Reply to Anonymous Referee #2

This study presents an evaluation of the Unified Model (with the CASIM module) using wave cloud aircraft observations. Multiple heterogeneous freezing parametrizations are examined and compared with the observations. In addition, a large number of the KiD model simulations are used to examine the impact of the mountain wave period and cloud-top temperature on the cloud evolution. Finally, the authors provide a conceptual model based on the acquired knowledge, which estimates the wave cloud-driven redistribution of water vapor.
I find this manuscript quite comprehensive and well written. I like the conceptual model idea, although it is obviously case-dependent (as suggested by the authors as well). However, I think that additional work needs to be done before this manuscript can be published in ACP (Major revisions).

Major comments

- "*UM results: I am not convinced that there is a very good agreement between the UM and the observations (l. 239-244). The statement in the text is subjective, that is, in terms of percentage, the specific humidity errors are very large, on the order of up to several tens of percent (see fig. 2b). Same for the vertical velocity amplitude - errors of 1 m/s can be on the order of ~50% relative to the observations (e.g., leg 2 – fig. 2b). As the UM simulations serve as a benchmark for the KiD simulations, the implications of a weak agreement between the UM and the observations could be significant.*"

  Reply: It is challenging to take a global model analysis and nest down to 250m resolution to have the waves exactly match 4-5 hours into the simulation. There are limitations imposed by the initial analysis fields, the representation of the orography, drag and dynamics as well as the microphysics. We give numerical values to indicate how close the simulation is to observations To our knowledge this is the first ever study, where such a direct comparison of model and observations has been attempted for orographic wave clouds.

  The specific humidity in Fig. 2 is about 10% different for the interpolated values compared to the measured values, which is well within the predicted variability. It is very unlikely for the model to predict the observed humidity at exactly the time and location of the observations. However, the model values taken over the one hour interval around the measurements contain the observed specific humidity values. As the model suggests a quite significant temporal variability of the upstream moisture profile and no continuous information (in time or vertical profiles) is available, it is very hard to speculate how the observed differences affect the condensate content at points inside the cloud. The aircraft measurements do not allow for a quasi-Lagrangian approach, which would be necessary for robust assessment of the modelled condensate content.

  Regarding the vertical velocity (Fig. 3), there are differences in the vertical velocity field. Peak-to-peak magnitudes are captured to within 30% and the wavelength appear similar although admittedly harder to quantitatively specify when only a couple of wavelengths are observed. However, the peak velocity is not the most important aspect on its own, rather the absolute height displacement experienced by a parcel (which will be linked to the maximum velocity and wavelength). This is difficult to ascertain for the observations due to only sampling at one level. In addition, aircraft measured vertical velocities can have systematic errors of up to several tenths of a metre per second (see Field et al. 2012: The absolute accuracy of the vertical wind measured from an aircraft is limited by the accuracy at which the aircraft angle of attack and height above ground is known. Typically this would result in a systematic error on the order of tenths of a meter per second). Based on this uncertainty of observations and the uncertainty in numerical model predictions mentioned above deviation between model and observations of around 1 ms^{-1} is judged as very good.

  The matching of the humidity curve between the model and observation (Fig. 4a-c) is probably the best test of the combined representation of thermodynamic structure and dynamic evolution of the model. It can be seen in all three passes that the model specific humidity west of 105°E is generally within 10% of the observed value (which itself has an error of about 1-3%), and east of

105°E. where the differences in ice treatment become more important, the modelled specific humidity is within 10-30% of the observed value.

The UM simulations are not strictly used as a benchmark for the KiD simulations. The comparison of the UM to the observations serves only to indicate that the CASIM microphysics seem to be able to roughly capture the general cloud microphysical evolution within the cloud. Bearing the above discussed levels of agreement and the challenges for a more vigorous assessment in mind we believe this conclusion is valid. The KiD simulations in turn are used to expand on the UM simulations in order to sample a larger section of the relevant parameter space. As already pointed out in the conclusion and discussion section a more comprehensive measurement campaign is needed to provide true observational constraints on the UM simulations as well as the conceptual model. This is more prominently highlighted in the discussion section now.

