General comments:

The paper presents an interesting study that statistically compares polarimetric radar observations from three storm systems in northwestern Germany to ensemble modeling results. I enjoyed reading the paper and found the results to be interesting, valuable, and very worthy of publication. I do, however, have several comments that I feel would improve the manuscript. I will address these in more detail in the next section.

We are very thankful for the reviewer's comments. Below, we address the reviewer's specific comments (in bold blue).

Individual comments:

1) While I do have a background in polarimetric radar observations and, to some degree,the use of polarimetric radar data in numerical models, I am unfamiliar with most of the models used in this study. Upon my first reading, I must admit that I was a little overwhelmed by the numerous acronyms that were being introduced and found myself continually going back to refresh my memory. After gett8ng a few pages in, I skipped to the back of the paper to see if there was perhaps an appendix that summarized the list of acronyms. Is this something the authors might consider?

We think it's a valid suggestion. In the revised manuscript, we have moved much of the acronyms to the Abbreviation section, which has improved the text flow in the manuscript. The acronyms are also better described here.

2) In section 3.2, I really don't feel like I had a good understanding of how the 20 ensemble members for each case were obtained. That is, it states the 20 ensemble members represent "uncertainties in model physics and lateral boundary conditions by combining five model physics perturbations with four global models are used for the initial and lateral boundary conditions". I found this description to be a little vague. Can you be more specific about what those five model physics perturbations and, more importantly, four global models are? And the results from those runs are used as the initial and boundary conditions for the COSMO runs? Also, what does COSMO-DE stand for? COSMO is introduced earlier but, looking back into the paper, I was unable to find what COSMO-DE referred to.

COSMO-DE is a high resolution (~2.8 km) configuration of the COSMO model encompassing the entire extent of Germany including some neighboring countries; thus "DE" is the code for Germany (Deutschland). The 20 ensemble members of COSMO-DE EPS can be divided into 4 subsets of 5 members each. The 4 subsets represent the different global models used for lateral boundary conditions and initialization: the Integrated Forecast System of ECMWF (IFS) [Janssen and Bidlot, 2002], the global model of DWD (GME) [Majewski et al., 2002], the Global Forecast System of NCEP (GFS) [Environmental Modeling Center, 2003] and the Unified Model of the UK Met Office (UM) [Staniforth et al., 2006]. Each of the 5 members in the subsets use different parameter sets in their parameterizations. These parameter sets govern the entrainment rate for shallow convection, the critical value for normalized oversaturation, the scaling factor of the laminar boundary layer for heat, and the asymptotic mixing length of turbulence. For details we refer to Gebhardt et al. (2011) and Peralta et al. (2012), which are also cited in the manuscript.

The hourly output from the COSMO-DE EPS provided by DWD is then used for the model runs in this study. For clarity, the description in Section 3.2 has been modified:

Ln 154: "The hourly model output from the 20 ensemble members of the COSMO-DE Ensemble Prediction System (EPS: Gebhardt et al. 2011; Peralta et al. 2012) provided by DWD is used for the model runs in this study. The COSMO-DE is a high resolution (~2.8 km) configuration of the COSMO model encompassing the entire extent of Germany. The 20 ensemble members of COSMO-DE EPS can

be divided into 4 subsets of 5 members each. The 4 subsets represent different global models: the Integrated Forecast System of ECMWF (IFS) [Janssen and Bidlot, 2002], the global model of DWD (GME) [Majewski et al., 2002], the Global Forecast System of NCEP (GFS) [Environmental Modeling Center, 2003] and the Unified Model of the UK Met Office (UM) [Staniforth et al., 2006], used to vary the boundary conditions of the COSMO-DE. Each subset of the 5 members is then perturbed by varying a set of parameters that control the physics parameterization of the COSMO model. "

At the beginning of section 5.3, there is a short discussion of clustering. Clustering, I believe, refers to how combined plots of two polarimetric variables will cluster in multidimensional space. This seems totally unrelated to computing convective area fractions of a single radar variable, such as reflectivity. Also, can the authors provide a more complete description of convective area fraction (CAF)? I know this is a concept that has been used in numerous papers, but without description I am often left wondering if the convective area fraction is with respect to the grid being used, or with respect to all reflectivity points (for example) above a certain dBZ threshold? It seems to me that a CAF can be defined in many different ways. Also, how is CAF impacted if, for example, a portion of the system that is being studied is moving off of the grid over which the CAF is being sampled?

Yes, the reviewer is correct, here clustering refers to how multiple polarimetric variables will cluster in multidimensional space. This clustering also depends on the stages of storm evolution. So, here we use the temporal evolution of the convective area fraction to identify the development stage of the storm in the measurements and in the synthetic radar data. This is then used to minimize effects of the mismatches in space and time when comparing both.

In this study the convective area fraction (CAF) is the area fraction of the storm with Z_H>40 dBZ at 2 km height a.g.l. divided by the total area of the storm which encompasses the grid points of the storm with Z_H>0 dBZ at the same height. Only CAF evolutions were compared for which the storm stayed within the domain. This restriction affected e.g., Case 1 and 2 when for some members the storm approached the boundaries in the last 30 minutes. Due to extended sampling time used in Case 3 the compared time interval is reduced because the storm moved off the grid in the simulations.

