[For Topical Editor] We appreciate the topical editor's comments on the manuscript. Please find our response to each of your comments below.

- *I agree with the reviewers' suggestion about elaborating on the implementation and use of BUMP, which is an essential contribution to this work.*

We have made additional revisions as described in our responses to reviewer 1, comment 3, and reviewer 2, comments a and b.

- *The authors should briefly explain or show a figure to quantitatively discuss how tuning the horizontal lengths of \psi and \chi by half changes the velocity variance.*

We have changed lines **246-252** to clarify how the implied velocity variance changes with changes in the $\delta\psi$ and $\delta\chi_u$ correlation lengths. (Since the velocity variance is proportional to the second derivative at the origin of the $\delta\psi$ and $\delta\chi_u$ correlation function, the velocity variance is inversely proportional to the square of the correlation length.)

Previous text:
Since the implied velocity variance depends on the second derivative at the origin of the $\delta\psi$ (and $\delta\chi$) correlation (Lorenc, 1981; Daley, 1985), the diagnosed covariances greatly underestimate the velocity variance relative to the statistics of the of original training data. Reducing the horizontal correlation length for $\delta\psi$ and $\delta\chi_u$ increases the velocity variances, though at the expense of further underestimating the correlations at larger separations.

New text:
The implied velocity variance in the modeled covariance (Eq. 2) is proportional to the second derivative at the origin of the $\delta\psi$ (and $\delta\chi$) correlation (Lorenc, 1981; Daley, 1985). That is, $\delta\psi$ correlations that are more strongly peaked at the origin will produce larger velocity variance even if the $\delta\psi$ variance is fixed. Thus, the modeled covariances greatly underestimate the velocity variance relative to the statistics of the original training data. Reducing the horizontal correlation length for $\delta\psi$ and $\delta\chi_u$ by a factor of two increases the second derivative of their correlation, and therefore the velocity variance, by a factor of 4, leading to a better fit to the velocity variance in the training data.

- *Also, please address the issue of the poorer performance of the 3DVar Qv analysis compared to 3DEnVar (Fig. 13). This may explain why using sophisticated moisture variables can help.*

Thank you for your suggestions. We have added the following sentences in lines 327-328 (section 5.2):
"**It is notable that the larger Qv RMSE for 3DVar lasts until 6 day forecasts (Fig. 14f). This might be because the moisture variable is univariate in current B design (section 3.1) together with relatively less observation amount for moisture.**" . Two literatures for moisture variables are mentioned in section 6, so they are not mentioned here.

[For reviewer 1] We appreciate the reviewer's comments on the manuscript. Please find our response to each of your comments below.

1. *Section 2.3: I very much appreciate the author's detailed responses to my previous comments regarding the background of this new incremental approach for updating p, roh_d, and theta_d. Nevertheless, I think revisions should be extended to the second paragraph as well. It is important to point out that the 3-dimensional pressure p is a prognostic variable in MPAS but not being used as an analysis variable in JEDI here in this manuscript. As such, after each data assimilation analysis, pressure is "re-diagnosed" from integrating T, q, and p_s from surface to upper levels via hydrostatic balance, resulting in a pressure that is different from the background forecast (i.e., a non-hydrostatic pressure). This discretization errors can exist even if there is no analysis increment from DA. The way the second paragraph is written does not state this very clearly, although it was very clearly explained in the author's response and should be included in the manuscript.*
   However, after reading the author's responses, I have another question. If there is no analysis increment from data assimilation, why not just use the background file as the analysis file, why is there a need to propose a specific treatment to this zero-increment situation? Or perhaps when you say "increment from DA is zero" you meant zero increment for ps only while increments for other analysis variables are non-zero, as such, you can't just replace analysis file with background file.

We have revised both paragraphs, considering the reviewer's comments from both rounds of reviews. We now emphasize the central point that Liu et al. (2022) enforced hydrostatic balance on the full, analyzed fields, while the new approach used in this paper applies hydrostatic balance only to increments from the background fields. We have also omitted some confusing details, such as the role of pressure as an intermediary variable in the transformations (note that pressure is not a prognostic variable in MPAS), and the fact that the incremental approach avoids any changes to the background state in regions where there are no observations.

