Note to anonymous referee #2,
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 #2

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 #2, comment #1: The authors aim to show the relevance of using the modRSW model as a tool for mimicking key aspects of convective scale data assimilation in order to justify the transfer of knowledge from a simplified and cheaper setup to an operational configuration. In my opinion, this is a very important and complicated topic which is often overlooked, because it is not straightforward how to tackle it. I think the authors made a good attempt and I encourage the publication of this article, after the authors have considered the following points.

General comments:

1) The topic of this article is very tricky, since much of the relevance of the setup with the modRSW will depend on the purpose of the research. I think the authors should put more focus on the type of research that would be and would not be appropriate with the modRSW. For example, the authors note based on the snapshots that DA can recover the location of convection but struggles with the intensity. We know that operational convective scale DA does have problems with location errors. So for research that seeks to deal with location errors the modRSW may not be suitable. Another important topic among toy model users is non-Gaussianity and positivity constraints on hydrometeors. Is the non-Gaussianity and non-linearity in the modRSW comparable to an operational model? Does the rain get negative values and does it influence the DA results in a similar way as in an operational setup? I encourage the authors to discuss what type of research would be and would not be appropriate with the modRSW.

Response to comment #1: The reviewer raises an interesting point regarding the type of research that can be conducted with the modRSW model. A non-exhaustive list of possibilities include: 1) research on strategic choices when assimilating a growing number of satellite observations; 2) research on comparing different data assimilation schemes in the presence of convection and precipitation at the convective-scale; 3) research on the representativity error (for which both the imperfect model scenario and the flexible observing system used in our setup can be particularly useful); and 4) research on data compression techniques (see Fowler, 2019) to optimise the amount of information that can be extrapolated from a growing body of observations.
In this regard, we would like to mention that (1) has been in fact the focus of Cantarello's PhD thesis (2021), in which an updated (i.e. isentropic) version of the modRSW model – the

ismodRSW model – has been used to conduct idealised satellite data assimilation experiments (see also Cantarello et al., 2022, and Bokhove et al., 2022).

R2, c2: 2) Comparing this modRSW setup to an operational setup skips the natural step of comparing to an idealised setup with operational model. I think it would be helpful to design a similar idealalised setup with an operational model to compare to the modRSW setup, to distinguish between model-caused differences and any other errors sources that come with the use of real observations. After all, in this work we are interested in the relevance of the modRSW, so we want to isolate its role in the DA experiments. Could the authors provide some thoughts on this matter?

Response to comment #2: Unfortunately, the University of Leeds' authors did not have access to an operational model during their research. Also, running the Met Office convective-scale research workflow would have been too complex a task for this research project. Nonetheless, we would like to point out that the data assimilation setup described in our work is largely based on the DA configuration used by the Met Office in their MOGREPS-G system, described in a recently accepted QJRMS paper by Inverarity et al. (2023).

R2, c3: L365: Observations → members , right?

Response to comment #3: This error will be fixed in the revised manuscript.

R2, c4: L370: By discarding negative observations, one creates a positive bias. Is this bias comparable to operational convective scale DA? As mentioned in general comment 1), non-Gaussianity and non-negativity is a popular topic among toy-model users, so I think this point should be explored more elaborately.

Response to comment #4: We appreciate the reviewer's comment, and although we have not discarded negative observations, but rather adjusted their values, we agree that this introduces a positive bias to the observation errors.
Regarding the comparison with the bias found in an operational convective-scale DA system, we cannot answer this question quantitatively. However, our approach can be viewed as a simple form of quality assurance analogous to that which is applied to pre-processed observations in operational NWP. Furthermore, a Gaussian error model is not strictly appropriate when applied to non-negative quantities, but is nevertheless convenient in order to apply a Kalman filter.

R2, c5: L412: the OID of 0.18, is for real observation experiments I assume. Would we expect a lower value for an idealised setup with operational model? As mentioned in general comment 2) shouldn't that be the value to compare the modRSW setup to?

Response to comment #5: The quoted OID was for a global NWP system. The actual value is expected to change depending on the quality of the forecast model (formulation accuracy and resolution, for example) and due to the observation coverage. As such, there is not a single value to compare to, leading to the quoted range, which is an

estimate. We are not aware of a comparable figure for kilometre-scale operational data assimilation systems. We will clarify this point in the text.

R2, c6: L421: I don't fully understand how the thresholds of 20 and 40% are chosen, given the numbers mentioned in the previous paragraph.

Response to comment #6: We acknowledge that the thresholds for the OID mentioned in the text are somehow arbitrary, although they are intended to lead to a configuration where the prior forecast is the dominant contributor, with observations still having a reasonable influence on the updated forecast. We will clarify this in the text.