Review of "*Assessment of the data assimilation framework for the Rapid Refresh Forecast System v0.1 and impacts on forecasts of a convective storm case study*"

The authors use a relatively new modeling system to produce a set of forecasts of a convective-scale event. Sensitivity experiments are performed to assess the impact of different data assimilation choices with standard verification metrics. The authors do a nice job of explaining the new system, and providing justification for their choices, as well as incorporating a diversity of verification approaches. My major concerns with the study are the overlap with prior work (e.g., Tong et al. (2020) used similar model and DA systems) and the lack of additional cases for analysis. I've provided some minor comments below that I believe need to be addressed before recommending acceptance.

We thank the reviewer for the comments and suggestions, contributions which improve the quality of the final version of the manuscript. Our responses are noted below. Changes to the document are highlighted in blue.

Indeed this study has some overlap with prior work such as Tong et al. (2020). As pointed out by the reviewer, the main overlapping is regarding the use of the limited area model (LAM) capability based on the Finite-Volume Cubed-Sphere (FV3) dynamical core (FV3LAM) and the Gridpoint Statistical Interpolation (GSI) analysis system. However, it should be noted that, up to the moment this manuscript was submitted, the only study in which both of these systems were used is Tong et al. (2020). Therefore, the present study intends to fill part of the gap in the literature by providing a description of the initial data assimilation infrastructure and assessing the analyses and forecasts produced with the prototype RRFS system. Some of the functionalities and configurations tested, such as the planetary boundary layer (PBL) pseudo-observations function of GSI, the localization scale radius, and cycling strategy, are based on currently operational Rapid Refresh (RAP) and High Resolution Rapid Refresh (HRRR) modeling systems. Thus, this study informs developers of a set of configurations that lead to better convection forecasts and points out aspects that require more tuning and improvements inside the prototype system. We are aware that additional case studies covering a wider variety of convection modes and time frames are needed in order to obtain more robust results, which is the subject of future studies. However, and as mentioned in the document, we intended to explore more functionalities and configurations on a single case in order to establish an understanding of baseline sensitivities, which can help future RRFS implementations.

Minor comments

- **I suggest the authors provide more detail about how this study differs compared to Tong et al. (2020), especially the model configuration and design choices. There are a lot of similarities, including the use of FV3 (termed the FV3-SAR in that study, which I believe is the same model that is the core component of the UFS-SRW), variational and hybrid DA with GSI, similar physics choices, and a similar convective-storm case study approach (although for a different case). Can the authors describe how that work ties in with the current set of experiments?**

- As mentioned by the reviewer, the main overlapping is regarding the use of the FV3LAM and the GSI analysis system. Other similarities that could be mentioned are the model horizontal grid spacing of 3 km; numerical experiments executed assessing the hybrid weight of the ensemble background error covariance (BEC) with values of 0% (pure 3DVar), 75% and 100% (pure ensemble based BEC); similar microphysics parameterization (Thompson Aerosol-Aware; Thompson and Eidhammer (2014)); and similar convective case study. Nevertheless, our study greatly differs from Tong et al. (2020) in two main aspects: the direct assimilation of reflectivity data and the testing of the ensemble Kalman filter (EnKF) data assimilation method. In our study we assimilate radar radial velocity and the vertical azimuth display derived from radar radial velocity, but, we do not assimilate any reflectivity data. Meanwhile, we test functionalities in GSI such as the PBL pseudo-observations function and supersaturation removal that were not considered in Tong et al. (2020). Additionally, we focus on the hybrid 3DEnVar method in GSI since it is used in operational RAP and HRRR systems. For operational RRFS, development is underway to incorporate the EnKF into the hybrid data assimilation system for the first implementation. In order to account for the reviewer's comment, lines 110-113 were added to the manuscript and lines 90-93 were re-phrased.

Lines 90-93: "which studied the impact of the direct assimilation of radar radial velocity and reflectivity using the hybrid three dimensional ensemble–variational data assimilation (3DEnVar) and ensemble Kalman filter (EnKF) algorithms within the Gridpoint Statistical Interpolation (GSI; e.g., Wu et al., 2002; Kleist et al., 2009)."

