**Response to Anonymous Referee #1**

We'd like to express our gratitude to the reviewer for their insightful review and we believe that the revised paper is significantly improved thanks to their comments and suggestions.

This manuscript explores a very interesting topic with many outstanding questions, namely the controls on mixed phase hydrometeor properties in deep convection. A large number of sensitivity simulations are performed of an MCS case near Darwin, Australia, for which there are ground radar and aircraft in situ measurements for comparison. Given the unique observational dataset and important topic, the manuscript is certainly worthy of eventually being published in ACP, but there are many major issues that need to be addressed before this can happen, and it will likely take the authors a long time to address all of these issues in a satisfactory manner. The most important of these issues are the flawed methodology for comparing very limited observations with very ample model output and the large amount of unsubstantiated assertions that are passed off as conclusions without evidence. These and other issues are discussed in much more detail below, and a number of suggestions are offered that provide paths forward to overcoming the major shortcomings of the manuscript.

Major Comments

1. The comparison of a single sounding with the model sounding is nowhere close to representative of environmental differences between the model and observations. In fact, the observed sounding is a classic "onion" sounding in a stratiform region where the low level air is completely stable and mid level air is dried out because of the mesoscale downdraft. This is not the air that is feeding the system (it thermodynamically cannot be since it is stable), and convective cloud base from lifted boundary layer parcels would be below 1 km, as it nearly always is in Darwin active monsoon conditions. In the stratiform region, where these soundings are taken, the cloud base is typically around the melting level, which is where the soundings approximately show it. The humidity profile will vary depending on where you take the sounding in the stratiform region, so you also cannot draw conclusions about upper tropospheric humidity. The likelihood that the model sounding is in a stratiform region location that is exactly like the one observed is practically zero, so no conclusions regarding model environmental biases can be drawn from this comparison. The winds are also not representative and examination of CPOL radial velocity shows that low-mid level winds are quite variables because of the MCS forming in a trough convergence region and the mesoscale circulations induced by the stratiform precipitation. Therefore, you should remove all conclusions based on comparison of these soundings. The prior Darwin sounding at 12Z (attached as Fig. 1) before the system initiates shows a classically active monsoon environment and one that is probably similar to the one that the convection develops in 6 hours later, so you can compare that to the model, but it is still not okay to draw conclusions about model environmental biases from one sounding because humidity, winds, and instability are highly variable across mesoscale domains (you can prove this to yourself by plotting them using the model output), so if you choose to include a comparison of 12Z soundings, you should plot a spread of model soundings outside of clouds and precipitation and place the observed sounding in this spread. If the spread covers the one observed sounding, you cannot conclude that there are biases in model environmental representation. Otherwise, simply remove the comparison of observed and modeled soundings. There could be environmental representation biases, but it is nearly impossible to show that given the available observations, and this is not the purpose of the manuscript anyway.

The comparison of the model with the sounding has been removed.

2. The timing and location of this observed MCS are very important for how it needs to be compared to model output. After the initial deep convective stage when convection is most intense, a large stratiform region forms and the most intense convective cells push westward outside of CPOL coverage before the aircraft even begins sampling the system. The aircraft then samples the remnant stratiform region and weak convection that is not representative of the convection that forms the MCS. It is unlikely that the simulations reproduced this lifecycle (in fact Figure 3 shows that they did not), but this lifecycle strongly impacts interpretation of comparisons between model output and observed reflectivity and aircraft observations in some of the figures (i.e., is some of the model error because of a different system evolution in terms of timing and location?). Therefore, the figures showing statistical comparisons would be greatly aided by showing observed and simulated (just pick a representative simulation – the time series show that they have similar evolutions) low level reflectivity during a couple times between the initial intense convection and the decaying stages when the aircraft was making observations.

Additional figures and discussion have been included that describe the plan view of OLR for the observations and control model, as well as the 2.5 km radar reflectivity fields from the radar and control simulation.

The added text for the OLR reads: Comparison of the modelled outgoing longwave radiation (OLR) with the satellite observations in Figure 2 show that in general, the control simulation represents the lifecycle of the MCS fairly well. The location of the mostly oceanic convective cells look reasonable, however, the modelled MCS is larger and composed of more numerous and deeper convective clouds than what was observed in the pixel level satellite OLR data and seen in the low level radar reflectivity fields shown in Figure 3. The model also produces more convection over the Tiwi Islands than what was observed at 17:30 UTC. As the MCS transitions from a developing-mature system through to a mature-decaying system, the observed reduction of deep convective cells with time is simulated, although the OLR remains significantly underestimated. During the research flight at 23:30 UTC, the modelled MCS shows cloud positioned in a similar location to that observed with respect to the MCS structure, however, the modelled cloud is shifted somewhat to the northeast.

3. In the paragraph that starts at the bottom of page 9 and continues onto page 10, I disagree with your reasoning that the underestimate in precipitation at later times is a result of drier low-mid levels. First, you can't determine whether they are drier or not given the available observations, and second, Figure 3 shows that the simulated MCS develops about 2 hours earlier than the observed one. If you shift the simulated precipitation time series to 2 hours later, then the evolution of the precipitation is very similar in the simulations and observations. In the second part of this paragraph, you state that lack of stratiform rainfall is not caused by excessive evaporation (even though earlier in the paragraph you partly blame drier low-mid level air) and instead blame overly strong convection that detrains too high in the troposphere. This could be going on, but you show no evidence of low biased stratiform rainfall or overly strong convection, so this is purely speculation and should be removed unless you add evidence to support it.

The references to the moisture bias have been removed in accordance with comment 1. With respect to the evolution of the simulated MCS see the response to the point above. The additional figures of the plan views of radar reflectivity and OLR support the results that the model produces overly strong convection that detrains too high, and the lack of stratiform rainfall is evident in the radar reflectivity figures.

4. In Figure 3d, the satellite retrieved OLR looks incorrect. I checked the satellite observations between 18 and 21Z and they show OLR less than 125 W m-2 covering the

entire domain (see attached Fig. 2 for 21Z OLR), whereas your figure shows 160 W m-2. Therefore, your conclusions on page 11, lines 16-24 are incorrect. Perhaps you are averaging over too large of a region for the comparison?

This figure has been removed and has been replaced by the plan views of the higher resolution OLR observations (that you showed in your review, rather than the coarse resolution observations that were used) at 4 different times.

5. Because so many model sensitivities are examined, I don't think that any single sensitivity is given the detail that it deserves to understand the mechanisms behind changes in model output. This leads to a lot of speculation throughout the manuscript without much evidence shown. Some speculation is fine, but the speculation is passed off as facts in a number of spots including in the conclusions. Here is a list of examples:
a. On page 10, lines 26-31, your explanation regarding differences in RH profiles between simulations with different ice PSDs may be reasonable, but you do not provide evidence showing riming rates connected to latent heating connected to convective updraft strength.
Unfortunately the model output is not available to analyse the latent heating generated from the riming rates. As such the discussion here has been revised to read:
The higher RH in the simulations using the generic ice PSD could be due to the larger, faster falling particles in the levels below 12 km removing more of the LWC via riming, which would allow for greater supersaturation. More riming would release more latent heat, which along with the larger ice particles being more effectively off-loaded, could lead to the generation of stronger updrafts with less entrainment and higher RH in the upper troposphere.
b. On Pages 10-11, you discuss convective entrainment but you are showing domain mean horizontal mass divergence in Figure 5, which incorporates all regions (convective, stratiform, neither) so you can't assume that differences in mass divergence profiles are related to convection alone. Furthermore, more than entrainment impacts convection. The location of the convection (differing surrounding environment) and the low level convective forcing influence the strength of the convective updrafts, and mass divergence incorporates all updrafts and more, so one simulation can simply have more updrafts reaching a certain height level than another simulation, but the entrainment and strength characteristics of the updrafts may be the same. To claim what you claim, you'd have to isolate convective updrafts (perhaps by using a vertical velocity threshold) and compute their buoyancy and detrainment. I also don't understand your argument on lines 3-6. Why would simulations that have the least mass divergence at upper levels be consistent with updrafts that penetrate higher and higher mean cloud tops?
The horizontal mass divergence figure has been revised to show the mass divergence for the convective updrafts with vertical velocity > 1 m $s^{-1}$. The key results shown do not change. Included in this figure are additional panels that show the convective updraft buoyancy plotted as a function of equivalent potential temperature. These figures support the results deduced from the horizontal mass divergence: greater turbulent mixing at 6 km produces many more occurrences of convective updrafts with reduced equivalent potential temperature, indicative of increased entrainment, and; at 14 km a simulation with smaller ice particle sizes shows considerably fewer occurrences of high equivalent potential temperature, indicative of greater entrainment. Further to this, the figure also includes the histograms of convective updraft buoyancy that show a greater number of occurrences of more positively buoyant clouds at 14 km for the simulations that have larger sized ice particles, supporting the result that less mass divergence represents less entrainment with more positively buoyant updrafts that penetrate higher. This additional reasoning has been added to the manuscript. See the response to comment 5d below about the analysis of environment differences.
c. On page 18, lines 12-14, differences in entrainment and water loading may impact the convective updraft strength and max reflectivity profile, but this is speculation and the correlation between lines in Figure 11c and Figure 10b is far from perfect. To show this, you

could plot these variables vs. one another to provide evidence. Another cause of simulation differences includes possible differences in the positioning and/or timing of convection. For example, for 17-18 UTC, the max reflectivity profile comparison looks quite different than for 23-24 UTC. If entrainment and water loading buoyancy differences caused by the turbulence or microphysics parameterizations are primarily controlling updraft strength and max reflectivity, then why is this the case?

This discussion is focussed on explaining the differences between 3 simulations, not all simulations, and these 3 simulations show a correlation between maximum reflectivity profiles and maximum vertical motion. These 3 cases all use the same ice PSD and only differ in their dynamical and turbulence parameterisations. The comment regarding entrainment and water loading was described to be the “likely” reason and is supported by the results in Figures 5 (see response to comment c above) and the accumulated water contents, as described in the text.

See responses to comment d below (for differences in environment) and minor comment 18 (for differences in max dBZ at 17 – 18 UTC).

d. On page 18, lines 27-29, how do you know extra latent heating is occurring without compensation by entrainment or water loading in the ENDGame simulation? Latent heating is one component of buoyancy, but the environment could also be different.

Analysing the vertically integrated moist static energy for the simulations across the time period 12 – 24 UTC, shows that the large scale environment is very similar across all of the simulations with the differences being < 0.8 K (when normalised by the specific heat capacity of air). The precipitable water differences are also small, around 1 mm, demonstrating that environment changes are unlikely to be responsible for the differences seen. However, since there could be a contribution, the sentence has been modified to read:

In this simulation it seems as though the stronger and deeper updrafts are able to generate enough latent heating that this effect on buoyancy is larger than that of entrainment and water loading as compared to the other cases.

e. On page 21, lines 4-6, why can’t increases in IWC with vertical velocity be the result of higher vertical velocities lofting more condensate upward?

This sentence explains why there is an increase of ice in this temperature regime, as compared to the warmer regimes where the IWC does not increase with vertical velocity. Since all regimes have advection of ice, the difference is caused by the heterogeneous freezing that occurs in this regime and not the others. The sentence has been revised to clarify this.

f. On page 21, lines 17-22, how can you draw any conclusion regarding change in IWC with height in observations with so few samples? If you look data from all of the flights and RASTA, they would disprove this result. Furthermore, where do the simulations support the drop in IWC between -20 to -10_C and -30 to -20_C? The distributions for both temperature regimes look very similar.

The observations from all of the Darwin flights have been added to this figure. The results also show a general trend to reduce the IWC for a given vertical velocity for the coldest regime analysed, but as with the simulations, the reduction is subtle. Because of this the discussion has been deleted.

g. On page 22, lines 23-25, why do you bring up the aerosol invigoration effect if your figures do not support it? For example, Figure 11c shows weaker max vertical velocities when cloud droplet number concentration is increased while Figure 16 shows that total ice mass is not changed.

This has been deleted.

h. On page 23, lines 13-16, I don’t see a change in 90th percentile cloud vertical velocity in Figure 11, but they aren’t as relevant as convective vertical velocity anyway, since it is in convective updrafts (not reflected in 90th percentile cloud upward motion of 0.2 m/s) where Hallett-Mossop is operating. If you examine the max vertical velocity in Figure 11, which is convective, it shows a decrease in vertical velocity by including Hallet-Mossop. Also, how do you know that including Hallett-Mossop increases latent heating? Can you show this?

This sentence has been deleted.

i. On page 24, lines 19-21, you claim that a bimodal PSD representation or a larger observational dataset to generate a more applicable PSD parameterization that correctly represents snow sizes. This is not necessarily true, and I don't see any evidence presented that the two modes of the ice size distribution are important to represent. In fact, the simulated ice size distribution is already bimodal or trimodal because of 2-3 separate ice categories. The fact is that a single-moment scheme will always struggle if it has to represent convective regions dominated by small ice particles and stratiform regions dominated by aggregating large ice particles. This instead suggests that a two-moment scheme that predicts number concentration in addition to mass is needed, and even then, as you show in the manuscript, microphysical and turbulent processes need to be properly parameterized as well, since they impact the predicted PSD moments that define the PSD.

