RESPONSE TO REVIEWER 1:

We thank you for carefully reading our manuscript and giving useful comments. We revised the manuscript based on your comments. Our responses to your comments are described in the following, where your comments are italicized.

Major comments 1

In the original MGBF design (Purser et al., 2022), the filter is applied hierarchically across multiple resolutions (g_1, g_2, …, g_n), with each level contributing to the final covariance operator. This multiscale construction is central to MGBF's ability to approximate broad localization functions and capture anisotropic or spatially inhomogeneous structures. The process involves adjoint and direct filtering at each grid level (see Eq. 18 and Purser et al., MWR 2022, p. 722), and the results are additively combined (Eqs. 16–17), ensuring smoothness, self-adjointness, and scalability.

In contrast, the present manuscript adopts a significant simplification: filtering is applied only at the coarsest filter grid, with no filtering at finer levels. This is a clear deviation from the original formulation, and although the authors mention it is for computational efficiency (Lines 99 and 306), the implications of this choice are not adequately discussed. Specifically, the manuscript should examine:

- *How this approximation affects the effective shape of the localization function, especially for short localization length scales (e.g., 20 km);*
- *Whether it risks degraded performance (e.g., loss of sharpness or spurious correlations) in such cases;*
- *Whether the approximation is acceptable only in certain regimes, such as large-scale SDL with long localization radii, or whether it generalizes more broadly.*

Clarifying these points would help readers understand the trade-offs and limitations of this modified implementation.

Response:

In MGBF-based localization, the interval of the filter grid should be smaller than the localization length. Therefore, the computational cost of the analysis with small localization length in MGBF is not necessarily smaller than that in RF. This limitation is both in single-scale localization and SDL. We will explain this trade-off in the revised manuscript. The effect of the coarser filter grid with large localization length is negligible, which is shown as difference between MGBF00 (the filter grid is the same as the analysis grid) and MGBF04 (the filter grid is coarser than the analysis grid) in Fig. 4.

Minor comments 1

Line 55, The current title of Section 2.1, "Ensemble-variational (EnVar) data assimilation", does not reflect the fact that this subsection includes a detailed mathematical formulation of scale-dependent localization (SDL) as applied in the GSI-based 3DEnVar system. In particular, Eqs. (3) and (4) describe the *decomposition of ensemble perturbations across multiple spatial scales and the corresponding block-structured localization matrix.*

Since SDL is a significant methodological feature of the paper, both in terms of formulation and in experimental comparisons (e.g., RFSDL vs. MGBF04SDL), I recommend updating the subsection title to something more precise, such as "2.1 Ensemble-variational (EnVar) data assimilation with scale-dependent localization".

Response:

We will add "with scale-dependent localization (SDL)" in the title of this subsection.

Minor comments 2

In Line 105, the manuscript states that interpolations are performed "from g_1 to the analysis grid g_0." Since g_1 is referred to as the "finest filter grid," it may be misinterpreted as having equal or higher resolution than g_0. However, based on Table 2, g_1 can in fact be coarser than the analysis grid (e.g., in MGBF03–04). I suggest the authors clarify the resolution relationship between g_0 and g_1 to avoid potential confusion.

Response:

As you pointed out, the finest filter grid g_1 can be coarser than the analysis grid g_0. We will explain it in the revised manuscript.

Minor comments 3

In Line 137, the authors mention that the analysis grid resolution is twice as coarse as the FV3LAM model grid (i.e., 6 km vs. 3 km), but do not provide any justification or discussion of this design choice. Since this resolution difference could affect the representativeness or accuracy of of the ensemble background error representation, localization, and filter application (especially given the role of multigrid interpolation in MGBF), it would be helpful if the authors could clarify:

- *The rationale for using a coarser analysis grid (e.g., computational efficiency, memory constraints, etc.),*
- *Whether this design introduces any limitations or trade-offs in terms of representativeness or localization sharpness,*
- *And whether the MGBF design is sensitive to the resolution mismatch between the filter grid and the model grid.*

Response:

We adopted the coarser analysis grid to reduce the computation cost. As the limitation, it makes the resolution of analysis increments coarser and prevents to set the smaller localization length than the grid interval. We will explain it in the revised manuscript. The interval of the analysis grid is sensitive to the optimal localization length, but it is equally sensitive in both RF and MGBF because the resolution of analysis increments does not depend on the filter grid.

Minor comments 4

Table 2: The symbol "–" appears in several columns (e.g., "Number of the finest filter grids", "Weight of (g_1, g_2, g_3, g_4)", filter specifications), but its exact meaning is not defined. It is unclear whether "–" indicates "not applicable," "not used," "same as previous case," or "no filtering applied." To improve clarity and reproducibility, I suggest the authors include a footnote or caption line in Table 2 to explicitly define what "–" represents in each context.

Response:

It indicates no filtering. We will add the explanation in this table.

Minor comments 5

Lines 248‑250 and elsewhere: The sentence beginning with "Nevertheless, the difference from RF…" is grammatically correct, but a bit hard to follow due to its length and repeated comparative structure. With multiple experiments and color-coded references mentioned together, the logical comparison becomes difficult to parse.

I suggest breaking it into two simpler sentences or rephrasing it for clarity. For example:

"MGBF04σ showed a smaller deviation from RF than MGBF04. Similarly, MGBF04σSDL was closer to RFSDL than MGBF04SDL."

In fact, similar long and repetitive sentence constructions appear in several other places in the manuscript. I recommend that the authors go through the manuscript to revise such sentences for improved readability and flow.

Response:

Thank you for the suggestion to make the sentence clearer. We will revise it as your suggestion.

Minor comments 6

Figure 9: there seems to be a mismatch between the panel labels and their descriptions in the caption. Based on the plotted content, panels (a) and (b) appear to show RMSE and bias for temperature, while (c) and (d) show RMSE and bias for horizontal wind. However, the caption currently states that (a, c) are temperature and (b, d) are wind, which appears to be incorrect.

Response:

Thank you for pointing out the mismatch between the labels and the caption. We will correct it.

Minor comments 7

Lines 305-306: The sentence "… and showed how to prevent the computational problem found in applying it" reads a bit awkwardly. The phrase "prevent the computational problem" is not the best fit here, since the issue already occurred during implementation.

Response:

We changed this phrase to "make the computation faster than the RF."