Changes to manuscript: We added some extra text including the reasoning above in section 3.1 and 3.2 on why we think the match between UM and observations is good. In the conclusion a paragraph has been added regarding the validity of the conceptual model given the levels of agreement between model and observations as well as on the possibility to obtain more vigorous constraints on the model from observations. (modifications / extra text: lines 250-253, 260-261, 263-273, 664-675)

- "*Homogeneous freezing regime: given the fact that for a large part, the heterogeneous freezing parametrization is examined here, the UM simulations, and hence, also a large fraction of KiD simulations, are "contaminated" by homogeneous freezing in the top of the cloud layer (e.g., fig. 7), which obviously non-linearly impacts the underlying heterogamous cloud layer. The authors did not refer to the homogeneous freezing parametrization in the UM and how well it corresponds with the parametrization in the KiD model. Now, I presume that the UM is too complicated to vary the homogeneous freezing parametrization, but I suspect that the influence of other parametrizations can be examined in the KiD model. An additional approach to address large parts of this comment would be to run the UM initialized without homogeneous freezing influence on the cloud layer (e.g., by offsetting the temperature profile to higher values while retaining the RH profile). I wonder how much would the results of this study change in that case (e.g., deviations of the UM from the observations)?*"

  Reply: We have now included a reference for the used homogeneous freezing parameterisation in the CASIM description. CASIM is used in the UM and KiD, i.e. both models have exactly the same microphysics representation. Additionally, a new simulation has been performed, in which homogeneous freezing is switched off. As is evident from the results (which are now included), homogeneously formed ice crystals do have no impact on the cloud microphysics in the updraft region of the cloud. However, they significantly contribute to the ice crystal mass and number concentration in the downdraft region (as was already stated in the original manuscript). Hence the main conclusions in the paper regarding the evaluation of heterogeneous freezing parameterisations is valid irrespective of homogeneous freezing. Discussion of this new simulation and the implications have been included in the new manuscript, where appropriate.
  Of course, differences in homogeneous freezing will have an impact on the sedimentation fluxes (as will uncertainties in the diameter fallspeed relationship). It is beyond the scope of the present study to investigated the uncertainty of sedimentation fluxes due to these issues. Also sedimentation fluxes cannot be verified with currently available observational data. In so far the KiD model results should only point to interesting parts of the phase-space that should be sampled in future campaigns to provide constraints on sedimentation fluxes. In the conclusions we included a stronger statement alerting readers to these additional sources of uncertainty.
  Changes to manuscript: modifications / extra text: lines 214-218, 304-307, 320-322

- *Deposition nucleation is not mentioned at all in the text, although the current understanding is that its efficiency significantly increases when we approach the homogeneous freezing regime (e.g., Kanji et al., 2017, https://journals.ametsoc.org/doi/full/10.1175/AMSMONOGRAPHS-D-16-0006.1). I understand that we would typically except immersion freezing to still account for most* *of the nucleation, but the deposition mode should still be mentioned in the introduction as well as in the model description, even if it is eventually omitted from the simulations.*

Reply: Deposition nucleation is currently not represented in CASIM. We added a few sentences on deposition nucleation in the introduction and the CASIM description. However, as the reviewer already states, this nucleation mode is most likely not very relevant in orographic wave clouds.
Changes to manuscript: modifications / extra text: lines 64, 79, 210-211

- *Negligible impact of rain processes in the KiD simulations and neglecting rain processes in the conceptual model: the minor impact of rain could be driven by: a. the dominance of the homogeneous ice precipitation from cloud top, b. high cloud droplet number concentration that leads to weak collision-coalescence in the model, and/or c. influence of the collision-coalescence kernel implemented in the model. I am not convinced that rain processes are indeed negligible and whether they should be neglected in the conceptual model. If necessary, I presume that adding a "rain component" to the conceptual model should not be a difficult task, given previous conceptual models for warm clouds (l. 435-436).*

  Reply: The CASIM microphysics used in both the UM and in the KiD model do include rain formation processes (autoconversion and accretion, following Khairoutdinov and Kogan (2000)). The maximum rain mass mixing ratio in the UM simulations is more than a magnitude smaller than that of either liquid or ice in the UM ($<10^{-5}$ kgkg^{-1} compared to 10^{-4} kgkg^{-1}) and even smaller than that in the KiD simulations ($<10^{-10}$ kgkg^{-1}). In general, bulk microphysics schemes produce rain too early when compared to a detailed size resolved microphysics scheme (Hill et al. 2015). We think the time air parcels spend in the updraft region is too short for significant rain formation. Most air parcels spend less than 15 min in the updraft region (based on the trajectory analysis, not shown). According to considerations of typical timescales for rain production (Seifert and Stevens, 2008; Miltenberger et al. 2015) this is too short for significant rain formation.
  We have checked the simulation results again for the impact of rain and cloud droplet sedimentation for the overall downward moisture transport. The contribution from rain sedimentation is below 1% for all cloud top temperatures. However, sedimentation of cloud droplets adds considerably to the total downward moisture transport for cloud top temperatures warmer than -32°C. For colder temperatures homogeneously formed ice dominates the downward moisture transport. As the sedimentation flux for cloud droplets depends only on the advective timescale, i.e. the time period used in the simulations, but if included in the downward moisture transport confuses the fitting of the timescales, we suggest adding this separately.
  We have included this discussion in section 4 of the manuscript.
  Changes to manuscript: modifications / extra text: lines 472-481, 527-540

