-deposition processes (except diffusion which is explicitly calculated). We thus propose to add a mixing term which is easy to implement and which actually agrees with the recent study of Ollivier et al. (2025). Adding this term improves the model - observations agreement. We also show that introducing this mixing term makes sense with respect to several mechanisms already observed in the upper firn such as the snow reworking (under the wind influence) and the wind or temperature induced mixing of water vapor. We keep the philosophy of the VFC approach and do not introduce a mechanistical description which may be seen indeed as a limitation and justifies the need of more mechanistical models in complement to the simple VFC approach. We will explain better in the revised text the limitations of the simple VFC approach and the choice of the simple representation of the post-deposition processes in the noise and mixing components.

4. *Finally, I encourage the authors to moderate several statements that imply a direct physical estimate of an 8 cm mixing layer. The analysis demonstrates that an effective mixing scale of approximately 8 cm provides the best agreement within the current modelling framework. However, given the uncertainties associated with model structure, forcing data, chronology, and optimization criteria, it would be more appropriate to describe this value as a model-derived effective parameter rather than a uniquely constrained physical property of the firn.*

Many thanks for this comment which is very sound. We will modify the discussion and conclusion accordingly to clearly state the limitations of our approach with the simple model approach chosen here. This approach should ideally be combined with more process-based models.