The mention of the bimodal PSD has been deleted. Instead the text is modified to discuss the better ability of double moment microphysics schemes to represent the observed PSD variability, as suggested.

j. On page 26, lines 4-12, you say that you show convective updraft buoyancy, but you don't show this or latent heating in the manuscript. Everything related to convective buoyancy and entrainment/detrainment is speculation.

See response to comment 5b.

k. On page 26, lines 15-23, you don't have a figure where it is possible to discern the slope of reflectivity above the melting level. This is not shown by Figure 6, which shows that the coverage of different reflectivity thresholds is different in simulations and observations, but doesn't show profiles of reflectivity. Furthermore, the slope of snow mean size in Figure 4c looks similar in observations and simulations using the generic PSD and the difference in diameters for 0.5 g m-3 in Figure 17 is not robust and strongly affected by very few observation samples between 0 and -5_C. So overall, I don't see a lot of evidence that implicit aggregation based on the shifting temperature- dependent PSD is too weak.

This discussion has been removed.

l. Of your 4 listed model shortcomings on page 28, "too much rain above the freezing level", "too little entrainment", "increases the stratiform cloud and rain area", and "too efficient depositional growth" are all statements that are not supported by any evidence shown. They are speculation for explaining the figures that you show, but they are not the only possible explanations for the figures that you show.

The depositional growth statement has been removed based on comment 8 below.

With respect to the model having too much rain above the freezing level, this is shown in the comparison of the observed radar reflectivity fractional area coverages with the control model. The > 40 dBZ areas in the model (that are not seen in the observations) are almost exclusively due to rain, as confirmed by producing the same figure when the only hydrometeor category used is rain. The aircraft observations also support the lack of supercooled water, which is produced by both cloud water and rain in the model at the times when the aircraft flew.

We agree with the point about too little entrainment. This sentence has been revised to read: Too little stratiform rain area is increased with increased turbulent mixing.
An additional row of panels is now included in the reflectivity fractional area coverages figure for the simulation that has increased turbulent mixing. This shows an increase in the stratiform cloud and rain compared to the control simulation.

6. By heterogeneous rain freezing, do you mean heterogeneous nucleation by ice nuclei or all freezing mechanisms other than homogeneous freezing? This is unclear in the text. You state that because including heterogeneous rain freezing produces better agreement between observations and simulations, it must be important in tropical convective cloud systems (e.g., page 15, lines 11-12), but the simulation including heterogeneous rain

freezing only slightly improves on the simulation without it, getting nowhere near observations. With such a difference between the simulation and observations, can you confidently trust that a change in the model is reflective of a change in the real world? For example, what if real tropical convective updrafts loft fewer raindrops than the model does for a given updraft strength. Then the effect of heterogeneous rain freezing in the model will have a larger impact than in real life.

The text has been revised to clarify that the heterogeneous rain freezing is heterogeneous nucleation by ice nuclei.

We agree that there is no way to definitively conclude from these simulations that the effects of the addition of this process are expressed in the model in the same way as they are in the real world. That is why the statement that you refer to suggested, rather than concluded, that this process is important. We have added the caveat here that reads:  However, given the errors in the dynamics and microphysics in the model for this case, further study is required to better understand the effects of this process.

7. The discussion about cloud base on page 19 is incorrect since the inferred cloud base from the stratiform sounding (as discussed in point #1) is incorrect, so I suggest removing this discussion. Cloud base for rising low level air is certainly not 3 km. The argument in lines 15-17 does not make sense to me either. Latent heating by condensation can make air buoyant, but only if this heating makes the air warmer than the environment, which is never guaranteed. Buoyancy accelerates air, so vertical velocity is a function of vertically integrated buoyancy. Therefore, any peak in updraft strength will occur at higher altitudes than peak buoyancy and peak buoyancy is often offset from peak latent heating. In this paragraph and later discussions in the manuscript referencing Figure 11, there is also confusing wording equating in-cloud upward vertical velocity with convective updraft vertical velocity. These are not the same. The 90th percentile upward vertical velocity in Figure 11e is _ 0.2 m/s, which can easily be achieved in many non-convective cloud types. To confine your analysis to convective updrafts would require some minimum threshold vertical velocity of 1-2 m/s.

The cloud base and associated buoyancy discussions have been removed. The later references to the Figure 11 percentiles and convective updrafts have been deleted.

8. Be careful interpreting aircraft humidity measurements in convective updrafts. Such measurements can and often do have large errors. Because of this and the small number of updraft samples biasing any statistical comparison, I would not trust any of your conclusions in the second paragraph on page 23.

Based on this comment we analysed the RH observations from all of the Darwin flights. This analysis confirmed that there are erroneous observations and, therefore, this figure and discussion have been removed.

9. Your reasoning on page 24, lines 10-15, doesn’t make sense to me. For the generic ice PSD, if mean sizes are overestimated for IWC > 0.5 g m-3, that means that this PSD has larger concentrations of large particles than observed, not smaller as is stated. This is the only way that mean sizes can be larger for a given IWC.

The sentence has been revised as suggested.

10. The overall text could be shortened and streamlined. It reads like a “stream of consciousness” at times, which makes finding the key points difficult. This is particularly true because of the large number of sensitivity simulations that you want to describe. I

recommend cutting out minor points so that the readers do not get so easily distracted away from the key points. One way to do this is to simply focus on the couple of model component changes that create the biggest effects for whatever variable you are examining. This would also free up space to show evidence supporting your theories (as listed in point #5) for why these specific changes cause the observed effects. You could also cut out some of the simulations if they don't make much of a difference and just say that they don't make a difference. This would unclutter the plots.

The text describing the simulation results has been significantly reduced to focus on the key points. We decided to leave all of the simulations in the figures so that the interested reader can examine the results for each of the cases tested. Due to the addition of more detailed descriptions of the observations and previous studies (comment from the other reviewer), the overall length of the paper has reduced by 2 pages and 3 figures.

11. For comparisons between model output using 1-km grid spacing and 1-Hz aircraft observations (_150-m sampling), do you average the aircraft observations to a 1-km grid before making comparisons? If not, please do this and include this information in the manuscript. Also include information for how the vertical velocity is retrieved from aircraft measurements, how water vapor is subtracted out of IKP evaporator probe measurements, and why IKP retrievals are assumed to be IWC rather that TWC (a combination of liquid and ice). If they are rather used as TWC, then making comparisons to simulated TWC (IWC + LWC) would potentially change some of the conclusions in the manuscript.

All of the observations are averaged to a 1 km grid before any analysis. The following text has been added to the paper in the section describing the observations:

Since the IKP-2 measures the total water content, liquid water and water vapour contributions should be subtracted to obtain IWC. Unfortunately, the hot-wire LWC sensor on the aircraft was unable to measure LWC below about 10% of the IWC in mixed phase conditions, and LWC levels exceeding this value were very rare. Fortunately the Goodrich Ice Detector could be used to detect the presence of liquid water. Two such regions in two very short flight segments for this case, research flight 23, were identified at -10°C, and these regions have been excluded from the analysis. The minimum detectable IWC of the IKP-2 is determined by the noise level of the water vapour measurements of the IKP-2 and background probes. This resulting noise level of the subtraction of the background humidity from the IKP-2 humidity is a function of temperature: it is about 0.1 gm-3 at -10°C, dropping rapidly to about 0.005 gm-3 at -50°C. Since most data were taken at temperatures colder than about -25°C, a minimum IWC of 0.05 gm-3 was chosen as the threshold to include in our analysis.

Two sources of vertical velocity are used from the Falcon 20. Position, orientation and speed of the aircraft are measured by a GPS-coupled Inertial Navigation System. The 3-D air motion vector relative to the aircraft is measured by Rosemount 1221 differential pressures transducer connected to a Rosemount 858 flow angle sensor mounted at the tip of the boom, ahead of the aircraft, and by a pitot tube which is part of the standard equipment of the aircraft. Wind in local geographical coordinates is computed as the sum of the air speed vector relative to the aircraft, and the aircraft velocity vector relative to the ground. Both computations use classical formulas in the airborne measurement field described in Bange et al. (2013). The other vertical air velocity measurement used is retrieved from the multi-beam cloud radar observations using the 3D wind retrieval technique described in Protat and Zawadzki (1999), and we use the technique described in Protat and Williams (2011) to separate terminal fall speed and vertical air velocity. Comparisons near flight altitude with the aircraft in-situ vertical velocity measurements show that the vertical velocity retrieval is accurate to within 0.3 m s-1. All observations are averaged to the model 1 km grid.

We also note that the significant overestimate of IWC by the model means that whether the aircraft IWC is taken as IWC or TWC will not change the conclusions from the model-aircraft comparisons.

12. The comparisons of model output with aircraft observations are not robust because of the low observational sample size in updrafts and downdrafts (e.g., Figures 11c, 12, 15, 17). In fact, the aircraft only penetrated 4 updraft cores at -12_C, 1 at -18_C, and then flew through the edges of a few others around -25_C. You admit as much in a few places in the manuscript, but then attempt to draw conclusions from the comparison about which simulations are most realistic, which isn't possible in convective updrafts or downdrafts for this case alone. Therefore, the plots with these comparisons are not appropriate since the model output is a mean relationship with many samples (essentially a population mean) while the observations are but one, likely unrepresentative sample. There are two ways that this issue can be corrected: a. Include aircraft data from the other field campaign flights to make the sample size more robust. These are different cases, but the sampling for this one case is already biased anyway as mentioned in point #2. Furthermore, the aircraft avoided cells with lightning (the most intense cells) in all cases and the most intense cells in this 18 February case had plenty of lightning, so no matter what, the aircraft is always sampling convection in all flights that is weaker than the most intense convection in this case. Furthermore, as your coauthors know, there are RASTA W-band radar retrievals of vertical velocity and IWC that can be used at temperatures colder than -20_C and would increase the observational sample size to make comparisons with model output more robust. b. Sample the model output with pseudo-flight tracks (E-W or N-S is fine) and limit the total sample size to the same as that observed. Do this a number of times to get a population of samples that are each directly comparable to the observed sample. Then the observed sample can be compared to the distribution of samples drawn from the model to see if it fits into the model spread or not. If it does, you cannot say that the model is wrong. If it doesn't, then you can say that the simulation and observations are different. Without this method, any conclusions drawn on the difference between the model output and aircraft observations are unfounded.

As suggested, the model and aircraft comparisons now include the observations from all of the Darwin research flights. The RASTA derived vertical velocity has also been used.
Additional text has been added to the beginning of the section comparing the simulations to the aircraft. It reads:
Due to the small sample size of observations from the single research flight on 18/02/2014, the observations from 18 of the Darwin HIWC flights have been used to allow for a more robust comparison of the model to the observations (Fig. 12 and 14). The majority of the flight time for these cases was in clouds with temperatures < -10 $^{0}$C and vertical motions within the range of -2 to 2 m s-1. Therefore, when comparing the model to the aircraft observations the focus is on this subset of cloud conditions as there are limited observational samples outside of these ranges.

The text describing the comparison of the simulations to the aircraft observations has been modified accordingly, but we note that apart from the increasing IWC in the downdrafts, the main conclusions have not changed.

13. You restate many of the results in the conclusions section making it rather long (4 pages). I suggest cutting much of this repetitive text out and focusing on key general points like you attempt to do at the very end of the conclusions section.

The conclusions section has been almost halved and now focusses on the general key points.

Minor Comments

1. On Page 6, lines 17-18, you say that graupel formation does not including freezing rain. Do you mean heterogeneous freezing of rain by ice nuclei? Surely, if a raindrop homogeneously freezes or freezes through contact with an ice particle, it should go in the graupel category, no?

This has been revised to read heterogeneous freezing of rain by ice nuclei.

2. On the bottom of page 7, you should also note whether the particle probes have anti-shattering tips or not.

The use of anti-shattering tips has been added to this discussion.

3. On page 8, line 11, you should note the resolution of the peak ice water content since ice water contents strongly depend on resolution.

The resolution of 1 s has been added.

4. On page 8, lines 24-26: The problems with moisture related to domain size are related to periodic lateral boundaries, but you use a nested simulation where moisture can leave the innermost domains, so I'm unsure as to why this discussion is relevant. As I note in major comment #1 though, your conclusion that the model has a moisture bias is not robust because the soundings are not representative, so I would remove all discussion of it or replace it with the comparison I suggest.

This discussion has been removed.

5. For your comparisons in Section 3.1, please state whether you are using the full model domain or the CPOL domain defined by the range ring in Figure 1 to calculate model domain mean quantities.

Text has been added to specify that these comparisons use the radar domain.

6. On page 9, lines 21-23: I'm not sure why you cite Fridlind et al. (2012) here to say that the simulated domain mean precipitation rate is outside of the radar-derived precipitation rate range of uncertainty. You also don't show the uncertainty range. If you examined that, why not show it using vertical bars in Figure 3a?

