Editor comments to the author:

I have now (finally) two re-reviews of your paper. I apologize again for the delay in obtaining these reports.
As you see, the comments on your paper are rather different. I also note that some of the issues (e.g. literature) could have been raised earlier. Nonetheless, I would like to ask you to address the comments of rev. #3 in a revised version. Perhaps you could make the "way forward" and "what is new" clearer in your paper.

We are very thankful for the editor's comments. Below, we address the reviewer #3's specific comments (in bold blue) and make an effort to present the "way forward" and "what is new" more clearly in the revised manuscript.

Reviewer #3

Overarching comments

This study examines three case studies of convective storms observed with two X-Band radars in Germany, comparing synthetic polarimetric observations simulated by an ensemble of convective permitting model simulations with observations. The forward operator is applied to the model output with a scheme to correct for propagation effects at X-Band (attenuation, differential attenuation). The study compares the evolution of surface rain rates, convective area fraction, selected time snapshots of polarimetric variables and hydrometeor mixing ratios, and time and space averaged cumulative frequency by altitude diagrams (CFADs) of polarimetric variables to conclude potential shortcomings of the coupled model simulations and forward operator technique.

While the end-to-end forward simulation of polarimetric observables from radar is indeed a potential pathway to diagnose model shortcomings (which is major goal of this study), the findings are qualitative in the sense that due to the necessary compromises in sampling the synthetic observations and the observations themselves that may suffer from significant sampling and representativeness biases, as well as errors due to the application of the forward operators. The authors note some of these potential uncertainties, but do not deal with them. Accounting for these uncertainties, it is not clear what the path forward is in understanding the issues with the model (is there a problem with the kinematics/thermodynamics/microphysics, or is it a shortcoming of the model setup or initial conditions?) towards ultimately improving the state of the art of cloud modeling. Can the authors specify the path forward?

We very much agree with the reviewer that there is a considerable knowledge gap in the fusion of radar polarimetry with atmospheric models. The uncertainties in the assumptions made in the forward operator (FO) and attenuation correction of observations really challenge and question the effectiveness of the model evaluation approaches with polarimetric radar data. However, we have to note here that the assumptions made in the FO

and attenuation corrections applied to the observations are based on the state of art knowledge. We acknowledge these limitations to exercise caution while interpreting the model evaluation, and thereby focus more on patterns to directly evaluate the shortcomings on the modeled signatures of the microphysical processes.

While ensemble simulation for multiple convective storms carried out in this study already shows the importance of initial and lateral boundary conditions for the improvement of simulated precipitating cloud system, a way forward would be to have additional ensemble FO with a combination of hydrometeor parameters to explore the full spread of synthetic polarimetric moments. Earlier, an extensive sensitivity study with the hydrometeor parameters in the FO was also conducted for a stratiform case over the same modelling domain (Shrestha et al. 2021). The model was found to exhibit a low bias in the polarimetric moments above the melting layer, where snow was found to dominate, but none of the alternative shape and orientation setups for snow could provide sufficiently strong polarimetric signals to reproduce observed signals at these heights. This sensitivity study, thus helps to point towards missing shortcomings in the cloud microphysical scheme and/or the scattering estimates using T-matrix calculations. Such low biases above the melting layer are also observed in this study.

Besides, evaluation of synthetic patterns of polarimetric signatures with observation (e.g., Z_{DR} /K_{DP} columns) already helps in identifying the possible deficiencies in the cloud microphysical schemes and the FO. So, in addition to ensemble FO, model evaluation should particularly focus on the dominant polarimetric signatures of the precipitating cloud systems. The model was particularly found to underestimate convective area fraction, width and magnitude of Z_{DR} columns, associated with small sized supercooled raindrops, which could be a result of the fixed CN concentrations and shape parameters of cloud drop size distribution, and the water content determination of the ice hydrometeors in the FO.
So, either the use of regional measurements of CN/IN concentrations, or sensitivity study with large scale aerosol perturbations or use of prognostic aerosol/trace gasses module could be a way forward to minimize the uncertainty in polarimetric signatures due to aerosols. Therefore, sensitivity studies with different parameters for cloud activation/ice nucleation and direct use of a chemistry transport model for prognostic aerosol simulation for the same case study have also been undertaken and are in the process of being submitted. Besides, work also needs to be undertaken to develop additional parameterization in the FO to determine the water content of the ice hydrometeors above the melting layer (e.g., water content for wet growth of hail etc). Furthermore, the modelling and observational community need to work more closely together (as fostered in PROM; Trömel et al. 2021) to achieve the above objectives.

