Note to anonymous referee #1,

Please note that we have decided to withdraw the current version of the manuscript from the peer-review process and to submit a revised version at a later date. See the response to the editor for more details.

Response to ‘comment on gmd-2022-269’ by anonymous referee #1

We would like to thank the reviewer for their interesting and relevant comments. We have addressed their remarks below, on a comment-by-comment basis. We have also indicated in the text how we intend to modify or integrate the manuscript accordingly.

Referee #1, comment #1: This manuscript aims to show the ability of the DA experiments using modRSW to imitate some behaviors of an operational NWP so that one can utilize such inexpensive configurations to study DA in an operational-like environment. The topic is attractive, especially for researchers not in an operational center and who cannot run a real operational NWP due to computational resource limitations. This work is valuable, but the experiments are not well-designed concerning imitating the convective-scale cycling DA. My recommendation is a major revision before publication.

1. The authors perform many experiments and give the experimental configurations relevant to the operational environment in terms of criteria such as grid spacing, ensemble size, localization, inflation, and RMSE. But what are the authors' suggestions on the improvement of localization and covariance inflation? If a new localization scheme or inflation scheme produces good results but cannot satisfy the criteria listed in this manuscript, should they be excluded from the operational NWP development? Showing the ability to imitate an operational-like environment is a good start, but further suggestions are also necessary. At least, the authors should tell readers what criteria should not be violated and what criteria can be adjusted when readers try to improve the DA performance in the simulated operational environment. Is a larger OID (>60%) not acceptable for imitating the convective-scale DA?

Response to comment #1: As the referee has correctly stated, the ability to imitate an operational system is an important first step in evaluating the relevance of an idealised model for NWP research. Our paper is primarily interested in demonstrating the credibility of the modRSW model as a tool to investigate convective-scale NWP data assimilation, rather than proposing specific suggestions for the implementation of an Ensemble Kalman filter. Therefore, we don't think that further suggestions are necessary on this matter. For the sake of clarity, we will add some text in the introduction to make this clearer.

Regarding the violation and the adjustment of the tuning and validation criteria mentioned in the paper, the two validation criteria are, for example, somewhat flexible. For OID, the aim is to avoid having a very low (minimal use of observations) or very high (minimal use of the previous forecast) OID. Error doubling times are also somewhat subjective, as they depend on the model resolution and the number and spatial and temporal resolution of observations. In this case, our suggestion is that a doubling time smaller than 12 hours is credible for convective-scale systems. The tuning criteria, on the other hand, should be satisfied by at least some of the experiments. For instance, if no experiment has spread close to the RMSE of the ensemble mean, then the observing setup should be rejected. We will revisit section 2 of the manuscript to clarify these points further.

R1, c2: 2. As far as I know, most operation centers use the variational DA algorithm or the variational-based hybrid algorithm. Only a few centers use the pure EnKF algorithm. Giving a reason for choosing the pure EnKF algorithm to imitate the operational environment is necessary. I do not ask for conducting experiments with a variational DA algorithm, but a brief discussion on the selection should be helpful.

Response to comment #2: The referee is right in saying that most operational centres running NWP models use a variational data assimilation algorithm rather than a pure Ensemble Kalman filter. However, the Local Ensemble Transform Kalman filter (LETKF), which is close in concept to an EnKF, is common and popular in the data assimilation community and used operationally by DWD in its KENDA system, for example (see Schraff et al., 2016). In addition, the DA configuration used in our work reflects in some respects the one used at the Met Office for the MOGREPS-G system, as discussed in the recently accepted paper by Inverarity et al. (2023). Besides, the Deterministic Ensemble Kalman filter that we have used in our study is particularly useful because it allows us to implement model-space localisation, which is typical of variational DA systems, together with parallel observation processing.
All in all, we want to stress again that our work is only intended to demonstrate the viability of the modRSW model for data assimilation in general rather than proposing a specific algorithm. We will expand slightly section 3.3 to better clarify these points.

R1, c3: 3. With respect to the convective-scale DA, the model and DA configurations are not so representative.
(1) The precipitation procedure in modRSW is more like a cumulus parameterization scheme that estimates the precipitation according to the large-scale thermodynamic environment (see the high correlation between r and h in Figure 6), while a feature of convective-scale NWP is using a microphysics scheme that explicitly simulates the physical procedures in the cloud. The difference in complexity is a gap between a convective-scale NWP and a synoptic-scale NWP. Heavy rain may occur with no large-scale forcing. So I think referring the DA experiments using modRSW as "convective-scale" DA is not proper.

Response to comment #3: We agree with the reviewer that the convection and precipitation scheme used in our model is a simple one, and we are aware that it does not include all the features of convective-scale processes. However, adding microphysics would increase the level of complexity of the modRSW model, defeating the purpose of having a simple and inexpensive model. Moreover, the cumulus-convection model developed in Wursch and Craig (2014) – for which the modRSW model constitutes an improvement – was later used in a number of data assimilation studies, all mentioning convective-scale applications (e.g. Ruckstuhl & Janjić, 2018).

R1, c4: (2) The observation density is too sparse for the convective-scale DA, especially in the case of assimilating radar and satellite data. The resolution of radar data is often 1 or 2 km. The lack of high-resolution observations is a flaw in the experiments aiming to imitate the convective-scale DA. Using multiscale observations is also a feature of convective scale DA and is not considered or discussed in the manuscript.