The uncertainty referred to here is the uncertainty of the rainfall retrieval that considers things like the sensitivity of the radar and calibration issues.

7. On page 10, line 23, in Figure 4, and throughout the manuscript, when you say "mean ice particle sizes", how are mean sizes calculated? Are these mass-weighted mean diameters or something else? Please clarify this throughout the manuscript.

The only measure of mean size used is the mass weighted mean diameter. This has been clarified here and elsewhere.

8. On page 11 and for Figure 3, how do you define cloud top in simulations?

The figure of cloud top heights has been removed.

9. On page 12, line 21, 23, and 29: A C-band radar cannot observe cloud top or the fraction of the domain covered by hydrometeors since it is only sensitive to precipitation sized hydrometeors, so clarify this by referring to the reflectivity echo coverage.

Modified as suggested.

10. How can you tell that the control simulation evolves from scattered to more organized convection with stratiform regions from Figure 6? I suggest showing this as I state in major comment #2.

See response to major comment 2.

11. On page 13, lines 27-28, the excess large particles above the freezing level can also be related to insufficient representation of the rain DSD, warm rain processes, and/or rain sedimentation (representation of fall speeds and size of updrafts being too large).

This has been modified to read: The simulated rain above the freezing level that is not observed suggests that the model has faster updrafts than observed, which loft large rain particles upwards and/or the heterogeneous freezing of rain that is not represented in the model is an important process in tropical convection and/or other errors in the representation of the rain DSD.

12. On page 13, line 31: This is true of raindrops and cloud drops, but the lower temperature limit should be 0_C as many raindrops freeze quickly at relatively warm temperatures from contacting entrained ice particles starting at 0_C.

This has been left unchanged as the observational evidence cited has a lower limit of -6 °C.

13. On page 14, lines 16-19, I doubt this is the reason for the non-prominent bright band in observations. It is much more plausible that the radar beam smears the bright band out because this data is taken from volumetric scans and more data is far away from the radar than close to it (because of radar coverage increasing as range ring radius squared). Despite this, you still see a bump at 4 km height corresponding to the bright band.

Thank you for this information. The text has been modified to read: The lack of a predominant bright band in the observations is likely due to the data being collected from volumetric scans, however, there are slightly higher reflectivities seen at 4 km indicating a bright band.

14. On page 15, lines 14-16, single moment schemes typically do increase the number concentration as IWC increases. Aggregation is a decrease in number concentration for no change or an increase in IWC. This can also be diagnostically represented in single moment schemes by altering the PSD as a function of temperature though. For example, the Thompson microphysics scheme (Thompson et al. 2008) commonly produces the best agreement with observed stratiform reflectivity profile above the melting level. Two-moment schemes can explicitly represent aggregation through predicting the number concentration, but also typically overestimate reflectivity aloft because other factors include excessive size sorting, mass-size relationships, and the assumed PSD shape.

This sentence has been deleted.

15. On page 17, line 4, the aircraft observations are mostly in stratiform precipitation (plot the flight track on top of the CPOL reflectivity and you'll see this clearly) even though the aircraft penetrates a few weak deep convective cores. The highest concentrations are found in convective cores, not in stratiform regions, so having convective observations does not make

them lesser than the ones in Field et al. (2007), which also include convective observations. The observations in Field et al. (2007), however, may suffer from ice shattering artifacts, so they may not be directly comparable to these new aircraft observations that mitigate and control for shattering.

With regards to the first part of this comment, the text has been revised to read: The observations in this case may be in a different type of cloud environment from the data used to construct the Field parameterisation, as suggested by the observed number concentration being below the lower range shown in Field et al. (2007).

As was stated, the data used in this comparison was only for particles > 100 microns in diameter to be consistent with the data used to derive the Field et al. (2007) parameterisation. They did this to minimise the effects of shattering. Because of the use of this minimum diameter, the effects of shattering should not significantly bias the comparison.

16. From Fig. 10, it looks like there is an issue in limiting hydrometeor sizes to realistic values in the microphysics scheme you are using. A rain reflectivity of 75 dBZ is physically impossible because raindrops begin breaking apart at large sizes. In the real world, rain reflectivities are limited to less than _55-60 dBZ. Some schemes implement limits on the slope of the rain DSD, and that may need to be done for this scheme.

Thank you for providing this information that is useful for future model development.

17. On page 17, line 18, the observed decrease in max reflectivity above 2 km may also be from raindrops falling through weak updrafts and collecting cloud droplets in the classic warm rain process.

Yes this could also be occurring and has been added to the text.

18. On page 17, lines 22-24: This is true that different subgrid turbulent mixing decreases max reflectivity, but only for 23-24 UTC and not for 17-18 UTC. Why?

Analysing the maximum updrafts at the earlier times shows that the difference between the simulations at this time is much smaller than the later times, and the updrafts are stronger with all simulations showing > 20 m $s^{-1}$ in the upper troposphere. The stronger updrafts allows for very large particles to be advected to the upper levels in all of the simulations resulting in little difference in maximum dBZ at these times.
The text has been modified to read: There is little spread in the maximum reflectivity profile across the simulations at 17 – 18 UTC, with strong updrafts > 20 m s-1 in all simulations (not shown) that allows large particles in all simulations to be advected into the upper troposphere.

19. On page 17, lines 24-27, I can’t clearly see the reduction in max reflectivity caused by implementing the heterogeneous rain freezing parameterization. Perhaps increase the symbol sizes so that the different lines can be seen more clearly.

The figure has been replotted with larger symbol sizes.

20. On page 19, lines 19-21, the upper level vertical velocity peak is also a result of vertical velocity being related to vertically integrated buoyancy. CAPE is usually distributed over a significant depth and the updraft will accelerate as CAPE is used up, primarily being limited by entrainment and opposing pressure gradients. Freezing of condensate and unloading of condensate simply help to push the peak higher.

This sentence has been revised to read:
The upper level updraft peak has been observed (e.g. May and Rajopadhyaya 1999) and is argued to be due to the deep column of convectively available potential energy in the tropics, coupled with latent heat released by freezing condensate and the unloading of hydrometeors, both of which increase parcel buoyancy.

21. On page 20, lines 23-24, you state that the reduction in rain by heterogeneous freezing reduces accretion of cloud water and thus increases the cloud water mass. Why don't the graupel particles formed by the freezing raindrops accrete the cloud water through riming? Is this related to lower cloud droplet collection efficiency by graupel than rain?

Yes thank you for picking up on this, changes between the accretion of rain and riming of graupel due to differences in the size distributions affect the cloud water removal. This has been modified to read: This is due to the reduction in the riming of cloud water by graupel as compared to the accretion of cloud water by rain.

22. On page 20, lines 25-28, how do fast fall speeds of particles help to generate downdrafts? I think of the loading and evaporation, mostly relating to rain in the tropics, as primary drivers. Do fast fall speeds impact loading and evaporation? Also, on lines 28-30, why does more accumulated graupel mass being correlated with the largest IWC in downdrafts support the argument that fast graupel fall speeds generate downdrafts? Do the strongest downdrafts have the most graupel? If so, that would be a supportive argument.

This has been revised to read: … where the suggestion is that these larger particles help to generate downdrafts through mass loading.
Analysing the IWC for the downdrafts in the warmest regime shows that the largest source of ice is indeed graupel. The text has been revised to read: This argument is supported by analysis of the downdraft IWC that shows that the majority of the ice in the downdrafts is graupel. For example in the control simulation, 82% of the ice mass is graupel for the warmest regime downdraft of 5 m $s^{-1}$.

23. On page 22, lines 29-30, I don't see a reduction in total accumulated ice mass in Figure 16. Am I missing something?

This refers to the "accumulated amount of aggregate mass" not the total (aggregate + crystal + graupel) ice mass.

24. On page 25, line 5, you claim that the simulations capture the timing of the deepest convection well, but Figure 3 suggests that the simulations initiate and organization deep convection earlier than observed, as you suggest on lines 9-10.

While the simulations do produce deep convection in the radar domain earlier than observed, the timing of the deepest convection observed at 17 – 18 UTC is also when the greatest amount of deep convection occurs in the simulations, as shown for example in OLR plan views and the statistical radar coverage figure, which shows the more vertically aligned contours in the simulations after 17 UTC. The sentence has been modified to read: Analysing 12 hours of observed and simulated radar reflectivity has shown that the simulations capture the intensification and decay of convective strength associated with the lifecycle of the MCS, with the timing of the greatest amount of deep convection represented well.

25. On page 25, lines 16-19, what is your definition of “large” particles? Reflectivity is more sensitive to large particles than small particles but a large number of small particles can give the same reflectivity as a small number of large particles, so it seems that you are using an arbitrary reflectivity value here to define large vs. small particles.

This sentence has been deleted.

26. On page 25, line 32, and page 26, line 2, you mention the percentiles of updraft speed, but your figure shows 90th percentile cloud upward motion, which isn’t necessarily correlated with max reflectivity since most of the cloud volume is not convective updrafts where the max reflectivities are occurring.

The reference to the $90^{th}$ percentile has been deleted.

27. On page 26, lines 24-25, do you mean that the heterogeneous rain freezing parameterization reduces raindrops above the freezing level rather than reducing the lofting of raindrops? A freezing mechanism shouldn’t impact raindrops lofting above 0_C, right?

This has been modified to read: The beneficial impact of including a rain heterogeneous freezing parameterisation was shown through the reduction of large raindrops above the freezing level, which was not observed by the radar or aircraft and supports previous observations that show that most drops in oceanic convection freeze between -6 and -18 $^{o}$C (Stith et al. 2002).

28. On page 26, lines 26-28, raindrops not being lofted above the freezing level cannot be detected by radar reflectivity and the aircraft was clearly observing the MCS during its decaying stage, not its mature stage, based on the time series shown in Figure 3. Updrafts, even weak ones, commonly loft raindrops above the 0_C level, but it is true that most of them freeze rather quickly. That is different though than what you state here, that raindrops are not lofted above the 0_C level, which is not supportable from available observations.

See the point above.

**Response to Anonymous Referee #2**

We'd like to express our gratitude to the reviewer for their insightful review and we believe that the revised paper is significantly improved thanks to their comments and suggestions.

The authors compare simulations of a tropical MCS observed during a recent airborne field campaign with the in situ measurements between 0 and –40 C, where liquid water and ice could coexist (although there appears to be no liquid in the observations). There is substantial uncertainty as to how MCS updraft microphysical processes operate in nature, and improving process-level knowledge is a worthy research goal within the scope of ACP. Observations from multiple campaign flights have been reported by Leroy et al. (2015), as cited, but this appears to be the first analysis of the relationship of dynamics and microphysics observed during a flight. Overall, I am an interested reader, but I found it difficult to maintain attention on such a long paper for several reasons. First, it appears that the baseline simulation simply does not capture the event well at all, contrary to the authors' claims (in the abstract for instance), and sensitivity tests have similar errors across the board (e.g. Fig. 15). Second, several aspects of the observations appear notably odd (such as large updrafts without any additional ice content), but the authors focus on narrow elements of the observations without explaining why such apparent oddities are present. These factors combined make it difficult to be interested in nearly twenty figures comparing the simulations and observations, and even lead this reader to feel that the sensitivity tests may be futile or ill-conceived because the simulations are so far off the mark. Below I suggest the major steps that could help develop this manuscript in my estimation. Minor comments are then listed in case they are helpful.

Major comments

1. The MCS evolution in observations and simulations needs significantly more description. Highly averaged satellite data in Fig. 3 indicate that there are plenty of images that could be shown to us to see what OLR evolution looks like in the observations and the simulations. The reader needs to see these to understand if the simulated system appears far too large (in addition to being far too cold on top) compared with the observations. Is this a system coming off the ocean in observations and simulations? Is it of a similar size and duration? I would recommend showing OLR images before, during and after the aircraft sampling times used in this paper, both observed and simulated. It feels decidedly odd that these were omitted. This needs to be remedied and re-reviewed.

A timeseries of the enhanced IR imagery has been added, along with plan views of the OLR from the observations and the control simulation at 4 different times throughout the MCS lifecycle. The text describing the MCS has been expanded to read: Comparison of the modelled outgoing longwave radiation (OLR) with the satellite observations in Figure 2 show that in general, the control simulation represents the lifecycle of the MCS fairly well. The location of the mostly oceanic convective cells look reasonable, however, the modelled MCS is larger and composed of more numerous and deeper convective clouds than what was observed in the pixel level satellite OLR data and seen in the low level radar reflectivity fields shown in Figure 3. The model also produces more convection over the Tiwi Islands than what was observed at 17:30 UTC. As the MCS transitions from a developing-mature system through to a mature-decaying system, the observed reduction of deep convective cells with time is simulated, although the OLR remains significantly underestimated. During the research flight time at 23:30 UTC, the modelled MCS shows cloud positioned in a similar location to that observed with respect to the MCS structure, however, the modelled cloud is shifted somewhat to the northeast.