Please see the revised paragraphs at lines 93-102:
"**In Liu et al. (2022), ρ_d and θ_d are computed from the analyzed T, p_s, and q by assuming hydrostatic balance. Here, we instead compute increments for ρ_d, and θ_d (i.e., $\delta\rho_d$, and $\delta\theta_d$) from the increments δT, δp_s, and δq. This approach, which is implemented by linearizing the corresponding calculations of Liu et al. (2022, steps 3 and 4 of their section 3.3), assumes hydrostatic balance only for the increments and not the full, analyzed fields.**

Assuming hydrostatic balance just for the increments is preferable because that balance is only approximate and, moreover, the discretized form of hydrostatic balance used in the variable transformation is not precisely equivalent to that implied by the discrete MPAS equations. Since the hydrostatic integral is computed from the surface upward, differences between the incremental and full-fields formulations can be expected to accumulate with height. Consistent with this, JEDI-MPAS cycling experiments (not shown) using the new, incremental update for ρ_d and θ_d exhibit reduced temperature bias in the stratosphere, especially near the model top." .

2. *Section 3.2: Regarding the training dataset, the author stated in their response that they actually have used the MPAS model's own forecast samples to diagnose the B parameters and compared with those from the NCEP GFS forecast samples and found overall similarity except for the error standard deviation having larger differences. Furthermore, the resulting one-month cycling experiments shown reduced temperature and wind RMSE in the upper levels for the 6-h forecasts when MPAS model's own forecast samples were used to diagnose B parameters. I think this is worth mentioning in the revised manuscript even though it is from a recent version of JEDI-MPAS source code that is different from the one used here.*

Thank you for the reviewer's suggestion. We have added the following statement to the end of the first paragraph in section 3.2 (lines 166-170).
"**With a recent (early June 2023) version of JEDI-MPAS source code after initial submission of this paper, we have trained the static B parameters from MPAS model's own forecast with the same methodology described here. The overall B statistics diagnosed from MPAS-based samples were similar to that from GFS-based samples reported here, except for the error standard deviations in the stratosphere, which were larger for MPAS-based samples. In the one-month cycling experiment, this led to a reduction in temperature and wind RMSEs in 6 hour forecasts in the stratosphere.**" .

3. *I agree with the other reviewer's comment that this manuscript can be at a higher level by considering to address the BUMP implementation in more details, especially from the algorithmic perspective. The authors have taken the suggestions and provided more details in sections 3.1 and 3.2 of the revised manuscript, which is very nice. I think this revised manuscript can be further elevated by including a flowchart or diagram to illustrate the BUMP's implementation, highlighting the correspondence between each of the operator in equation (2) with its associated BUMP driver(s). For example, C, the block-diagonal correlation matrix in equation (2) involves the use of NICAS and HDIAG, while Σ, the diagonal matrix of standard deviations involves the use of VAR and NICAS. Including a flowchart or diagram will also make this manuscript more educational and attract more readers.*

Thank you for your suggestion. We have added a simple diagram to represent the operation shown in equation (2) as figure 1. The following text was added at the end of section 3.1 (lines 158-159) :
"**Figure 1 shows a diagram for Eq. 2 with corresponding BUMP drivers and MPAS-specific linear variable change.**" .

[For reviewer 2] We appreciate the reviewer's comments on the manuscript. Please find our response to each of your comments below.

a) *At the third step of HDIAG, a horizontal average is performed over all mesh nodes. I wonder if it is possible to average parts of mesh nodes to achieve spatially varying statistics.*

Indeed, HDIAG can generate the *local* statistics, but we missed this in the steps of HDIAG. We have changed lines 195-202 as follows:

"The third step is a horizontal averaging of these raw correlations**, either over all the mesh nodes or over local neighborhoods**. The average is binned depending on the level and the horizontal separation for the horizontal correlation, and depending on the concerned levels for the vertical correlation. As a final step, HDIAG fits a Gaspari and Cohn (1999) function for each averaged correlation curve. Thus, we obtain horizontal and vertical correlation length-scale values for each level. **If the averaging and curve fitting steps are performed over local neighborhoods, an extra interpolation step is necessary to obtain 3D fields of length-scales on the model grid.** These **length-scale profiles or 3D fields** can be stored and provided to NICAS in order to model the spatial correlation operator. **In this study, the local correlation lengths were obtained from raw statistics within 3000 km radius for a given diagnostic point.**" .

b) *How is the cost of HDIAG compared to the preexisting methods in calculating these correlation statistics?*

Any method needs to diagnose some statistics (sample correlation) from samples. BUMP HDIAG does this on the subsampled grid (which is beneficial in a computational aspect), then interpolate those back to the full grid (this requires an additional cost, but marginal compared to benefit from subsampling).