The authors would like to thank the reviewer for their insightful comments. A point by point response to the reviewer's comments, along with changes made to the manuscript as a result, are included below.

R1. The advantage of basing forward calculations on empirical relationships, as opposed to fundamental radiative transfer and scattering theory, is not well established in the manuscript. One justification for the approach in the paper is that using empirical relationships means that one does not need to make assumptions about scatterers (e.g., a spherical assumption), but this simply exchanges a known assumption with an assumption (or set of assumptions) hidden in the empirical relationships. And most of these empirical relationships are actually retrievals, just inverted! If we're going to do forward calculations based on retrievals, we might as well just use the retrievals on the observations and cast all the quantities in terms of geophysical variables, which are easier to interpret. This approach seems like a big step back compared to performing fundamental radiative transfer/scattering calculations on the model fields, which yields an independent forward calculation of the observational fields. Furthermore, the assumptions in the empirical relationships may not be consistent with the assumptions in the model cloud microphysical parameterization (e.g., the assumed distributions). Consistent forward calculations of model variables should use assumptions consistent with the cloud-physics scheme in the model.

A1. In response to this comment by the reviewer we now elaborate within the manuscript on the reasoning behind our approach:

"Hydrometeor properties that impact backscattering include size, phase, composition, geometrical shape, orientation and bulk density. Were plausible representations for these hydrometeor properties available as part of the model formulation, fundamental radiative scattering transfer calculations would be the most accurate way to transform model hydrometeor properties to observables. However, in most GCMs such detailed hydrometeor information is highly simplified (e.g., fixed particle size distribution shapes) or not explicitly represented (e.g., orientation and realistic geometrical shape), complicating the process of performing direct radiative scattering transfer calculations. Chepfer et al. (2008) proposed an approach by which lidar backscattered power can be forward-simulated using model output hydrometeor effective radius. Their approach, based on Mie theory, relies on the assumption that cloud particles (both liquid and ice) are spherical and requires additional assumptions about hydrometeor size distributions and scattering efficiencies. Similarly, the COSP (Bodas-Salcedo et al., 2011) and ARM Cloud Radar Simulator for GCMs (Zhang et al., 2017) packages both use QuickBeam for the estimation of radar backscattered power (i.e., radar reflectivity; Haynes et al., 2007). QuickBeam computes radar reflectivity using Mie theory again under the assumption that all hydrometeor species are spherical and by making additional assumptions about the shape of hydrometeor size distributions as well as mass-size and diameter-density relationships. While some of these assumptions may be consistent with the assumptions in model cloud microphysical parameterizations, some are not adequately realistic (e.g., spherical ice) or complete for accurate backscattering estimation and it is typically very difficult to establish the sensitivity of results to all such assumptions.

To avoid having to make ad hoc assumptions about hydrometeor shapes, orientations, and compositions, which are properties that also remain poorly documented in nature, $(GO)^2$-SIM employs empirical relationships to convert model output to observables. These empirical relationships based on observations, direct or retrieved with their own sets of underlying assumptions, are expected to capture at least part of the natural variability in hydrometeor properties. Additionally empirical relationships are computationally less expensive to implement than direct radiative scattering calculations, thus enabling the estimation of an ensemble of backscattering calculations using a range of assumptions in an effort to quantify part of the backscattering uncertainty (see Sec. 7). The empirical relationships proposed require few model inputs, potentially enhancing consistency in applying $(GO)^2$-SIM to models with differing microphysics scheme assumptions and complexity. Section 6 will show that, while the empirical relationships employed in $(GO)^2$-SIM may not be as exact as direct radiative scattering calculations, they produce backscattering estimates of sufficient accuracy for hydrometeor phase classification, which is the main purpose of $(GO)^2$-SIM at this time."

R2. The manuscript advocates a phase determination that is solely in forward-calculation space and fairly well articulates the reason for this. However, this approach does not take advantage of knowing the actual hydrometeor fields, and therefore this discards a great deal of potentially useful information. Is there any way the approach in the manuscript can take some advantage of the fields in hydrometeor (model) space?