2. It is difficult to continue this review without understanding how the system simulated relates to the system observed in terms of overall shape and top OLR structure. Right now, it appears to me, based on the figures shown, that the observed system is weak (Fig. 11), with low cloud top heights and surprisingly warm OLR (Fig. 3). Is this even an MCS? The simulations on the other hand do look like an MCS in terms of OLR and updraft strengths, but cloud top height seems low to me for a tropical MCS at 12.5 km with hardly any change with time. How is cloud top height defined in the observations and simulations? How do the underlying structure of cloud top heights observed and simulated compare, and what is the uncertainty in differences of definition between observations and simulations?

Apologies for making your reviewing job difficult because of these omissions. Please see the response above and note that we no longer include the cloud top height comparison due to, as you point out, difficulties in consistent definitions between satellite and models. Instead we describe the structure of the OLR as detailed in the point above. Also note that we now use the much higher resolution pixel level OLR observations, rather than the coarse resolution observations. This change shows lower observed OLR of around 120 W $m^{-2}$.

3. I will continue by assuming that the observed system is a small, weak system and the simulated system is a big, strong MCS, as appears to be the case from all indications in Figs. 3 and 11. Moving to the objective of this study, the title of the paper refers to phase composition, but this topic is not clearly explored. Only Figs. 13 and 14 (really one figure together) show liquid water content as a function of updraft velocity, but as far as I can tell there are no measurements of phase and no other analysis of phase.
Can the authors explain why they chose to focus on phase composition and why with this data set and this case study in particular? Also, what is a "high ice water content"? The updrafts shown here seem to have low ice water content. The authors refer to some other papers, but those seem to be focused on radar reflectivity.

In the introduction the description of the aims has been expanded to read:
The aims of this study are twofold: firstly to test different configurations of the dynamics, turbulence and microphysical formulations in the model to determine those that best represent tropical convective cloud systems and to understand the sensitivities in the modelled cloud and dynamical properties to these changes, and; secondly to determine what process control the phase composition and ice water content in the model. As mentioned previously, observations of HIWC (defined here as > 2 g $m^{-3}$ at 1 km resolution) typically occur in glaciated conditions. However, as will be shown, the model is unable to replicate this and instead produces mixed-phase clouds under the same temperature regimes. For this reason we examine what processes control the modelled phase composition in order to understand how the model produces HIWC. This understanding will aid in improving the representation of these clouds in the model and produce a better forecasting capability.

4. The authors seem to view this modeling study as an exercise in manipulating their simplified microphysics (primarily) to better agree with the observations (unsuccessfully I would say) without investigating whether processes are actually likely to be active based on the observations. For instance, the absence of an observed bright band leads to a suggestion that particles are heavily rimed. (I think a tropical MCS should have a bright band, which to me seems another indication that the observed systems is not really an MCS. If the authors had a bright band simulator, I expect the simulated case would have one.) Later, graupel is removed from the model. What do the observed particle images look like? Do they indicate heavy rime? Is graupel observed? Leroy et al. (2015) show particle images, so I assume that they exist for this flight. Please describe what is known about the hydrometeor particles based on the flight data.

With respect to the bright band, the description of the lack of a bright band was in error. Based on the other reviewer's comment the text has been modified to read: The lack of a predominant bright band in the observations is likely due to the data being collected from volumetric scans, however, there are slightly higher reflectivities noticeable at 4 km indicating the presence of a bright band.

A discussion has been added to the section describing the MCS that reports on the presence of graupel and the observed particle images. It reads as:
There was almost no supercooled water detected during the flight, even at -10 $^{o}$C, and graupel was intermittently observed. The absence of supercooled water coupled with the occasional presence of graupel is due to the system being sampled at the mature-decaying stage, where the supercooled water had been consumed in the production of graupel. Most of the time the particle images were of dense ice aggregates at flight level, except within some convective cores where graupel was observed, as also indicated by strong W-band attenuation.

5. Past literature on updraft microphysics seems to be largely ignored, as do particle size distributions themselves. The last sentence of the paper concludes that there is a need to better represent the "observed bimodal ice size distribution" but we are never shown a size distribution in the paper, either observed or simulated. How do we know that either observed or simulated are or are not bimodal and that this is important?

Based on a comment from the other reviewer, the mention of the bimodal size distribution has been deleted. Instead we retain the focus in this paper on the mass-weighted mean diameters and discuss the advantages of using a double moment microphysics scheme in representing the observed PSD variability. We have also added some discussion on updraft microphysics from other studies and note that detailed PSD studies from this campaign are currently underway. The additional text reads:
This contrasts with the lack of dependence of mean ice particle size on IWC that has been observed in earlier flights over Darwin and Cayenne in 2010 – 2012 (Fridlind et al. 2015) but agrees with more recent findings by Leroy et al. (2015). These findings show similar results to those documented by Gayet et al. (2012), with high concentrations of ice crystals occurring in regions of ice water content > 1 g $m^{-3}$ sustained for at least 100 s at Darwin (Leroy et al. 2015) and > 0.3 g $m^{-3}$ in the over shooting convection in the midlatitudes in Western Europe (Gayet et al. 2012). Gayet et al. (2012) proposed that the high concentration of ice crystals that appeared as chain-like aggregates of frozen drops, could be generated by strong updrafts lofting supercooled droplets that freeze homogeneously. However, using updraft parcel model simulations, Ackerman et al. (2015) showed that this process produced a smaller median mass area equivalent diameter than is observed. They proposed a number of other possible microphysical pathways to explain the observations including the Hallett-Mossop process and a large source of heterogeneous ice nuclei coupled with the shattering of water droplets when they freeze.

6. The concern of the authors with model dynamics is likely well founded. Some discussion of past model resolution studies would be helpful. Question: why bother with this exercise if the resolution of this model is too coarse to properly represent the updrafts observed, given that such updrafts are the only location where phase composition is interesting? If the authors do believe that the updrafts are grossly misrepresented dynamically, why spend so much time examining details of what occurs within them microphysically? Do the authors have evidence that this model is adequate to sufficiently represent uprafts being compared with observations? Why should I not conclude that this is the wrong tool to study phase composition in a tropical MCS?

The discussion on past studies of model resolution and the effect on updrafts has been expanded. It reads: …These values are well outside the range of maximum vertical velocities presented for oceanic convection by Heymsfield et al. (2010) and agree with other studies showing excessive tropical vertical velocities simulated by convection permitting models. Hanley et al. (2014) demonstrated that the UM with a grid length of 1.5 km simulated convective cells that were too intense and were initiated too early, as was also shown by Varble et al. (2014a), suggesting that convection is under resolved at grid lengths of order 1 km. Improved initiation time was shown by Hanley et al. (2014) to occur when the grid length was reduced to 500 and 200 m. However, the intensity of the convective cells was not necessarily improved, with the results being case-dependent. Varble et al. (2014a) also showed that in the tropics the intensity of the updrafts remained overestimated even at the 100 m grid length. Both of these studies suggest that there are missing processes in the model and/or the interactions between convective dynamics and microphysics are incorrectly represented.

We also note that recent cloud-resolving model intercomparison studies of tropical convection use a similar horizontal grid length to what is used in this study (e.g. Fridlind et al. 2012; Varble et al. 2014a,b). Some of these recent studies focus on convective updraft properties, which as described in the introduction, is important because these models are used to develop convection parameterisations for coarser resolution models. Therefore, a detailed understanding of how these models represent convective updraft processes is necessary.

7. Throughout the abstract, broad claims are made that are not clearly limited to these particular simulations. For instance, the last sentence of the abstract states that "... the entrainment and buoyancy of the air parcels is controlled by the ice particle sizes, demonstrating the importance of the microphysical processes on the convective dynamics." I think the authors mean in this particular system simulated by their particular model, which does not appear to resemble the system observed as far as I can tell.
The statements made in the abstract need to be more carefully delineated to refer to their particular model with coarse resolution and one-moment microphysics, especially given the apparently poor resemblance of results to observations in almost every way shown (e.g., updrafts, reflectivities, OLR, ice mean size, ice water content, and ice water content versus updraft strength). I credit the authors with showing these myriad flaws of their simulations (that is truly useful), but I would be more interested to see conclusions related to what model factors need to be changed to improve the simulations rather than conclusions about whether ice size controls updraft strength, given the unrealistic nature of the simulations.

The abstract has been revised to read:

Simulations of tropical convection from an operational numerical weather prediction model are evaluated with the focus on the model's ability to simulate the observed high ice water contents associated with the outflow of deep convection, and to investigate the modelled processes that control the phase composition of tropical convective clouds. The 1 km horizontal grid length model that uses a single moment microphysics scheme simulates the intensification and decay of convective strength across the mesoscale convective system. However, deep convection is produced too early, the OLR is underestimated and the areas with reflectivities > 30 dBZ are overestimated due to too much rain above the freezing level, stronger updrafts and larger particle sizes in the model. The inclusion of a heterogeneous rain freezing parameterisation and the use of different ice size distributions show better agreement with the observed reflectivity distributions, however, this simulation still produces a broader profile with many high reflectivity outliers demonstrating the greater occurrence of convective cells in the simulations. Examining the phase composition shows that the amount of liquid and ice in the modelled convective updrafts is controlled by: the size of the ice particles, with larger particles growing more efficiently through riming, producing larger IWC; the efficiency of the warm rain process, with greater cloud water contents being available to support larger ice growth rates, and; exclusion or limitation of graupel growth, with more mass contained in slower falling snow particles resulting in an increase of in-cloud residence times and more efficient removal of LWC. In this simulated case using a 1 km grid length model, horizontal mass divergence in the mixed-phase regions of convective updrafts is most sensitive to the turbulence formulation. Greater mixing of environmental air into cloudy updrafts in the region of -30 to 0 degrees Celsius produces more mass divergence indicative of greater entrainment, which generates a larger stratiform rain area. Above these levels in the purely ice region of the simulated updrafts, the convective updraft buoyancy is controlled by the ice particle sizes, demonstrating the importance of the microphysical processes on the convective dynamics in this simulated case study using a single moment microphysics scheme. The single moment microphysics scheme in the model is unable to simulate the observed reduction of mean mass-weighted ice diameter as the ice water content increases. The inability of the model to represent the observed variability of the ice size distribution would be improved with the use of a double moment microphysics scheme.

Minor comments

1. Page 8, line 14: How well are cloud bases observed by satellite? Cloud base throughout this system is at 3 km? That seems quite high to me for a tropical MCS.
Over ocean?

This paragraph has been deleted based on comments from the other reviewer.

2. Page 9, line 32: CloudSat IWP uncertainty is less than 25%?

This sentence refers to a comparison that was made between the tropical IWP derived from VISST and that from CloudSat. In the cited study, the comparison showed that VISST derived IWP was underestimated compared to the CloudSat derived IWP by 25%. But we take the point that CloudSat has its own uncertainties and have modified the text to read:
The observed IWP is only valid for the daytime from about 22:30 UTC or 8 am local time, and while the simulations with the generic PSD parameterisation compare well with the satellite derived value, the comparison of VISST IWP with CloudSat in tropical regions was shown by Waliser et al. (2009) to be underestimated by 25%, likely due to the maximum retrieved optical depth being limited to 128. Together with the CloudSat uncertainties (30% bias, 80% root mean square error; Heymsfield et al. 2008), this suggests that the modelled domain mean IWP may be underestimated from 22:30 – 23:30 UTC.

3. Page 11, first paragraph: There is a lot of discussion of divergence and convergence here, but to me the peaks above 15 km in Fig. 5 look like oscillatory gravity waves.
What evidence do the authors have that the peaks in motion above 12 km are not dominated by oscillatory motions?

Analysing vertical velocity profiles of the convective cells shows a smooth profile up to about 16 km, with oscillatory motions above this height. This finding also fits with the PDF of cloud top heights at this time that shows a distinct change in the distribution at 16 km. We note this in the revised manuscript.

4. Page 16, line 1: Both rain and ice appear bimodal to me; could they be related to one another?

Thank you for pointing this out. The text has been revised to state that the PDF is bimodal. Looking at the observed PDF distribution at heights in between 6 and 2.5 km shows that the bimodality does not persist throughout the vertical and, therefore, they do not appear to be related.

5. Figure 15: These observations need some explanation. There is a 10 m/s updraft with less than 90% RHI between –20 and –30 degrees C? Is there a problem with the observations? Fig. 12 shows IWC remaining low to 15 m/s at 0 to –5 degrees C. I think a section should be devoted to noting and explaining such features when these observations are first shown. Are they somehow atypical? Is this strange strong updraft(s) associated with some aspects of the chaotic and odd diameter trends shown in Fig. 17?

Based on this comment we analysed the RH observations from all of the Darwin flights. This analysis confirms that there are erroneous observations and, therefore, this figure and discussion have been removed.