- *Importance of periods longer than 1000 s (stated in the abstract and l. 554-557): do not believe this conclusion, as the authors did not consider different obstacle heights (eta values I presume). As a result, the impact of the (likely maximum) vertical velocity is nearly ignored in the parameter space evaluation, although it should definitely impact the longevity of the wave cloud, among other factors, via ice processes, especially in the homogeneous freezing regime, where we would expect to lose all condensated droplets relatively fast.*

  Reply: The in-cloud timescale essentially scales with the upstream flow velocity and the obstacle width if linear gravity wave theory is used (e.g. Miltenberger et al. 2015). A timescale 1800s corresponds to a mountain width of 18 (54) km assuming a horizontal wind speed of 10 (30) ms^{-1}. Note we are considering here mid-level, isolated wave clouds and not thick orographic clouds producing significant surface precipitation (in contrast to previous conceptual models for orographic precipitation such as e.g. Smith and Barstad 2004 or Miltenberger et al. 2015). For the former typically the width of single mountains is essential, while for the latter typically the width of the mountain range, instead of the width of isolated mountains is more representative. For lee-wave clouds typical wavelength reported in literature are shorter than 20 km (Grubisic et al. 2008), i.e. are well covered by the time period range investigated here. For cap clouds there are to our knowledge no estimates available, but the likely spatial extend estimated from photographs is also on the order of 10-50 km.

Of course, for clouds containing ice crystals (or other hydrometeors with long evaporation timescales) the mountain height influences the cloud extend by effecting how much water is condensed and how long the ice crystals can survive in sub-saturated regions. This effect is independent of the wavelength (or time period of the wave clouds). Although it would be interesting to investigate this, systematically investigating this effect would require a substantial amount of additional simulations (at least doubling the 45000 simulations already analysed). This is beyond the scope of the present paper.
We have included a paragraph in the description of the KiD set-up as well as in the discussion section pertaining to this issue.
Changes to manuscript: modifications / extra text: lines 172-180, 644-650

Minor comments:

- *l. 49-52 – That is a rather complex sentence. I suggest rephrasing or breaking into two separate sentences.*

  Reply: The sentence has been split into three sentences.

- *l. 109 – ICE-L acronym definition is missing.*

  Reply: added.

- *l. 109 - please add parentheses to the citations.*

  Reply: done

- *l. 109 – inconsistent date (November 17th) with the following sections and figures.*

  Reply: corrected. Thanks for spotting this.

- *l. 127-130 – These two sentences seem redundant - repeating information already provided in the Introduction.*

  Reply: removed.

- *l. 146 - Suggest consistency in the number of fractional coordinate digits.*

  Reply: done

- *l. 147 – focusses –> focuses*

  Reply: corrected

- *l. 167 - Please define the source for this eta value - I will presume that it is equivalent to the obstacle (mountain) height*

  Reply: This value is based on the mean maximum η value of trajectories in the UM model, which pass through the wave cloud. A sentence stating this has been added. η is roughly equivalent to the obstacle height close to the obstacle, but varies with height above ground according to the vertical structure of the gravity wave.

- *l. 168 – 32.1 K - Suggest consistency about the temperature units (C instead of K).*

  Reply: Sorry for the confusion. We have carefully checked (and corrected) the temperature units in the manuscript.

- *l. 173 – Eq. 2 I do not find consistency in the definition of the different z ranges.*

  Reply: corrected.

- *l. 175 – (nothing to revise here) – Figure 9 is very nice.*

  Reply: Thank you very much :)

- *l. 201-202 – What is the parametrization for ice hydrometeor fall velocity used in UM and KiD? This may have a substantial impact on the results presented below.*

  Reply: Both models use a (the same) relation between ice diameter and fallspeed. The equations and parameters are now provided in the CASIM description. We added also a comment on the sensitivity of sedimentation fluxes to the parameterisation of fall velocities.

- *l. 242 – I presume "basis" should be "bias"*

Reply: corrected.

- *l. 252-253 - If the correction is made for 0.02 g/kg, what is the importance of 0.0001 g/kg in fig. 1?*

  Reply: Thanks for spotting this inconsistency. The new Fig.1 shows now only qi / ql larger than 0.02 g/kg.

- *l. 282 - I can only see some sort of an ni agreement in fig. 5f. I can't interpret the max ni intersection in panel e as model-observations agreement. Using that parameter as a measure of model performance could be misleading.*

  Reply: We added a sentence to better describe the agreement (or lack thereof) in ice crystal number concentrations

- *l. 321 - "data .. is . . ." - "Data" is the plural form of "datum" - please correct the text and figure captions accordingly (is –> are, etc.)*

  Reply: corrected.