A2. As articulated in the manuscript our goal is "[…] development of a phase classification algorithm that can be applied to observables, forward-simulated and real." This explains why we avoided developing a hydrometeor-phase classifier dependent on model output quantities that are not accessible via observations. Rather, we take advantage of the fields in model space by using them to 1) evaluate the ability of Doppler velocity and Doppler spectral width observations to be used for hydrometeor phase classification (a concept which was developed empirically and was not formally evaluated) and to 2) select optimum classification thresholds to minimize false detection in model space.

This reasoning is expressed in the following modified manuscript excerpts:

"While the thresholds used for the radar reflectivity, lidar backscattered power, and lidar depolarization ratio are generally accepted by the remote sensing community, the same cannot be said about the radar Doppler velocity and Doppler spectral width thresholds suggested by Shupe (2007). Because simulated mixing ratios of liquid and ice hydrometeors are known in the $(GO)^2$-SIM framework, the use and choice of all such thresholds for phase classification can be evaluated using joint frequency of occurrence histograms of hydrometeor mixing ratios for a single species and forward-simulated observable values (resulting from all hydrometeor types; Fig. 6)."

"The objectively determined thresholds, based on model output mixing ratios, optimize the performance of the hydrometeor phase classification algorithm and are expected to generate the best (by minimizing false detection) hydrometeor phase classifications. Results using these objective flexible thresholds are compared in Sec. 6.4 to results using the fixed empirical thresholds of Shupe (2007)."

"The performance of the objectively determined flexible phase-classification thresholds (illustrated using colored dashed lines and shading in Fig. 7) is examined against those empirically derived by Shupe (2007) with one exception (illustrated using grey lines in Fig. 7). The modification to Shupe (2007) is that radar reflectivity larger than 5 dBZ are not associated with the snow category since introducing this assumption was found to increase hydrometeor-phase misclassification (not shown). From Fig. 7 it is apparent that both sets of thresholds are very similar. We estimate that hydrometeor phase frequency of occurrence produced by both threshold sets are within 6.1 % of each other and that the fixed empirical thresholds modified from Shupe (2007) only produce phase misclassification in an additional 0.7 % of hydrometeor-containing grid cells (compare Table 1b to Table 1c). These results suggest that the use of lidar-radar threshold-based techniques for hydrometeor-phase classification depends little on the choice of thresholds."

R3. Constructing an ensemble of forward calculations based on different empirical relationships is a good idea, but it is a stretch to portray it as quantifying uncertainty. The authors have no way to know to what extent the results from these calculations actually map to the PDF of possible outcomes. It is useful but is not statistically defensible to call it UQ. The authors should much more carefully word this claim.

A3. The authors agree with the reviewer that the 576 forward-simulations performed do not cover the entire range of possible scattering assumptions. The following manuscript changes reflect this reality:

"Additionally empirical relationships are computationally less expensive to implement than direct radiative scattering calculations, thus enabling the estimation of an ensemble of backscattering calculations using a range of assumptions in an effort to quantify part of the backscattering uncertainty (see Sec. 7)."

"$(GO)^2$-SIM performs an uncertainty assessment by performing an ensemble of 576 forward simulations based on 18 different empirical relationships (relationships are listed in Table 2). While the relationships used do not cover the entire range of possible backscattering assumptions, they represent an attempt at uncertainty quantification and illustrate a framework for doing so. […] Nevertheless, we suggest using the full range of frequency of occurrences presented in Tables 1b,c for future model evaluation using observations and acknowledge that additional uncertainty is most likely present."

R4. The calculations are based on 30-minute instantaneous model hydrometeor fields. The article is focused on the actual forward calculations of the microphysical fields, but comparison of forward-model calculations and observations necessarily includes assumptions of spatial and temporal scale. Would the authors please discuss with a bit more detail on how the forward calculations (30-minute instantaneous calculations of lidar and radar fields) would be compared to observations? If nothing else, this would provide some guidance for readers using their forward simulator.

A4. We now elaborate more on this topic and provide an updated flow chart:

"A follow-up study will describe an approach by which vertical and temporal resampling of observations can help reduce the scale gap. Furthermore, it will be showed that, using simplified model evaluation targets based on three atmospheric regions separated by constant pressure levels, ground-based observations can be used for GCM hydrometeor-phase evaluation."