Most of the flight time was at temperatures colder than -10 °C and the limited number of samples affects the results for this temperature range. We now include the results for all of the Darwin flights to increase the sample size. However, there are still not a great deal of observations within this warmest temperature regime and the figure only includes the results of the compositing when there are more than 5 samples. The effect of this is to eliminate the chaotic trends. Additional text has been added to the beginning of this section that reads:

Due to the small sample size of observations from the single research flight on 18/02/2014, the observations from 18 of the Darwin HIWC flights have been used to allow for a more robust comparison of the model to the observations (Fig. 12 and 14). The majority of the flight time for these cases was in clouds with temperatures < -10 $^{o}$C and vertical motions within the range of -2 to 2 m $s^{-1}$. Therefore, when comparing the model to the aircraft observations the focus is on this subset of cloud conditions as there are limited observational samples outside of these ranges.

The text describing the comparison of the simulations to the aircraft observations has been modified accordingly, but we note that apart from the increasing IWC in the downdrafts, the main conclusions have not changed.

6. I found it difficult to follow and maintain interest after the jump from Fig. 12 to Fig. 16 on page 20.

This section has been significantly revised. Figures 12 and 14 are now represented by a single figure and Figures 13 and 15 have been removed. The text has been streamlined throughout to focus more on the key points.

# Controls on phase composition and ice water content in a convection permitting model simulation of a tropical mesoscale convective system

**C.N. Franklin[1], A. Protat[2], D. Leroy[3] and E. Fontaine[3]**

[1]{CSIRO, Aspendale, Victoria, Australia}

[2]{Bureau of Meteorology, Docklands, Victoria, Australia}

[3]{Laboratoire de Meteorologie Physique, Universite Blaise Pascal, Clermont-Ferrand, France}

Correspondence to: C. N. Franklin (charmaine.franklin@csiro.au)

## Abstract

Simulations of tropical convection from an operational numerical weather prediction model are evaluated with the focus on the model's ability to simulate the observed high ice water contents associated with the outflow of deep convection, and to investigate the modelled processes that control the phase composition of tropical convective clouds. The 1 km horizontal grid length model that uses a single moment microphysics ~~The~~scheme simulates the intensification and decay of convective strength across the mesoscale convective system. ~~lifecycle is simulated well,~~ H~~h~~owever, deep convection is produced too early, the OLR is underestimated and the areas with reflectivities > 30 dBZ are overestimated due to too much rain above the freezing level, stronger updrafts and larger particle sizes in the model. The inclusion of a heterogeneous rain freezing parameterisation and the use of different ice size distributions show better agreement with the observed reflectivity distributions, however, this simulation still produces a broader profile with many high reflectivity outliers demonstrating the greater occurrence of convective cells in the simulations. Examining the phase composition shows that the amount of liquid and ice in the modelled convective updrafts is controlled by: the size of the ice particles, with larger particles growing more efficiently through riming, producing larger IWC; the efficiency of the warm rain process, with greater cloud water contents being available to support larger ice growth rates, and; exclusion or limitation of graupel growth, with more mass contained in slower falling snow particles resulting in an increase of in-cloud residence times and more efficient removal of LWC. It is shown that the growth of ice is less dependent on vertical velocity than is liquid water, with the control on liquid water content being the updraft strength due to stronger updrafts having minimal entrainment and higher supersaturations. Larger liquid water contents are produced when cloud droplet number concentrations are increased or when a parameterisation of heterogeneous freezing of rain is included. These changes reduce the efficiency of the warm rain processes in the model generating greater supercooled liquid water contents. The control on ice water content in the model is the ice sizes and available liquid water, with the larger ice particles growing more efficiently via accretion and riming. Limiting or excluding graupel produces larger ice water contents for warmer temperatures due to the greater ice mass contained in slow falling snow particles. This results in longer in-cloud residence times and more efficient removal of liquid water. In It is demonstrated thatthis simulated case using a 1 km grid length model, horizontal mass divergenceentrainment in the mixed-phase regions of convective updrafts is most sensitive to the turbulence formulation in the model. Greater mixing of environmental air into cloudy updrafts in the region of -30 to 0 degrees Celsius produces more mass divergence indicative of greater entrainment, which generates more detrainment at these temperatures and the generation of a larger stratiform stratiform rain area. Above these levels in the purely ice region of the simulated updrafts, the convective updraft entrainment and buoyancy of air parcels is controlled by the ice particle sizes, demonstrating the importance of the microphysical processes on the convective dynamics in this simulated case study using a single moment microphysics scheme. The single moment microphysics scheme in the model is unable to simulate the observed reduction of mean mass-weighted ice diameter as the ice water content increases. The inability of the model to represent the observed variability of the ice size distribution would be improved with the use of a double moment microphysics scheme.

## 1 Introduction

Improving the simulation of tropical convective clouds in convection-permitting simulations is an important yet challenging endeavour. Forecasting centres are beginning to use operational numerical weather prediction models with horizontal grid spacing of order 1 km and while these models have been shown to improve the diurnal cycle of convection and the distribution of rain rates (e.g. Clark et a. 2007; Weusthoff et al. 2010), there are numerous deficiencies at these resolutions that impacts the accuracy of the forecasts and the confidence in using these models to help guide parameterisation development for coarser resolution models and develop retrieval algorithms for remotely sensed cloud properties (e.g. Del Genio and Wu 2010; Shige et al. 2009). One salient aspect of forecasting tropical meteorology is the high ice water contents that are responsible for numerous aircraft safety incidents as discussed by Fridlind et al. (2015). These incidents tend to occur in fully glaciated conditions in the vicinity of deep convection where high ice water contents can cause engine power loss (e.g. Lawson et al. 1998; Mason et al. 2006; Strapp et al. 2015). In recognition of this, an international field campaign called the High Ice Water Content (HIWC) study was conducted out of Darwin in the beginning of 2014 and provided a high quality database of ice cloud measurements associated with deep tropical convective systems. These observations are a valuable resource for evaluating convection permitting model simulations and cloud microphysical parameterisations. In this work cloud properties are evaluated from an operational model with the focus on the model's ability to simulate high ice water contents generated from the outflow of deep convection and to understand what modelled processes control the phase composition of the simulated tropical convective clouds.

Many previous convection permitting simulations of tropical convection have documented common biases amongst models including excessive reflectivities above the freezing level, lack of stratiform cloud and precipitation, and too much frozen condensate (e.g. Blossey et al. 2007; Lang et al. 2011; Fridlind et al. 2012; Varble et al. 2014a,b). Lang et al. (2011) modified a single moment microphysics scheme to reduce the biases in simulated radar reflectivities and ice sizes in convective systems and found better success in a weakly organised continental convective case compared to a stronger oceanic MCS. The reason could be due to dynamical errors in the model that had a greater influence on the microphysical characteristics in the simulations of stronger convection. Varble et al. (2014a) compared cloud resolving and limited area model simulations with the extensive database of observations from the Tropical Warm Pool-International Cloud Experiment. They found excessive vertical velocities even at 100 m horizontal grid spacings, and suggested that the overly intense updrafts are a product of interactions between the convective dynamics and microphysics. These strong updrafts transport condensate and moisture to the upper levels that contributes to the larger amount of frozen condensate seen in simulations, and the reduced detrainment at lower levels could play a role in the lack of generation of significant stratiform cloud and precipitation (Ferrier 1994; Tao et al. 1995; Morrison et al. 2009). In the operational model used in this study the microphysics scheme is a single moment bulk scheme. Model intercomparison studies have shown that double moment microphysics schemes do not necessarily perform better than single moment schemes, and in fact provided that the intercept parameters are not fixed and are able to vary, these more simple schemes can match or even outperform the more complex double moment schemes in their representation of cloud and rainfall properties (e.g. VanWeverberg et al. 2013; Varble et al. 2014b).

The aims of this study are twofold: firstly to test different configurations of the dynamics, turbulence and microphysical formulations in the model to determine those that best represent tropical convective cloud systems and to understand the sensitivities in the modelled cloud and dynamical properties to these changes, and; secondly to determine what process control the phase composition and ice water content in the model. –As mentioned previously, observations of HIWC (defined here as > 2 g m$^{-3}$ at 1 km resolution) typically occur in glaciated conditions. However, as will be shown, the model is unable to replicate this and instead produces mixed-phase clouds under the same temperature regimes. For this reason we examine what processes control the modelled phase composition in order to understand how the model produces HIWC. This understanding will aid in improving the representation of these clouds in the model and produce a better forecasting capability. The following section describes the model and observations used in this work. Section 3 compares the simulations with the available observations including: a time series comparison with the satellite data, comparison of the simulated radar reflectivity characteristics with those from the Darwin radar and an investigation into the controls on phase composition in the model and how the IWC and ice particle sizes compare with the in situ observations. This is followed by a summary of the results in section 4.

## 2 Description of the model and observations

The Met Office Unified Model (UM) version 8.5 is used to create a series of one-way nested simulations. The global model configuration GA6 (Walters et al. 2015) is the driving model, which uses the Even Newer Dynamics for General atmospheric modelling of the environment (ENDGame) dynamical core (Wood et al. 2014). The global model has a resolution of N512 (~ 25 km) with 70 vertical levels and is run with a 10 minute time step. The convection scheme is based on Gregory and Rowntree (1990) and uses a vertical velocity dependent convective available potential energy (CAPE) closure. The Prognostic Cloud Prognostic

Condensate (PC2) scheme of Wilson et al. (2008) is used with the microphysics scheme described by Wilson and Ballard (1999) but with numerous modifications including prognostic rain and graupel, cloud droplet settling and the Abel and Boutle (2012) rain drop size distribution. The boundary layer scheme used is based on Lock et al. (2000) and the radiative fluxes are determined by the Edwards and Slingo (1996) scheme. The global model is initialised at 00 UTC using the Australian Community Climate and Earth System Simulator (ACCESS; Puri et al. 2013) operational analysis for the case study date of February 18 2014.

The first nested simulation within the global model is a 4 km grid length simulation. These simulations are run with a 100 s time step and are forced at the boundaries every 30 minutes. At this resolution the Smith (1990) diagnostic cloud scheme is used where the critical relative humidity is 0.8 above 800 m and increases to 0.96 at the lowest model level. The cloud microphysical parameterisations are the same as the global model except that the generic ice particle size distribution (PSD) scheme of Field et al. (2007) is used. The convection scheme at this resolution has a modified CAPE closure that scales with grid-box area, which allows for more of the convective activity to be modelled explicitly. The other difference from the global model is the diffusion. While there is no horizontal diffusion in the global model, in the 4 km model this is modelled by a Smagorinsky (1963) type scheme and the vertical diffusion coefficients are determined using a scheme that blends those from the boundary layer scheme and the Smagorinsky scheme (Boutle et al. 2014). The older dynamics scheme (named New Dynamics; Davies et al. 2005) is used in the control model configuration, as that dynamical core was the one being used in the high resolution operational model forecasts for this version of the model. However, the effects of the dynamics are also tested by using ENDGame in a sensitivity experiment.

A suite of 1 km simulations are nested in the 4 km simulation that investigates the effects of the dynamics, turbulence and microphysical parameterisations on the simulations of tropical convective clouds. There are 80 vertical levels and the model is run with a time step of 30 s. The domain is 500 x 500 km$^2$ centred on the location of the Darwin radar (12.25 $^{o}$S, 131.04 $^{o}$E) as shown in Figure 1 and the convection is modelled explicitly. Given that the focus of this work is primarily on the cloud microphysics, a description of the scheme used in the model is provided, with the details of the other parameterisations available in the previously cited references. The microphysics scheme is described by Wilson and Ballard (1999) but with numerous modifications. The single moment scheme carries water in four variables:

vapour, liquid, ice and rain, with an additional graupel variable in the 1 and 4 km simulations. The 4 km and control version of the 1 km model use the generic ice particle size distribution of Field et al. (2007), where the aggregates and crystals are represented by a single prognostic aggregate variable. This parameterisation is based on the idea of relating moments of the size distribution to the second moment, which is directly proportional to the ice water content when mass is equal to the square of the particle size. In using this parameterisation there is no need to specify an intercept parameter for the PSD and instead the microphysical transfer rates are derived from the moment estimation parameterisation that is a function of ice water content and temperature. The mass-diameter relationships take the form of a power law

$$m(D) = aD^b \quad (1)$$

The particle size distributions are generalised gamma functions

$$N(D) = N_0 D^{\mu} e^{-\lambda D} \quad (2)$$

where $N_0$ is the intercept parameter, μ is the shape parameter and λ is the slope parameter. The coefficients for each hydrometeor species are given in Table 1, where the aggregate and crystal PSD coefficients are for the simulations that use an explicit PSD and not the generic ice PSD parameterisation. The explicit ice size distributions have a temperature-dependent intercept parameter that decreases with warming temperatures, representing larger particles and the effect of aggregation (Houze et al. 1979), where in Table 1

$$f(T) = \exp\left(-\frac{\max(Tc, -45°C)}{8.18°C}\right) \quad (3)$$

following Cox (1988) with $T_c$ the temperature in degrees Celsius. Fall speeds are parameterised from power laws with the coefficients for crystals and aggregates from Mitchell (1996), graupel from Ferrier (1994) and rain from Abel and Shipway (2007).