The following text has been added in the revised manuscript:
Ln 567: “For the 2-moment cloud microphysics scheme, the fixed CN concentrations and shape parameters of cloud drop size distribution could also be partly responsible for the overall too low storm intensities, thus regional measurements of CN/IN concentrations, surface precipitation and polarimetric radar data observations could be used together to

constrain the shape parameters of cloud droplets. While regional measurements of CN/IN concentrations might not be readily available, sensitivity study with large scale aerosol perturbations or use of prognostic aerosol/trace gasses module could be a way forward to minimize the uncertainty in polarimetric signatures due to aerosols. On the forward operator for 2-moment cloud microphysics scheme, the water content of the ice hydrometeors can strongly modulate the dielectric constant and hence the scattering properties. This information is not directly available in the forward operator - and the melting parameterization in the FO does not completely compensate for the scattering properties of the ice hydrometeors above the melting layer. So, future advancement in the FO should include parameterization for determining more accurate water content of the ice hydrometeors above the melting layer, which would help in obtaining more accurate dominant polarimetric signatures. Importantly, the prominent polarimetric signature of convective storms like the Z_{DR} column appears to be poorly resolved at km-scale simulations. Future model evaluations with polarimetric radar data should focus on hyper-resolution simulations to better resolve the three-dimensional motion and microphysical processes associated with multivariate polarimetric signatures as well as uncertainty estimates in the attenuation correction of polarimetric moments for convective cases."

Furthermore, the authors note large discrepancies between the snapshots comparisons as well as the statistical comparisons using CFADs. The reviewer notes that this approach has been taken in prior work (Pfeifer et al. 2008; Kumjian et al. 2019; Matsui et al. 2019), which leads to the logical question about what is new here? The present work does not present itself as a novel advance of the state of the topic under investigation. Nor does it present an improvement to methodology, the state of the art in modeling, nor the use of radar data. Thus, the reviewer has serious concerns about the need to publish this paper in the literature. I suggest that the authors take some time in revision to state what is new in this paper.

We are thankful for the reviewer's suggestion of previous works which have taken a similar approach for model evaluation with radar polarimetry. We have included these references in the revised manuscript also, and compared them in context to the findings from this study.

Even though first polarimetric forward operators have already been available several years ago, like SynPolRad introduced in Pfeifer et al. (2008), there is still a considerable knowledge gap in the fusion of radar polarimetry with atmospheric models. The acceptance of PROM project as a special priority programme by the German Research Foundation confirms the novelty and great potential exploiting radar polarimetric observations for a more detailed model evaluation and ensuing improvements (see PROM-overview paper: Trömel et al. 2022)- refinements of forward operator are still ongoing and mandatory for a full exploitation (like connection to scattering data base for better representation of the ice phase) and their full exploitation for model evaluation is still at the beginning, partly because polarimetric precipitation radar networks became just recently available (in Germany since 2016) as well as newly developed tools (e.g. QVPs). Furthermore, each atmospheric model has their own cloud microphysical schemes with different levels of sophistication. In this study, we use the Terrestrial Systems Modeling Platform to perform km-scale ensemble simulations in convection permitting mode, to evaluate 2-moment cloud

microphysics scheme (Seifert and Beheng 2006; hereafter SB2M) for multiple convective storms using a FO with high resolution X-band polarimetric radar data. The 2-moment scheme allows the possibility of aerosol-cloud-precipitation interaction studies and hence the possibility of aerosol effects on polarimetric quantities. The findings from this study points to such possible connections, which will be addressed in an upcoming sensitivity study with cloud activation and ice nucleation parameters. Further, the SB2M scheme is also a candidate for the ICON model used for operational weather forecasting by DWD. And, previous studies have mostly documented polarimetric signatures of convective storms in C or S band, while studies based on high resolution X-band are still gaining ground.

To particularly address the reviewer's concerns, the Introduction section has been revised to better clarify the motivation for this research. Importantly, we agree with the reviewer that the findings should be made clearer and associated/interpreted together with our challenges to present what is new in this study. This has been now discussed and presented more clearly in the revised manuscript in context to polarimetric signatures and statistical distributions of polarimetric variables.