The following text has been added in the revised manuscript:

Ln 270: "The total area of the storm for CAF estimate, includes the grid points of the storm with radar reflectivity >0 dBZ at 2 km height a.g.l. The time extent of the CAF evolution was chosen such that the storm is within the domain. However, due to variability in the ensemble members, some members are affected as part of the storm approaches the boundary in the last 30 minutes of CAF evolution for Case 1 and 2. And, due to extended sampling time used in Case 3, the CAF is partly impacted by the storm moving off the grid for the synthetic data."

3) In sections 5.3.1, 5.3.2, and 5.3.3, I am confused why the elevation angle 8.2 is used for a PPI for cases 1 and 3 (Figs. 5 and 9) and an elevation angle of 1.0 is used for case 2 (Fig. 7). Using an elevation angle of 8.2 for a PPI seems very unusual. Please explain why such high elevation angles are being used for these plots.

An 8.2° elevation angle for a PPI is unusual for near-surface quantitative precipitation estimation for which we use indeed PPIs measured at 0.5° or 1° elevation angle or terrain-following scans. In recent years, however, volume scans consisting of a series of PPIs measured at different elevations, mostly between 0.5° and 30°, became more popular in order to get a 3D picture of hydrometeors and microphysical processes e.g. for improved process understanding, model evaluation and data assimilation. Such volume scans also enable us to construct vertical cross-sections of convective systems (e.g. Fig. 5b). Choosing a PPI measured at higher elevations for Fig. 5a and Fig. 9a gives insights also of the measurements at different heights (~1 km, near melting layer and 2~3 km above melting layer) of the deep convective systems (radar

measures at increasing height with increasing distance from the radar). Together with the spatial extent and location of the system, these figures complement the cross-section in Fig. 5b and Fig. 9b. Case 2 instead, was less vertically extensive and further distant from the radar, so 1° scan was used to explore the low-level features.

For clarity on the use of the different elevation scans, we have added the following text in Section 3.3 of the revised manuscript:

Ln 182: "Both X-band Doppler radars produce volume scans consisting of a series of Plan Position Indicator scans (PPIs) measured at different elevations, mostly between 0.5° and 30°. The use of these multiple sweeps became more popular in recent years in order to get a 3D picture of the spatial distribution of hydrometeors and microphysical processes. These PPIs can be exploited for improved process understanding, model evaluation and data assimilation. And, such volume scans also enable us to construct vertical cross-sections of the convective systems."

Ln 211: "Based on the time and the distance of the storm from the radar for the different cases, PPIs measured at different elevation are used - to provide optimal insights in convective systems at different heights (~1 km, near melting layer and 2~3 km above melting layer)"

4) Overall comment on the figures, my philosophy has always been that figure captions should contain enough information that they could be "stand alone", i.e., that the reader should be able to fully interpret the figure without having to refer back to the text. That being said, I feel that much can be done in this manuscript to improve figure captions and, in a few cases, the figures as well. As an example, I felt that the caption describing the right most panel of Fig. 2 could have been much better, particularly the description of the rightmost panel.

The caption for Fig. 2 has been improved in the revised manuscript:

…"The right panel shows the frequency distribution of accumulated precipitation for each ensemble member (light grey dashed line) and observation (black dashed line). The inset in the right panel shows the domain average accumulated precipitation for each ensemble member (light grey color bar) and observation (black color bar) with one standard deviation (solid line above the bars)."

The captions for the remaining figures has also been improved, where applicable.

5) The text states that there were 104 GRDC stations, 36 were considered useful for this study. These figures show 22 or 23 stations, but the figure is very crowded with all of the stations grouped together in the rightmost 2/3rds of the figure with lots of unneeded and wasted white space on the left, etc. If several of the stations are not going to be used (presumably those that were to be plotted on the left side of each of the figures), I would suggest eliminating the "which space" and making the figure easier to read. Also, are there 20 asterisks representing the 20 ensembles plotted for each station with the some of them just overlayed on each other?

The model evaluation with GRDC station data has been removed from the revised manuscript to keep the focus on the model evaluation with polarimetric radar data. This was suggested by reviewer #2.

6) Figure 4: Time labels need to be improved.

We assume that the reviewer here refers to the missing second digit in the minutes of some labels. This has been fixed in the revised manuscript.

Specific comments:

1) Overall, the paper is well written. There are some very minor grammatical issues throughout the text. I'll make just a few suggestions here.

We are again very thankful for the reviewer's comments. Below, we address the reviewer's suggestions.

2) There is also some inconsistency throughout the paper on whether the word "modeled"and "modeling" should be spelled with one "l" or two "ll"s.

Fixed. We have now used "modelling" consistently.

3) Line 7: Remove "however".

Corrected.

4) Line 9: Suggest replacing "besides" with "in addition to".

Corrected.

5) Line 20: Suggest rewording from "Polarimetric radar observations provide ZDR,…" to"Besides ZDR, polarimetric radar observations provide…"

Corrected.

6) Line 32: Suggest changing "thus e.g. provides insight on new snow generation" to"thereby providing insight into the generation of new snow".

Corrected.

7) Line 33: Suggest changing "measure for the diversity" to "measure of the diversity".

Corrected.

8) Line 34: Change "These informations" to "This information".

Fixed.

9) Line 47: Remove "e.g."

Fixed.

10) Line 49: Suggest changing "operator and due" to "operator due".

Here, the "uncertainty in the model evaluation in radar space" stems from the "uncertainty in the assumptions made in the forward operator" and the "uncertainties of polarimetric radar measurements".

11) Line 50: Remove "e.g."

Fixed