Ice can be formed by homogeneous and heterogeneous nucleation processes. At -40 °C and below, homogeneous nucleation instantaneously converts all liquid water (both cloud water and rain) to ice. Heterogeneous nucleation requires cloud water to be present at temperatures at or below -10 °C. The process is dependent on relative humidity and the mass of the number of active nuclei produced from the temperature dependent function from Fletcher (1962). Once ice has been formed it can grow by vapour deposition, riming, collection and aggregation. The autoconversion of snow to graupel occurs when snow growth is dominated by riming, with the additional conditions that the snow mass threshold is exceeded and the temperature is below -4 °C. Once graupel has formed it grows by riming and collection. The ice hydrometeors experience sublimation, evaporation and melting. There are a number of graupel transfer terms that have not been included in the model as their rates are significantly smaller than the dominant processes (Wilkinson et al. 2013). The graupel terms not included are: deposition and sublimation; wet mode growth; collection of ice crystals; and heterogeneous freezing of rain by ice nuclei.

The control model (denoted as nd) in the set of 1km simulations uses the New Dynamics and the sensitivity to dynamical formulation is investigated by testing the ENDGame dynamical core in the simulation denoted eg. Modelling the vertical turbulent mixing using the 3D Smagorinsky scheme rather than the blended scheme used in the control simulation is labelled 3d. The other experiments test aspects of the microphysical parameterisations:

nopsd – Rather than use the generic ice PSD as in the control experiment, explicit PSDs are used for ice where the single ice prognostic is diagnostically split as a function of the temperature difference from cloud top into two categories to represent the smaller more numerous ice crystals and larger aggregates (Wilkinson et al. 2013).

qcf2 – As for nopsd but the crystals and aggregates are represented as two separate prognostic variables.

qcf2hm – As for qcf2 but with the inclusion of an ice splintering parameterisation that increases the deposition rate in the Hallett-Mosssop (1974) temperature zone of -3 to -8 °C. This parameterisation represents the increase in the ice particle number concentration due to ice splinter production during riming and is dependent on the supercooled liquid water content, and as such the riming rate, as well as the temperature that allows for increased deposition at temperatures colder than -8 °C due to the vertical transport of ice splinters (Cardwell et al. 2002).

qcf2ndrop500 – As for qcf2 but with an increase in the cloud droplet number concentration from 100 $cm^{-3}$ to 500 $cm^{-3}$.

qcf2sr2graupel – As for qcf2 but with the restriction that snow-rain collisions do not produce graupel.

qcf2noqgr – As for qcf2 but without the inclusion of graupel.

qcf2rainfreeze – As for qcf2 but with the inclusion of a heterogeneous rain freezing parameterisation based on the stochastic parameterisation of Bigg (1953) following Wisner et al. (1972). This process represents the heterogeneous freezing of rain by heterogeneous nucleation by ice nuclei.

qcf2raindsd – As for qcf2 but with the Marshall-Palmer (1948) rain drop size distribution.

The Darwin C-band polarimetric (CPOL) radar (Keenan et al. 1998) collects a 3D volume of observations out to a range of 150 km. The radar observations have been interpolated onto the model 1 km grid, and the analysis of radar reflectivities is for the area encompassed by the radius < 150 km from the radar (see Fig. 1). The precipitation rates derived from the radar reflectivity have uncertainties of 25% at rain rates greater than 10 mm $hr^{-1}$ and 100% for the lowest rain rates (Fridlind et al. 2012). The satellite observations of outgoing longwave radiation (OLR)~~, cloud top height~~ and ice water path (IWP) were derived from the geostationary satellite MTSAT-1R following Minnis and Smith (1998) and Minnis et al. (2008; 2011). Observations from the French Falcon 20 aircraft include~~are from research flight 23.~~ t~~T~~he ice water content (IWC) measurement~~ was~~ made with the isokinetic evaporator probe IKP-2 (Davison et al. 2009), and the ice particle size distribution reconstructed from images of individual particle~~s are~~ from the 2D-Stereo (Lawson et al. 2006) and precipitation imaging probes (Baumgardner et al. 2001). The particle probes were fitted with anti-shattering tips and the p~~P~~rocessing of the size observations accounted for any possible remaining ice shattering by consideration of the inter-arrival times and the ratio between the particle surface and lengths (Leroy et al. 2015). Since the IKP-2 measures the total water content, liquid water and water vapour contributions should be subtracted to obtain IWC. Unfortunately, the hot-wire liquid water content (LWC) sensor on the aircraft was unable to measure LWC below about 10% of the IWC in mixed phase conditions, and LWC levels exceeding this value were very rare.  Fortunately the Goodrich Ice Detector could be used to detect the presence of liquid water. Two such regions in two very short flight segments for this case, research flight 23, were identified at -10 °C, and these regions have been excluded from the analysis. The minimum detectable IWC of the IKP-2 is determined by the noise level of the water vapour measurements of the IKP-2 and background probes. This resulting noise level of the subtraction of the background humidity from the IKP-2 humidity is a function of temperature: it is about 0.1 g $m^{-3}$ at -10 °C, dropping rapidly to about 0.005 g $m^{-3}$ at -50 °C.  Since most data were taken at temperatures colder than about -25 °C, a minimum IWC of 0.05 g m$^{-3}$ was chosen as the threshold to include in our analysis.

Two sources of vertical velocity are used from the Falcon 20. Position, orientation and speed of the aircraft are measured by a GPS-coupled Inertial Navigation System. The 3D air motion vector relative to the aircraft is measured by Rosemount 1221 differential pressures transducer connected to a Rosemount 858 flow angle sensor mounted at the tip of the boom, ahead of the aircraft, and by a pitot tube which is part of the standard equipment of the aircraft. Wind in local geographical coordinates is computed as the sum of the air speed vector relative to the aircraft, and the aircraft velocity vector relative to the ground. Both computations use classical formulas in the airborne measurement field described in Bange et al. (2013). The other vertical air velocity measurement used is retrieved from the multi-beam cloud radar observations using the 3D wind retrieval technique described in Protat and Zawadzki (1999), and we use the technique described in Protat and Williams (2011) to separate terminal fall speed and vertical air velocity. Comparisons near flight altitude with the aircraft in-situ vertical velocity measurements show that the vertical velocity retrieval is accurate to within 0.3 m s$^{-1}$. All observations are averaged to the model 1 km grid.

## 3 Comparison of the simulations with observations

On February 18 2014 the monsoon trough was stalled near the base of the Top End with active conditions continuing about the northern coast. There was a deep moisture layer and low level convergence that produced a mesoscale convective system. At 14:30~~12~~ UTC, satellite imagery shows the convection around Darwin was somewhat isolated in nature, with a convective cell developing close to the radar~~ by 15 UTC~~ (Figure 2)~~ (not shown)~~. This convection developed into a larger organised oceanic mesoscale convective system by 18 UTC with deep convective cells producing cloud top temperatures of -80 $^{o}$C. A widespread region of anvil cloud produced from the outflow of deep convection was seen to develop from 18 UTC and persist for over 8 hours. The HIWC research flight penetrated convective cores in a region northeast of the radar at 22 – 24 UTC (Fig. 1) with peak ice water content up to 5 g m$^{-3}$ at 1 s resolution. There was almost no supercooled water detected during the flight, even at -10 $^{0}$C, and graupel was intermittently observed. The absence of supercooled water coupled with the occasional presence of graupel is due to the system being sampled at the mature-decaying stage, where the supercooled water had been consumed in the production of graupel. Most of the time the particle images were of dense ice aggregates at flight level, except within some convective cores where graupel was observed, as also indicated by strong W-band attenuation.

Comparison of the modelled outgoing longwave radiation (OLR) with the satellite observations in Figure 2 show that in general, the control simulation represents the lifecycle of the MCS fairly well. The location of the mostly oceanic convective cells look reasonable, however, the modelled MCS is larger and composed of more numerous and deeper convective clouds than what was observed in the pixel level satellite OLR data and seen in the low level radar reflectivity fields shown in Figure 3. The model also produces more convection over the Tiwi Islands than what was observed at 17:30 UTC. As the MCS transitions from a developing-mature system through to a mature-decaying system, the observed reduction of deep convective cells with time is simulated, although the OLR remains significantly underestimated. During the research flight at 23:30 UTC, the modelled MCS shows cloud positioned in a similar location to that observed with respect to the MCS structure, however, the modelled cloud is shifted somewhat to the northeast (Fig. 2h,l).

The sounding at 23 UTC (Figure 2) shows a temperature of 24 ºC at 70 m and an unstable environmental lapse rate, with the temperature gradient reducing at 700 hPa. This height corresponds to the typical cloud base in the region as observed by satellite at about 3 km and saturation is observed at the freezing level at 4.6 km (570 hPa). The control 1 km model shows a reasonable representation of the low level temperature up to 800 hPa, where the model is then warmer up to 600 hPa. This simulation is drier in the levels below 4 km and then has excessive moisture throughout the mid and upper troposphere, maintaining saturated air with a warm bias present from 400 hPa (7 km). The upper level moisture bias is not present in the global model simulation, however it is apparent in the 4 km simulation. This bias is seen in the relative humidity regardless of whether the individual model grid box at the sounding location is used as in Figure 1 or whether an area averaged domain is used as shown in Figure 4a. At this time the model simulates almost completely overcast conditions, which compares well to the satellite observed cloud cover of 95%. Excessive moisture in small domain simulations is a common error related to the limited domain size that does not allow for sufficient mesoscale organisation of convection and humidity (Bretherton et al. 2005). Given that the 4 km simulation also shows this error and the domain size in that case is 2000 x 2000 km$^2$, it seems that the upper tropospheric moisture errors in this case are not predominately driven by the domain size.

~~The observed winds tend to be from the south east in the lowest few kilometres and turn clockwise to persist as westerlies from 6 – 12 km. Above this height the wind shifts to be from the east with the largest wind speeds occurring above 14 – 15 km > 20 m s$^{-1}$ (note this is above the pressure range shown in Figure 2). These wind profiles tend to be associated with the active monsoon at Darwin where the migration of the monsoon trough reverses the large-scale circulation (Fein and Stephens 1987). The height of the largest vertical wind shear in the simulations is a couple of kilometres too high but the magnitude and direction of the strong upper level easterlies is represented well. The winds are too strong in the simulations between 1.5 and 4 km and do not have the same easterly component, however, above this level the wind speed is reasonably captured, with these deep westerly winds providing the source of moisture for the deep convective clouds observed and simulated.~~

### ~~3.1 Time series comparison with observations~~

The ~~domain~~ mean precipitation rates and ice water path (IWP~~) ~~) (Fig. 3) calculated for the radar domain shown in Figure 1,~~shown in Figure 3~~ demonstrate that a larger IWP implies a larger surface rainfall rate as seen in previous tropical studies (e.g. Liu and Curry 1999). The radar derived precipitation shows that the simulations overestimate the domain mean rainfall rate during the development stages of the MCS, and produce the peak in precipitation about 2 hours earlier than is observed. The model precipitation maximum occurs when the simulated convection is strongest, as measured by the largest domain mean vertical velocity at 500 hPa and the maximum vertical velocities. The observed domain mean rainfall maximum corresponds to the time when the domain mean cloud top height is highest (not shown), and together with the infrared satellite imagery~~observed brightness temperatures~~ (Figure 2~~not shown~~), suggests that the generation of significant anvil cloud occurs before the domain mean precipitation maximum, rather than when the convection is strongest as is the case in the simulations. Note that the simulated domain mean precipitation rate at both the earlier and later times is outside of the uncertainty range of the radar derived rainfall rate (Fridlind et al. 2012).

The underestimate in modelled surface rainfall for the later times when the MCS has matured is not due to an underestimate in the domain mean upper tropospheric cloud cover, as both the model and satellite observations show mostly overcast conditions, but rather the underestimate in condensate reaching below the freezing level (Figure 3f~~as will be demonstrated in the following subsection~~)~~, which is partly due to a drier lower troposphere as~~

~~shown in Figures 2 and 4~~. The observed IWP is only valid for the daytime from about 22:30 UTC or 8 am local time, and while the simulations with the generic PSD parameterisation compare well with the satellite derived value, the comparison of VISST IWP with CloudSat in tropical regions was shown by Waliser et al. (2009) to be underestimated by 25%, likely due to the maximum retrieved optical depth being limited to 128. Together with the CloudSat uncertainties (30% bias and 80% root mean square error; Heymsfield et al. 2008), t~~T~~his suggests that the modelled domain mean IWP may be underestimated from 22:30 – 23:30 UTC. Other studies have documented the lack of stratiform rainfall in convective-scale simulations and some attributed the error to excessive evaporation in single-moment microphysics schemes that use a constant intercept parameter in the rain DSD (Morrison et al. 2009). That is not the case in this work and rather the cause is likely due to overly strong convection (Figures 2 and 3d~~Sect 3.2.3~~) that detrains too high and does not produce enough condensate in the lower stratiform regions as has been shown by Ferrier (1994), Tao et al. (1995) and Morrison et al. (2009).