Ln 497: "The synthetic Z_{DR} column signature is a result of the supercooled raindrops only. The missing treatment of freezing raindrops (which do require an additional hydrometeor class) could further be contributing to deficiency in this polarimetric signature (Kumjian et al., 2014). And, to a certain extent, the absence of polarimetric signature contribution from wet growth of hail, which is not parameterized in the FO, could additionally be contributing to the deficiency in the shape and magnitude of the synthetic Z_{DR} column. Besides, the mean diameter size of the raindrops strongly controls the magnitude of polarimetric signature. A reason for the relatively small mean diameter size of supercooled raindrops could be due to high CN concentrations and the missing feedback between the CN concentration and shape parameters of cloud drop size distribution (Noppel et al., 2010). A sensitivity study with low CN concentrations for case one in fact produced high hail concentration, which increased the CAF, Z_{DR} and Z_H magnitudes of the storm (Trömel et al., 2021)."

Ln 527: "The Z_{DR} CFADs from the ensemble simulations exhibit narrow distributions with peak values near zero above the melting layer, which does not differ among the three case studies. It also exhibits bimodal peaks below the melting layer compared to unimodal distribution in observations. Similar bi-modal CFADs of Z_{DR} was also reported by Matsui et al. (2019) for a simulated mesoscale convective system over Southern Great Plains, USA using both spectral bin microphysics and single moment cloud microphysics scheme, while the observed CFADs of Z_{DR} exhibited a more smoother gradient below the melting layer as shown for the observation in this study as well. In their study, sensitivity studies with FO parameters also could not reproduce the distribution similar to the observations, while the effect of sensitivity study was found to differ between the two microphysics schemes. In this study, the model tends to strongly underestimate the maximum reflectivities for case one but generally it exhibits a broader distribution of Z_H for all three cases compared to the observations, with a peak around 30 dBZ above the melting layer. This higher reflectivity is caused by the dominance of graupel as discussed above. Consequently, the precipitation production by melting of graupel/hail below the melting level, as shown in the cross sections of model simulated hydrometeors for all cases, could explain the second Z_{DR} peak at approximately 2 dB in

the lower levels. This possibly indicates that the modeled mechanism of precipitation formation below the melting layer differs from the observation. Furthermore, the use of a functional form of drop size distribution in the FO leading to a unique mapping between modeled quantities and synthetic polarimetric quantities can create errors (Kumjian et al., 2019), which could also be partly contributing to this bi-modal peak behavior in the synthetic Z_{DR} CFADs."

Ln 544: "Thus, the observed variability in Z_{DR} and K_{DP} above the melting layer is underestimated in the synthetic polarimetric variables. Part of this reduced variability can be explained by the deficiencies of the FO. Earlier, an extensive sensitivity study with the hydrometeor parameters in the same FO was conducted for a stratiform case over the same modelling domain (Shrestha et al., 2021). In their study, the model was found to exhibit a low bias in the polarimetric moments above the melting layer, where snow was found to dominate, but none of the alternative shape and orientation setups for snow could provide sufficiently strong polarimetric signals to reproduce observed signals at these heights. The inability to reproduce the polarimetric characteristics of snow with T-Matrix also justifies the need for a scattering database. This issue needs to be revisited with more sophisticated forward operators available in the future (already planned in this project). For ρ_{HV}, the CFADs are poorly simulated by the model, probably due to the shortcomings in FO assumptions on diversity of hydrometeor shapes and orientation (Shrestha et al., 2021). Although, the synthetic ρ_{HV} exhibits a very homogeneous high value above the melting layer, it exhibits slightly reduced magnitude in locations with elevated Z_{DR}. This pattern was found to be consistent for all simulated case studies."

Specific comments

L133: remove "for those"

Done.

L228: "connected to" should be "associated with a"

Corrected.

L233: lightnings should be singular

Corrected.

L236: rephrase "connected with" again.

Done.

L241: this sentence should be rephrased or removed. It is sufficient to say that a warm front extended across central Germany. The "wave like feature" is also not apparent in Figure 2c.

The sentence has been removed.

L252: Do you quantify “gradient” or just accumulation frequency?

We only quantify the frequency distribution of accumulated precipitation. The sentence has been rephrased for clarity.

Ln 275: “Overall, the spatial pattern of ensemble averaged accumulated precipitation resembles the RADOLAN estimates, but the frequency distribution produced by the ensemble members underestimate high precipitation.”

L272: It is unusual to start a sentence with “And”
Removed and rephrased.

Figure 5: Label with Case 1 as in Fig 4.
Added in caption.

L332: “hails” should be “hail”
Fixed.

L338-9: L352: “upto” should be two words.
Fixed.

L475: “shift” -> “shifts”
Fixed.