Lines 110-113: "It is worth mentioning that despite some similarities with the work of Tong et al. (2020), in this study the focus is on the hybrid 3DEnVar method in GSI and configurations used in operational RAP and HRRR systems. For the operational RRFS, development is underway to incorporate the EnKF into the hybrid data assimilation system for its first implementation."

- **The RRFS will be an ensemble-based system, so generating and verifying ensembles seems like a good choice to assess the benefits of the various approaches. It may be useful to clarify why the authors only performed deterministic forecasts somewhere in the text (sorry if it's there and I missed it!).**

- The reviewer is right. Since RRFS is under development, RRFS v0.1 does not yet have the capability of generating ensemble forecasts. Lines 102-103 were added, as suggested.

  Lines 102-103: "While single, deterministic forecasts are produced and evaluated in this study using RRFS v0.1, it should be noted that future RRFS implementations will produce convection-allowing ensemble forecasts."

- **I really think this study would benefit from additional cases, especially when the authors argue at many points in the paper they are using the results to guide future configuration decisions. Some of the differences between the experiments seem very small, and may become more evident with a larger sample size. This could be considered a "fatal flaw" by some, but I think there's**

some merit in providing documentation of ongoing work leading up to the implementation of the future RRFS system in the form of this manuscript.

- We agree with the reviewer. However, and as mentioned in the document, the intention of this study is to explore a variety of functionalities and configurations on a single case in order to establish an understanding of baseline sensitivities. Upon establishment of initial baselines, retrospective cases and real-time experiments spanning weeks and seasons will be examined as RRFS advances toward operational implementation of RRFS.

Specific comments

Lines 40-42: This sentence implies that the addition of the WaveWatch model into the operational forecast somehow improved the low-level cold temperature bias observed in a prior version of the GFS. I don't think that's possible and I don't think the change notice referenced supports that claim. Please revise.

- The reviewer is right. The sentence was revised as suggested.

  Lines 40-41: "Within the UFS framework, the GFS was coupled with the WAVEWATCH III wave model in the operational upgrade of March 2021 (NWS, 2021)"

Line 87-91: I recommend removing these sentences. The number of studies that examine data assimilation for convection-allowing applications is too numerous to mention here, so describing these two specific studies is necessary, unless they are aspects of the work that are especially relevant to the current work.

- The sentences were removed as recommended.

Line 104: The authors should make clear that the eventual RRFS implementation will produce ensemble forecasts and not just a single deterministic forecast.

- Lines 103-104 were added in order to clarify the point made by the reviewer.

  Lines 103-104: "Single deterministic forecasts are produced and evaluated in this study using RRFS v0.1, however, it should be noted that future RRFS implementation will produce convection-allowing ensemble forecasts."

Line 153-161: I suggest moving the list of these parameterizations into a Table that can be referenced in the future, including names of schemes and associated studies that describe each scheme.

- As suggested, the list of parameterizations was moved to Table 1. The text in lines 151-153 was adjusted accordingly.

  Lines 151-153: "The RRFS_PHYv1a suite is based on physical schemes implemented in the operational RAP, HRRR, and GFS systems and is used in all simulations in this study. Table 1 presents the RRFS_PHYv1a physics parameterizations and associated studies that describe each scheme, based on CCPP (2021)."

Section 2.6: What do the authors mean by “workflow”? As written, the term is used rather generically, but I’m guessing that there is specific workflow software that is used that should be described in the text (this may be described later in the text, but the authors should bring this up earlier).

- As recommended, the specific workflow software used was referenced earlier in the text in lines 205-207.

  Lines 205-207: “. It is based on the UFS SRW application v1.0.0 (UFS Development Team, 2021) community workflow which uses the Rocoto workflow management system (https://github.com/christopherwharrop/rocoto/wiki/Documentation).”

Line 260-262: How were the MLCAPE and shear diagnostics computed? The text states they were “observed”, but there are no routine soundings typically available between 19-20 UTC in northeastern Oklahoma. If these values are from a model analysis, that should be stated (e.g., “The RAP analysis contained MLCAPE values of…).

- The reviewer is right. However, for this case study in particular, a sounding at 19 UTC is available in northeastern Oklahoma. The sounding with all the instability parameters can be accessed at https://www.spc.noaa.gov/exper/archive/event.php?date=20200504. Lines 255-256 were added to the document to clarify the source of the parameters.