~~-~~The greater IWP in the simulations that use the generic ice PSD parameterisation is associated with larger relative humidity in the upper troposphere (Figure 4a). In a study comparing different microphysics schemes, VanWeverberg et al. (2013) found the same result and associated the increased moisture with the sublimation of ice particles due to the scheme with the slowest ice fall speeds producing the greatest condensate and moisture. That is not the case for this current study where the larger IWP and relative humidity is produced by the microphysics configuration that produces larger mean mass-weighted particle sizes (Figure 4c) but similar ice fall speeds above about 12 km, with faster below this height. Figure 4b shows the fall speeds for the ice crystals and aggregates/snow particles. All simulations use the same formulation for snow, and even though the generic PSD only represents a single hydrometeor category there are two fall speeds used to enable a representation of both fast and slow sedimenting particles based on size. The method when using the generic PSD is described by Furtardo et al. (2014) where for narrow size distributions and small mean sizes the fall speed used is that shown for the ice crystals in Figure 4b, and for broader size distributions and larger mean sizes the snow fall speed is used (the cross over is around 600 μm). Looking at the mean mass-weighted ice~~ particles~~ diameter~~size~~s in Figures 4c and 4d shows larger sizes for the simulations that use the generic PSD, however, the slower ice crystal fall speed used in these cases produces a similar mean fall speed to the simulations that use two ice prognostics.

The higher RH in the simulations using the generic ice PSD could be~~is~~ due to the larger, faster falling particles in the levels below 12 km removing more of the LWC via riming~~ (explored later in Section 3.3)~~, which would allow~~s~~ for greater supersaturation. More riming would release~~s~~ more latent heat, which along with the larger ice particles being more effectively off-loaded, could lead to the ~~generat~~generation of~~es~~ stronger updrafts with less entrainment and higher RH in the upper troposphere. This is illustrated in the convective updraft (> 1 m $s^{-1}$) horizontal mass divergence profiles shown in Figure 5a. As discussed by Yuter and Houze (1995), the presence of decelerating updrafts and accelerating downdrafts can be largely explained by entrainment. Entrainment reduces the buoyancy of updrafts, slowing and eventually stopping the air parcel, which is where divergence is expected. In contrast, entrainment into downdrafts enhances evaporative cooling, increasing the downward mass transport and convergence. ~~The simulations that use the generic ice PSD produce less horizontal mass divergence in the levels above 12 km, suggesting reduced entrainment and deposition of mass at these heights. Instead updrafts in these simulations tend to penetrate higher, in agreement with Figure 3.~~ Note that above 16 km the vertical velocities show oscillatory motions consistent with gravity waves, and therefore, above this height the mass divergence appears to be driven by these waves.

Figure 5a shows that horizontal mass divergence~~entrainment~~ in the mixed-phase regions of the convective updrafts is the most sensitive to the turbulence formulation in the model, with the simulation with greater turbulent mixing (3d) showing greater mass divergence, indicative of~~and~~ greater entrainment, in the range of 5 – 8~~7~~ km. This contrasts with~~to~~ the upper ice-~~-~~only regions of the convective updrafts that show that the largest control on horizontal mass divergence~~entrainment and buoyancy~~ is the ice sizes. The simulations with smaller sized particles have more ~~horizontal~~ mass divergence above 12 km, indicating more entrainment and a larger reduction in the buoyancy in the upper levels of convective updrafts than the simulations with larger sized ice particles. This is confirmed by examining the convective updraft buoyancy properties at 14 km shown in Figure 5b and c. The buoyancy, $\Delta\ \theta_d$, is calculated from the difference in the density potential temperature (that includes condensate) from the slab mean for the convective updrafts with vertical velocity > 1 m $s^{-1}$. Comparing the equivalent potential temperature as a function of $\Delta\ \theta_d$ at 14 km (Fig. 5b) between simulations with larger and smaller ice sizes shows that for the positively buoyant updrafts, the simulation with smaller ice sizes has fewer occurrences of high $\theta_e$. This~~-~~ gives support to the argument derived from the convective updraft horizontal mass divergence that entrainment is larger in the upper ice-only convective updrafts when the ice sizes are smaller, although we do note that some of this difference could be due to differences in freezing. To analyse this in more detail, the histogram of convective updraft buoyancy (Fig. 5c) shows a greater number of occurrences of more positively buoyant clouds at 14 km for the simulations that have larger sized ice particles, supporting the argument that less horizontal mass divergence represents less entrainment with more positively buoyant updrafts that penetrate higher (as confirmed by examining the cloud top height distributions; not shown). Similarly, comparing $\theta_e$ as a function of $\Delta\ \theta_d$ at 6 km between the control simulation and the one that increases turbulent mixing, shows that the case with greater mixing has significantly more occurrences of low $\theta_e$, consistent with greater entrainment.

The satellite retrieved cloud top height shows a variation in domain mean of greater than 2 km over the 12 hours of the MCS lifecycle analysed (Fig. 3c). The simulations show typically only a 500 metre change, reducing from 12 – 24 UTC. While the domain mean cloud top height agrees reasonably well with the satellite observations, the outgoing longwave radiation (OLR) does not with the simulations reducing the OLR by 50 – 100 W m$^{-2}$ too much (Fig. 3d). The simulations that use the generic ice PSD have higher cloud tops with colder temperatures and greater IWP that produce lower OLR than the simulations that use explicit ice PSDs (20 – 30 W m$^{-2}$ lower) and the observations (~80 W m$^{-2}$ lower). The minimum observed OLR at 20 UTC is captured by most of the simulations, with the simulations then tending to increase OLR at a faster rate than is observed as the MCS structure matures to be composed of mostly stratiform cloud.

### 3.23.1 Radar reflectivity characteristics

The model hydrometeor fields have been converted into radar reflectivities by assuming Rayleigh scattering, with no consideration of the effects of attenuation or attempt to model the radar bright band. Due to the long wavelength of the CPOL radar (5.3 cm) modelled reflectivity is calculated following Hogan et al. (2006) where the reflectivity is considered proportional to mass squared

$$Z = R\int_0^\infty M(D)^2 N(D)dD \qquad (4)$$

where $R = 10^{18}\frac{|K|^2}{0.93}\left(\frac{6}{\pi\rho}\right)^2$, ρ is the particle density and the mass $M$ and particle size distribution *N(D)* are defined by (1) and (2). For cloud liquid water the reflectivity is calculated from the constant number concentration of 100 cm$^{-3}$ in the simulations with the size distribution $N(D) = PD^2 \exp^{-\lambda D}$, where $P = N/2\lambda^3$ following McBeath et al. (2014). The dielectric factor $|K|^2$ is set to 0.93 for water and 0.174 for ice. The particle densities used in the calculation of *R* are 1000 kg m$^{-3}$ for rain, 917 kg m$^{-3}$ for aggregates and crystals and 500 kg m$^{-3}$ for graupel. For the simulations that use the generic ice PSD parameterisation, the aggregate reflectivity is proportional to the 4th moment of the PSD, which is calculated from the Field et al. (2007) moment estimation parameterisation.

#### ~~3.2.1~~3.1.1 Statistical radar coverage analysis

To examine the temporal evolution of the mesoscale convective system and evaluate the modelled MCS lifecycle and the simulated reflectivities, a statistical coverage product has been produced following May and Lane (2009). The data used to construct the statistical product are reflectivity fields from CPOL and the simulations every 30 minutes for 12 hours from 12 – 24 UTC. At each height the fraction of the total area within the radar domain covered by reflectivity thresholds is calculated, with the thresholds chosen as 10, 20, 30 and 40 dBZ.

The observed statistical radar coverage product shown in Figure 6 illustrates the development of the MCS. At 12 UTC the radar domain has a low fractional area coverage of up to 0.15 for the 10 dBZ threshold, showing that at 12 UTC there were radar-detectable hydrometeors covering 5 – 15% of the radar sampling area between the lowest detectable altitude of 1.5 km and 8 km. Highest reflectivity echo~~cloud~~ tops of 11 km are seen in the > 10 dBZ fractional coverage at 17:30 UTC, which coincides with the time that the very cold cloud tops associated with deep convective cells were seen in the satellite imagery (Fig. 2). The maximum coverage of the domain by hydrometeors with reflectivities > 10 dBZ is 85% seen at 21 – 22 UTC, which is when the large anvil cloud shield appears a few hours after the deepest convection occurs. The observed areas of reflectivity > 10 dBZ are fairly uniform with height from 2 – 6 km, demonstrating little variability of the reflectivity echo~~hydrometeor~~ coverage from the low levels to a couple of kilometres~~m~~ above the freezing level. Fractional areas larger than 0.05 with reflectivities > 20 dBZ are mostly confined to below 6 km, with the maximum fraction of 0.65 occurring at 21 UTC at 4 km. The > 30 dBZ area is not greater than 10% until 16 UTC, and is maximum between 20:30 – 22 UTC at 4 km with a value of

0.35. There is no fractional area of the domain > 0.05 that contains observed reflectivities greater than 40 dBZ.

While the statistical radar coverage product produced for the control simulation does show a transition to~~from scattered to more organised convection with~~ widespread stratiform cloud regions, as shown by the peak < 10 dBZ coverage at 21 UTC, and predicts the timing of the deepest clouds generally well (Fig. 6), there are clear deficiencies in the simulated evolution of the MCS. There are much larger high dBZ fractional areas, deeper clouds occur too early in the simulation and there is a strong vertical gradient in the area coverage with height. The less uniform vertical area coverage shows that the simulated clouds have more variability in reflectivity with height compared to the observations. In coarse resolution models a common model error is too little detrainment at the freezing level (e.g. Franklin et al. 2013), however, in this convection permitting simulation the change in hydrometeor area with height is mainly due to too little stratiform cloud and rain area, which explains the reduction in area below the melting level and the convective-stratiform modelled ratio being skewed towards more convection than is observed (discussed in section 3.2.2).

A clear difference between the observations and the simulation is the > 20 dBZ reflectivity areas above the freezing level. The observations show some hydrometeors present 1 – 2 km above the freezing level that have reflectivities > 20 dBZ, but no areas that meet the minimum threshold of 5% that have reflectivities > 30 or 40 dBZ. The simulation on the other hand shows large > 20 dBZ fractional areas > 0.6 indicative of larger ice particles in the model than in the observations, which will be explored in detail later. The simulated reflectivity area > 30 dBZ above 5 km is due to the presence of both ice and rain, and the > 40 dBZ areas are almost exclusively due to rain. The simulated rain above the freezing level that is not observed suggests that~~ either~~ the model has faster updrafts than observed, which loft large rain particles upwards and/or the heterogeneous freezing of rain that is not represented in the model is an important process in tropical convection and/or other errors in the representation of the rain DSD. This ~~latter~~ result is what motivated the experiment with the addition of a heterogeneous rain freezing parameterisation as observations in oceanic convection have shown that most drops freeze between about -6 and -18 $^{o}$C (Stith et al. 2002, 2004; Heymsfield et al. 2009).

All simulations show the same main errors in the statistical radar coverage as the control~~ nd~~ case, nd~~ (not shown)~~. The simulation that uses a differing turbulent mixing formulation produces~~ing~~ the closest representation of the observed fractional areas for the dBZ thresholds of 10 and 20 dBZ~~, particularly~~ in the larger areas below the melting level (Fig. 6i, j). This can likely be attributed to~~ the~~ greater horizontal mass divergence~~detainment~~ between 5 and 8 km at the earlier convective times~~ (see Fig. 5d)~~ (Fig. 5)~~due to greater ~~, indicative of increased entrainment and mixing of environmental air in this simulation, which acts to increase the amount of IWC (Fig. 3 and 13) and the area of precipitation.

### ~~3.2.2~~3.1.2 Contoured frequency by altitude diagrams

The CPOL contoured frequency by altitude diagram~~s~~ (CFAD~~s~~) using the observations from 23 – 24 UTC every 30 minutes exhibits a fairly narrow distribution at the heights above the freezing level, with the altitude range of 12 – 13 km having little variability, reflecting the dominance of small ice particles growing primarily by deposition in the uppermost cloud levels (Figure 7a). Below 10 km the distribution shows increasing reflectivity with decreasing height as particles grow rapidly through aggregation, with reflectivities centred on the modal value of 10 dBZ. At altitudes below the melting level the distribution widens and the reflectivities extend from 5 – 35 dBZ with the largest occurrences around 30 dBZ. The lack of a predominant bright band in the observations is likely~~may ~~ due to the data being collected from volumetric scans, however, there are slightly higher reflectivities seen at 4 km indicating a bright band. ~~indicate that the particles were heavily rimed rather than aggregated, low density snowflakes due to differences in the dielectric constant and size as these particles melted into rain (e.g. Hogan et al. 2002).~~

The simulations all show the common errors of: clouds within these reflectivity regions extending too high, reflectivities that are too large between 4 – 6 km, greater reflectivity range below 4 km, and disjointed profiles due to separate hydrometeor categories. The simulations show more of a convective type profile with broader distributions above the freezing level compared to the observations. The more numerous high reflectivity outliers in the simulations indicate a larger number of deep convective cells and/or a smaller proportion of convective – stratiform area.