- *l. 330 - best agreement with DM10 and TB13 - how do you define the best agreement? In some aspects (e.g., absolute ni values), the log scale in fig. 6 is misleading because I would suggest that the best agreement is with DM10 and DM15.*

  Reply: You are right. We have expanded the text on this issue and also include the DM15 parameterisation.

- *l. 365 - 0.1 g/kg - these are mixing ratio units, not flux units. I suggest providing a more consistent terminology throughout the text (also related to the conceptual model).*

  Reply: Yes, you are right. These are the Lagrangian integral of the sedimentation fluxes. We have changed the terminology everywhere in the manuscript (using instead: total downward moisture transport).

- *l. 403 - a minus sign is missing for the temperature*

  Reply: corrected

- *l. 406 - I suggest adding a reminder to the reader about which two simulations are discussed here.*

  Reply: We are here not referring to two simulation in particular. For each wave period, cloud thickness, and cloud top temperature combination there are 20 simulations with different settings in the cloud microphysics. We compute the difference between any combination of these 20 simulations and show the maximum difference from these in Fig. 10b. We have rephrased the sentence to make this clearer.

- *l. 445-446 - I presume that the definition of Gpot is the integration of qv minus qs(Tmin), yes? Please clarify, or alternatively, describe eq. 4 earlier in the text.*

  Reply: Gpot is qv -qs (Tmin), no integration needed. We have added a sentence at the beginning of the paragraph to make this clear.

- *l. 473 " –saturation)" - redundant parenthesis*

  Reply: corrected.

- *l. 479 – inline equation - not all terms in this equation are defined.*

  Reply: added.

- *l. 492 – pre-scribed –> prescribed*

  Reply: corrected.

- *l. 500-501 & Fig. 13 - I suggest adding a panel that shows the difference between the conceptual model and the KiD model. The discrepancies I see in Figure 5 in the SI (especially in panels c and d, which are shown in log scale) suggest that the differences can be quite significant when the time period and Tct parameter space are examined. Also, shouldn't there be consistency re-*

garding the discussion/use of total water (qt) and water vapor (qv) throughout the text in figures (7-13)?

Reply: The relative difference between the conceptual model and the KiD model are shown in Fig. 13b. Errors are up to 30% of the downward flux. Also, we checked the use of total water and water vapour for consistency.

- *l. 504 - I presume that 'b' is missing when referring to fig. 13.*

  Reply: yes, corrected.

- *l. 538 - Except for the conceptual model, I did not encounter any discussion about the results based on varying cloud thickness. I suggest adding some information to the text and figures, or removing it from this discussion about the conclusions.*

  Reply: The KiD results are only shown for cloud thickness of 2 km, which roughly corresponds to the ICE-L cloud. This is mentioned in the figure captions. We added a reference to the cloud thickness, when discussing the Fig. 10.

- *Fig. 1 - Suggest changing the cyan, blue, and gray curve colors - they mixed with the colormap. Also, the units in the colorbar are confusing here. Also, What altitude do the black contours represent? This should be specified.*

  Reply: todo

- *Fig. 1 and discussion in the text - Is 10ˆ-7 kg/kg above the aircraft instrumentation uncertainty level for IWC and LWC? This should be discussed and justified in the text, i.e., what is the "true" extent of this cloud field given measurable justified quantities?*

  Reply: Thanks for spotting this inconsistency. The new Fig.1 shows now only qi / ql larger than 0.02 g/kg.

- *Fig. 2 - I suggest redefining the altitudes for each flight leg*

  Reply: Not sure what you mean here? Based on the comments from reviewer 2, we have added labels (A, B, C) to the different flight legs, which are used throughout the text and the altitude of which is specified in the text and the Fig. 2 caption.

- *Fig. 4 - I'm having a tough time reading the axes labels*

  Reply: enlarged.

- *Fig. 3-5 - please provide a title for each panel stating the flight leg and/or altitude. At the moment, it is hard to follow the text.*

  Reply: done.

- *Fig. 5-6 – suggest adding a legend instead of directing the reader to fig. 4 every time.*

  Reply: done

- *Fig. 7 - blue for positive values in panels a,b is counter-intuitive. I suggest flipping this colormap.*

  Reply: We find the colourmap intuitive as is, as positive values mean moistening (blue) and negative values drying of the air parcels (red). We therefore decide not to implement this suggestion.

- *Fig. 9 - Please correct qv–>qt in the legend*

  Reply: done

- *Fig. 10 - The figure caption is not complete, e.g., last sentence, circle markers, the definition of the two simulations, etc. The varying contour colors in panels b,d are quite confusing.*

  Reply: We reformulated and added additional information to the figure caption. We have deliberately chosen different colour scales to alert the reader that one column of plots (a,c) are showing absolute (mean) values, whereas the right column plots (b,d) are shown the spread of the mean values. As this was intentional and we still believe using different colormaps is meaningful, we refrain from implementing the change in colourmap.