Response to comment #4: The choice made in this paper was to assimilate only conventional ground observations, for which a spatial density of 50 km is realistic. We are aware that convective-scale data assimilation involves the assimilation of other types of observations, such as satellites and radar reflectivity, however the simplicity of the modRSW model means that some observation types are harder to imitate than others. In this regard, some of the authors have already worked on an isentropic version of the modRSW model, the ismodRSW model, which is better suited for satellite data assimilation experiments. The ismodRSW model (together with its derivation) is discussed in two already published papers (Cantarello et al, 2022; Bokhove et al, 2022) and extensive idealised satellite data assimilation experiments are presented in Cantarello's PhD thesis (2021).
On the reviewer's suggestion, we have modified the observing system in order to imitate the assimilation of Doppler radial wind observations. We will run new experiments in which two bands of u observations are assimilated, alongside the usual h and r observations. We will update the new version of the manuscript with the new results and we will make reference to the above-mentioned papers and PhD thesis regarding the problems of imitating satellite observations using the modRSW model.

R1, c5: (3) The precipitation r in the manuscript is more like a simultaneous quantity, e.g., the precipitation rate in a time step. The accumulated precipitation, 3-h or 6-h, is used in real data assimilation. In this respect, the DA configuration in the manuscript does not imitate the real scenarios. This situation should be stated.

Response to comment #5: We will state this difference more clearly in the text. We will also note that the rain variable in the modRSW model is defined as a 'mass fraction' and therefore the comparison with accumulated precipitation over time is not straightforward.

R1, c6: (4) In convective-scale DA, we must face that many model variables are not directly observed. This issue implies that some model variables must be updated through crossvariable covariance. It is better to show results without r and h observations.

Response to comment #6: A set of 'data-denial' experiments, in which each subset of observations was excluded from the assimilation, had already been run in the past, although the results did not end up being included in the first version of the manuscript. Starting from the same DEnKF configuration considered in the paper (same number of ensemble members, same background error and filter parameters), and by excluding one set of observations at a time, our main finding is that assimilating the horizontal velocity is relatively more important to the system than the assimilation of h or r observations. In fact, both the spread and the error grow significantly in a scenario where u is not assimilated, while the overall observation impact decreases slightly. On the other hand, excluding h or r from the assimilation has a less dramatic impact on the performance of the DEnKF, probably because of the strong cross-correlation between these two variables. We will comment on this without providing supporting evidence.

R1, c7: (5) The operators of remote observations are often nonlinear, such as those of radar reflectivity and satellite observations. The inaccurate operator is also an issue in the convective-scale DA, but this issue is not discussed.

Response to comment #7: The reviewer is correct in saying that observation operators are often nonlinear in operational DA systems. A simpler choice has been made for this paper, also motivated by a stronger focus on the illustrative rather than the representative nature of the DA setup chosen for our experiments. However, as mentioned in the response to comment #4, a nonlinear observation operator for idealised satellite observations has been developed in Cantarello's PhD thesis (2021), in which an upgraded version of the modRSW model (the ismodRSW model) is used.

R1, c8: In general, the DA configurations in the manuscript are more suitable for a synoptic-scale DA study; many issues that the convective-scale DA has to face are not discussed. Since it is an idealized study, doing DA experiments with multiscale observations and without r and h observations should not be difficult. Observing the precipitation area with high resolution u observations should result in a much smaller RMSE, similar to radar data assimilation.

Response to comment #8: We believe our responses above are enough to justify the use of the modRSW model for 'convective-scale DA' and we hope we were able to address most of the reviewer's concerns. We will modify the manuscript accordingly and thank the reviewer for suggesting the use of high-resolution u observations.

R1, c9: L320-323: With respect to imitating LFC with Hc, I have some reservations.

Response to comment #9: We acknowledge that this is an approximation of much more complex thermodynamic processes, although this is the same description given in Würsch and Craig (2014).

R1, c10: L407: Is $\mathbf{K}$ in Equation (19) identical to $\mathbf{K}_e$ in Eq. (6b)? If so, use $\mathbf{K}_e$ please. If not, what is the difference?

Response to comment #10: We thank the reviewer for having noticed this error, which will be fixed in the text.

R1, c11: L421: If a new DA method or a new configuration has a much larger influence (OID) in convective-scale DA, what is the authors' suggestion?

Response to comment #11: We believe that different OID values would be acceptable as long as they are not too small or too large. Ultimately, idealised models can be used to demonstrate the credibility of a particular approach or configuration, as a precursor of testing in an operational trial with a panoply of diagnostics, rather than just the OID. We will add more text to clarify this point.

R1, c12: L606: "the limitations of the EnKF" What are the limitations?

Response to comment #12: The main limitations of an EnKF are usually the imposition (or assumption) of Gaussian distributions on an estimated covariance and the adoption of linear or weakly nonlinear error models and observation operators to work properly. We will add this to the text.

R1, c13: L611: Should “Fig. 11” be “Figure 11” at the beginning of a sentence?

Response to comment #13: We thank the reviewer for noticing this notation issue, which we will also address.

R1, c14: L616: It seems that “).” is missed after (Fig. 10.

Response to comment #14: We thank the reviewer for noticing the missing parenthesis. This will be fixed in the revised manuscript.

R1, c15: L623: “This shows the impact of a well-calibrated $\mathbf{P}^f_e$ matrix” How do authors define a well calibrated **P**?

Response to comment #15: This sentence has not been phrased correctly. What we meant was the impact of ‘inter-variable correlations’ in the $\mathbf{P}^f_e$ matrix. The text will be modified accordingly.