  Lines 255-256: “The instability parameters are based on the observed sounding at 19:00 UTC over Norman, OKlahoma (KOUN).”

Line 367: Are Oklahoma Mesonet observations assimilated?

- Yes, as can be noted in Figure 4 by the + symbol.

Line 543: Is model level 50 really located around 850 mb? Does that mean that there are 50 levels below 850 mb and only 14 levels above 850 mb (64 levels total)?

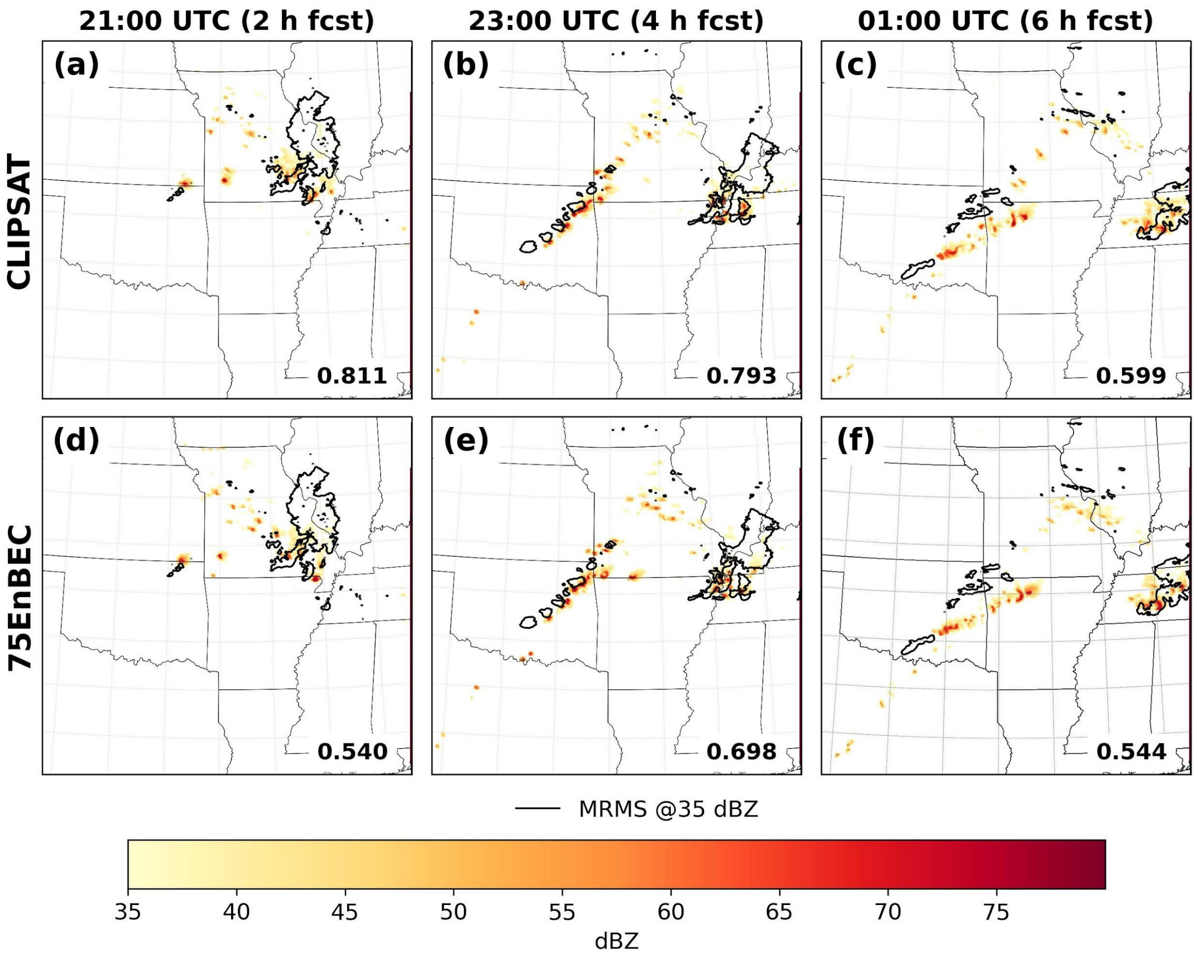
- We thank the reviewer for pointing this out. In FV3LAM, the hybrid model levels are inverted from 0 to 64, thus, model level 64 is close to the surface and level 0 is at the model top. The hybrid model level 50 in this study corresponds then to level 15 which is located around 850 mb. Figure 15 and its caption as well as line 538 were modified accordingly. In addition, the vertical resolution of the model was modified since 65 vertical layers were used instead of 64 (line 267).