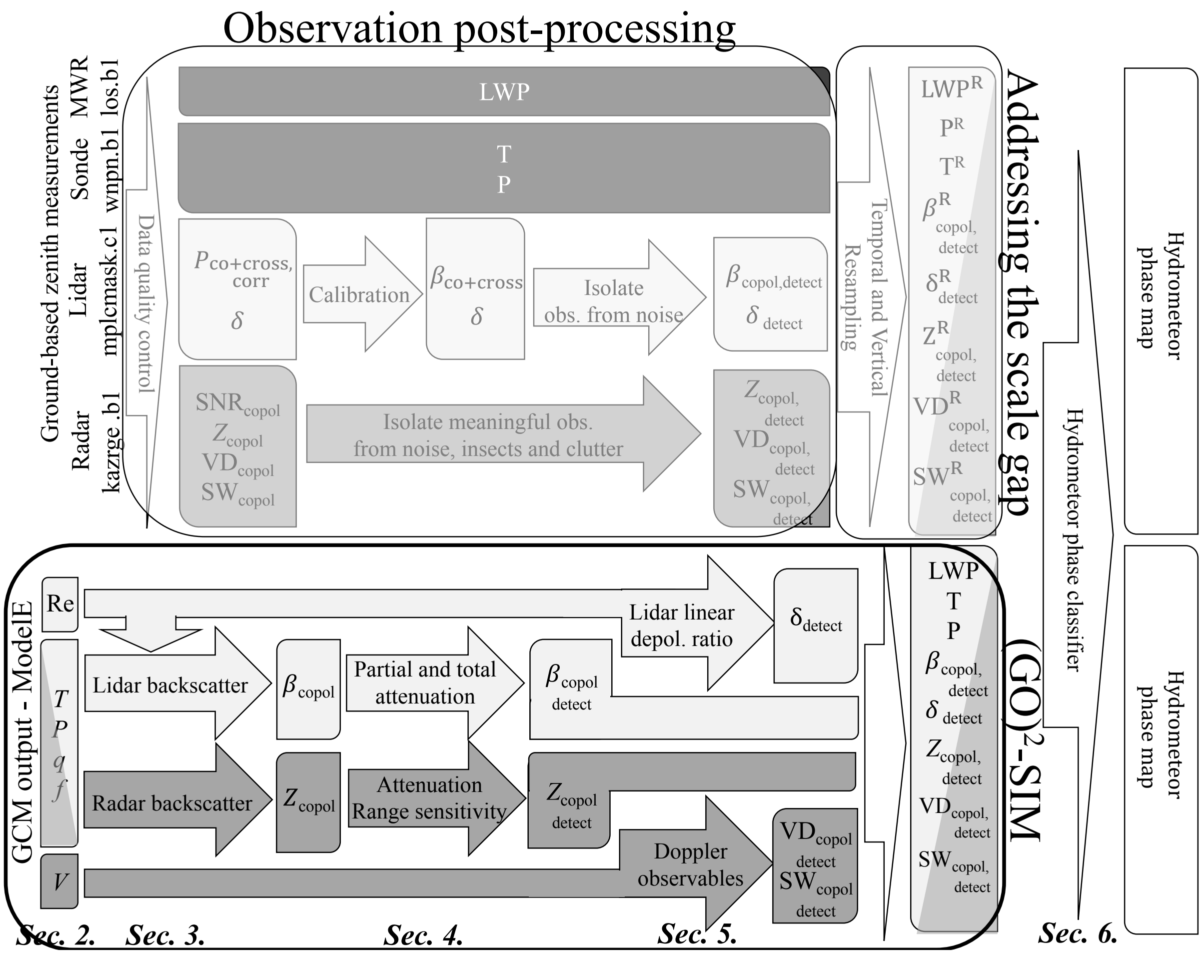

Figure 1. $(GO)^2$-SIM framework. $(GO)^2$-SIM emulates two types of remote sensors: Ka-band Doppler radars (dark gray shading) and 532 nm polarimetric lidars (light gray shading). It then tunes and applies a common phase-classification algorithm (white boxes) to both observed (upper section) and forward-simulated (bottom section) fields. Follow-on work will describe how observation can be post-processed and resampled to reduce the scale gap before model evaluation can be performed.