Lines 554-556: How are the MMI values so different at 21 UTC in CLIPSAT and 75EnBEC? To my eye, the figures look almost identical. The differences at 23 UTC look more significant, but the MMI values are more similar at this time. Can the authors explain why this is the case?

- We agree with the reviewer that in a visual and more subjective analysis the subplots for 21 UTC from both experiments look very similar (Figure 16a and d in the new version of the manuscript). In fact, that motivated the inclusion of the objective analysis in the figures and results discussions showing the MMI (F+O) results for reflectivity values larger than 35 dBZ. In order to rule out any error in the results shown in this figure, the Method for Object-Based Diagnostic Evaluation (MODE) was re-applied but results were the same as shown in the manuscript. The figure below shows Figure 16 but only for forecasts of composite reflectivity above 35 dBZ with the corresponding contour of observed reflectivity at 35 dBZ. It can be noted that the experiment that produces more isolated objects which do not match with observed objects are penalized in the MMI values. This justifies the lower MMI (F+O) value presented in Figure 16d for the 75EnBEC experiment at 21 UTC. Results from the CLIPSAT experiment shows that more predicted reflectivity areas match better the observed one and it is reflected in the larger MMI (F+O) value obtained. At 23 UTC, it can be observed that the MMI (F+O) value of the 75EnBEC experiment has greatly increased when compared to the MMI (F+O) value at 21 UTC. Yet, the experiment CLIPSAT shows a better reflectivity forecast at this hour with a larger MMI (F+O) value. In CLIPSAT there is a reduction of the spurious convection between north-central Arkansas and south-central Missouri, but most of the spurious convection shown in 75EnBEC over other regions is also observed in CLIPSAT. This may be the reason why the differences are not very significant in the MMI values. To account for the reviewer’s comment, lines 550-557 were added to the document and lines 626-628 were revised and modified.

Lines 550-557: “These values are greatly increased at 2 h forecast from 0.540 in 75EnBEC to 0.811 in CLIPSAT due to a better positioning and coverage of the reflectivity above 35 dBZ in areas over southeastern Missouri. Overall, more spurious convection over Missouri is shown in 75EnBEC which may have led to the lower MMI (F+O) at this forecast hour. At 4 h forecast, MMI (F+O) results show an increase from 0.698 in 75EnBEC to 0.793 in CLIPSAT, with a reduction of the spurious convection between north-central Arkansas and south-central Missouri in CLIPSAT. Nevertheless, most of the spurious convection shown in 75EnBEC over other regions is also observed in CLIPSAT, which may have penalized the MMI (F+O) values in the last experiment. At 6 h forecast, more similar MMI (F+O) values are found, but still less spurious convection is observed for lower reflectivity thresholds in CLIPSAT”

Lines 626-628: “At shorter forecast lead hours, it produces more skillful forecasts with a better positioning and coverage of the reflectivity above 35 dBZ and precipitation forecasts are as good as in experiments 75EnBEC and 100EnBEC.”



2, 4, and 6 h forecasts of composite reflectivity from experiments CLIPSAT (a, b, and c) and 75EnBEC (d, e, and f) initialized at 19:00 UTC on 4 May 2020. Solid black lines are the 35 dBZ reflectivity observation contours, valid at the forecast time, respectively. MMI (F+O) results for reflectivity values larger than 35 dBZ are shown in the lower right corner of each panel.

Figure A1: Why is this included as figure A1 and not Figure 11?

- As indicated, Figure A1 was removed from the Appendix and included as Figure 11 in the new version of the manuscript (see lines 489, 491, and 492). The figures below were renumbered accordingly.