The simulation with the different dynamical core, ENDGame shown in Figure 7c, shows higher clouds and a broader range of reflectivities at 14 – 16 km. This latter result suggests the presence of large particles being lofted into the upper cloud levels by intense convective cores, as can be seen by the 40 dBZ reflectivities at 17 km. The observations do show some sign of this lofting occurring at 11 – 12 km, however, the reflectivities are constrained to be <

20 dBZ. This feature can also be seen in the cases that include the ice splintering process, the limited graupel case and the increased droplet number concentration case. The simulations that use the generic ice PSD parameterisation (Fig. 7b and c) overestimate the occurrence of low reflectivities above 10 km and have a modal reflectivity at 6 – 8 km that is too low compared to the observations. Using explicit ice PSDs produces a closer match to the observed reflectivity distribution above 10 km, although the simulated clouds still have greater vertical extent, and~~.~~ t~~T~~he modal value of the reflectivities at 6 – 8 km with the explicit PSDs is approximately ~~1~~5 dBZ too large~~, which is greater than the observed value of 10 dBZ~~.

The inclusion of a heterogeneous rain freezing parameterisation reduces the number of occurrences of reflectivities > 20 dBZ between 5 and 10 km and reduces the cloud top heights. Both of these results agree better with the observations suggesting that this process may be~~is~~ important in tropical convective cloud systems. However, given the errors in the dynamics and microphysics in the model for this case, further study is required to better understand the effects of this process. Even in the simulation without graupel the reflectivities are overestimated at the melting level (not shown) and this is due to the ice aggregate PSD. ~~Unlike double moment microphysics schemes, single moment schemes cannot increase the number concentration as the IWC increases and is why the overestimation in reflectivity is seen, even without the contribution from graupel.~~

Focussing on the 2.5 km reflectivity distribution shown in Figure 8a allows an evaluation of the rain properties from the simulations, in particular the rain DSD. All simulations except for one use the Abel and Boutle (2012) rain DSD, with the remaining simulation testing the sensitivity of rain drop sizes by using the Marshall-Palmer (1948) DSD. The Abel and Boutle rain DSD represents the observed rain reflectivity distribution fairly well, however, the observed peak of 30 dBZ is underestimated and there are too many occurrences in the tails of the distribution. ~~The drier subcloud levels (Fig. 2. and 4) are likely to contribute to the underestimate of the peak reflectivity through enhanced evaporation but cannot explain the larger reflectivities that could result from the stronger convective dynamics as well as the prescribed rain sizes.~~ The contribution from the convective updrafts is demonstrated by the largest occurrences in the high reflectivity tail coming from the simulation with the different dynamical core. It is this ENDGame simulation that produces the strongest updrafts (Fig. 11) and is the least representation of the observed rain reflectivity distribution for the reflectivities

> 40 dBZ. The simulation using the Marshall-Palmer DSD peaks at too low a reflectivity at around 10 dBZ and produces too many small rain drops with low reflectivities.

At 6km the observations again show a bimodal~~single peak~~ reflectivity distribution, with the largest peak centred on approximately 16~~5~~ dBZ (Figure 8b). The simulations show a more complicated distribution at this height with multiple modes due to the presence of multiple hydrometeor species. The simulations that use the generic ice PSD parameterisation peak at -1 dBZ. When this parameterisation is not used and the explicit ice size distribution is used the peak is too high at 24 dBZ. When an additional ice prognostic is added this peak is reduced and compares better to the observations at 18 dBZ, however, the tail of the distribution in these cases is too long with too many occurrences at high reflectivities. While the tail of the distribution for the generic ice PSD cases is also too long, compared to the observed reflectivity distribution these cases represent the graupel reflectivities better than the cases that use the explicit PSD even though all cases use the same graupel PSD. The better graupel representation with the generic ice PSD coupled with the significantly larger occurrence of weak reflectivities around 0 dBZ is similar to the result found by Lang et al. (2011). They modified microphysics parameterisations to reduce the occurrence of excessive large reflectivities and found that this resulted in too many low reflectivities due to a shift in the reflectivity distribution, as is this case here when comparing the generic and explicit ice PSD cases. ~~They suggested that this may be due to entrainment and the sublimation of small ice particles resulting in the observed particle sizes and reflectivities being larger for the low reflectivity end of the distribution than seen in the simulations. This reasoning does not fit this case because the ice sizes from the simulations that use the generic PSD at this height are significantly larger than the simulations with the explicit ice PSD (Fig. 4) and the entrainment from the 3d simulation with the differing turbulent mixing is larger than the other cases that use the generic ice PSD (Fig. 5) yet the reflectivity distribution is very similar suggesting that reduced entrainment is not responsible.~~

To examine to what extent the generic ice PSD parameterisation is misrepresenting the observed reflectivities or how much the erroneous cloud dynamics are responsible for errors in the modelled reflectivities, the PSD moments derived from the generic PSD parameterisation using the observed IWC and temperature are shown in Figure 9. In calculating the predicted moments the observed mass-diameter relation was used, $m = 4.97\times10^{-3} D^{2.05}$, and the observed moments are calculated only for particle sizes >

100 µm in diameter and for IWC > $10^{-3}$ g $m^{-3}$ to be consistent with the data used to derive the Field et al. (2007) parameterisation. The $4^{th}$ moment is equivalent to radar reflectivity when mass is proportional to the square of the particle diameter, and it can be seen in Figure 9a that the slope of the parameterised reflectivity results in an overestimate of the larger reflectivities. The generic ice PSD parameterisation underestimates the zeroth and first moments and has a good representation of the third moment. The underestimate of the number concentration (Fig. 9d) is consistent with the overestimation of particle sizes and reflectivities. The observations in this case may be in a~~are sampled near convective cores, which is a~~ different type of cloud environment from the data used to construct the Field parameterisation, as suggest~~demonstrat~~ed by the observed number concentration being below the lower range shown in Field et al. (2007).

### ~~3.2.3~~3.1.3 Maximum reflectivity profiles and vertical velocities

In agreement with many previous studies (e.g. Blossey et al. 2007; Varble et al. 2011) the model overestimates the reflectivity above the freezing level as can be seen in the profiles of maximum reflectivity shown in Figure 10, as well as overestimating the rain reflectivities below 5 km. From the set of simulations it can be seen that graupel is not the sole cause of the significantly higher reflectivities as the simulation without graupel also displays this bias. The largest difference between simulated and observed maximum reflectivity during 23 – 24 UTC occurs above 7 km and increases with height for many of the simulations, with the difference between the simulation with the different dynamical core and the observations at 10 km equal to 40 dBZ. The observations show a decrease in the maximum reflectivity with height from approximately 2 km, whereas the simulations tend to show a more constant profile. The observed reduction in height may be due to large raindrops falling out of strong updrafts or due to raindrops falling through weak updrafts and growing due to the accretion of cloud droplets. The likely overestimate in updraft strength in the simulations (shown next) will advect the raindrops upwards allowing these particles to be collected by the existing ice, generating larger ice particles and maximum reflectivities above the freezing level, as well as acting as a source of latent heating to further fuel convective updrafts. The simulation that decreases the maximum reflectivity with height the most is the simulation with differing subgrid turbulent mixing (Figure 10b), which ~~tends to~~ suggests weaker updrafts. The addition of a rain heterogeneous freezing parameterisation follows the different turbulence simulation in reducing the maximum reflectivity from the freezing level up to 8 km, reflecting the reduction in rain and a better representation of the reflectivities.

At 17 – 18 UTC, when the greatest amount of deep convection occurs~~is the strongest~~ in all of the simulations and the coldest satellite derived cloud top temperatures are observed, the CPOL maximum reflectivity profile has a more constant profile with a slower reduction of reflectivity with height as compared to the later less convective times (Fig. 10). The observed 40 dBZ contour reaches 8 km in agreement with the results of Zipser et al. (2006) who showed that radar echoes of this strength rarely occur above 10 km. The profile of maximum reflectivity from the simulation that uses the new dynamical core shows essentially the same profile at these strong convective times as for the later times when the MCS has matured, unlike the observations and the majority of the simulations, suggesting that there is less variability in maximum updraft when using ENDGame. There is little spread in the maximum reflectivity profile across the simulations at 17 – 18 UTC, with strong updrafts > 20 m s$^{-1}$ in all simulations (not shown) that allows large particles to be advected into the upper troposphere. T~~with t~~here is a clear difference in the two simulations that limit or exclude graupel, demonstrating that at the time of strongest convection, the vertical advection of graupel is responsible for the largest error in the maximum reflectivities in the upper troposphere.

Comparing the control case with the cases that use a different dynamical core and different turbulent mixing parameterisation shows that the reduction in maximum reflectivity with height at 23 – 24 UTC is well correlated with the reduction in maximum vertical velocity shown in Figure 11b~~e~~. These cases all use the generic ice PSD and the differences are likely due to the different entrainment~~-~~ and water loading that affects the cloud buoyancy and the strength of the updrafts that advect large particles into the upper troposphere. The ENDGame simulation produces significantly larger maximum updrafts and has less accumulated ice water (see Fig. 13~~6~~). C~~, and c~~onversely there is greater accumulated IWC for the simulation with the different turbulent mixing parameterisation compared to the control case, supporting the argument that water loading differences likely contribute to the differences in ~~and associated lower~~ maximum vertical velocities and maximum reflectivities.

Comparing the differences in maximum vertical velocity across the simulations for the times 23 – 24 UTC shows that the largest sensitivity tends to come from the choice of dynamics and turbulence. The reduction in updraft strength at these times with the 3D Smagorinsky turbulence scheme is also achieved with the inclusion of a ~~rain~~ heterogeneous freezing rain parameterisation. Both of these cases tend to have larger ice water contents in strong updrafts (see Fig. 12) that will reduce buoyancy through the effect of water loading. While there is different sampling between the aircraft observations and the simulations, the aircraft observations of maximum updraft strength shown in Figure 11~~0~~ are smaller than the ENDGame simulation by as much as 20 m $s^{-1}$. In this simulation it seems as though the stronger and deeper updrafts are able to generate enough latent heating that this effect on buoyancy is larger than that of entrainment and water loading as compared to the other cases. The in-cloud mean vertical velocity for this simulation is also larger than the other cases from 4 – 8 km, as well as the 99$^{th}$ percentile of upward vertical motion (Figure 11). The shape of the mean updraft velocity is similar for the ENDGame case and the simulation without graupel, both showing greater mean updraft strength from 3~~cloud base~~ to ~~6 –~~7 km. These two simulations produce the largest domain mean rain rate (Fig. 3a) at these times and show that dynamical changes to the cloud system can be achieved through changes to the model's dynamical core and the cloud microphysics.

While the maximum updrafts produced by the simulations at these times are within the range of observed maximum tropical updrafts from other field campaigns at Darwin (e.g. < 25 m $s^{-1}$ in TWP-ICE; Varble et al. 2014a), the maximum updrafts produced throughout the MCS lifecycle are much larger and in excess of 50 m $s^{-1}$ for the ENDGame simulation at 17 – 18 UTC. These values are well outside the range of maximum vertical velocities presented for oceanic convection by Heymsfield et al. (2010) and agree with other studies showing excessive tropical vertical velocities simulated by convection permitting models. Hanley et al. (2014) demonstrated that the UM with a grid length of 1.5 km simulated convective cells that were too intense and were initiated too early, as was also shown by Varble et al. (2014a), suggesting that convection is under resolved at grid lengths of order 1 km. Improved initiation time was shown by Hanley et al. (2014) to occur when the grid length was reduced to 500 and 200 m. However, the intensity of the convective cells was not necessarily improved, with the results being case-dependent. Varble et al. (2014a) showed that in the tropics the intensity of the updrafts remained overestimated even at the 100 m grid length. Both of these studies suggest that there are missing processes in the model and/or the interactions between convective dynamics and microphysics are incorrectly represented. ~~(e.g Varble et al. 2014a).~~

~~The control simulation shows a large peak in the mean upwards vertical velocity and the 99th percentile at cloud base at approximately 3 km (Fig. 11). The in-cloud velocity statistics are calculated where cloud and/or ice water is present but does not include rain areas, and hence the peak in updraft strength at cloud base is associated with the buoyancy production generated by the condensation and latent heating of air that reaches saturation.~~ Most of the simulations show a double peak in vertical velocities with maxima at 3 km~~cloud base~~ and in the upper troposphere at about 13 km. The upper level updraft peak has been observed (e.g. May and Rajopadhyaya 1999) and is argued to be due ~~to~~to the deep column of convectively available potential energy in the tropics, coupled with latent heat released by freezing condensate and the unloading of hydrometeors, both of which increase parcel buoyancy. A bimodal peak has been observed but tends to be correlated with the freezing level rather than a couple of kilometres lower as~~and not cloud base as seen~~ in the simulations. The apparent lack of observational support for the low level~~cloud base~~ peak is likely due to the inability of many observations to distinguish between non-precipitating cloud and clear air, and dual profiler measurements during TWP-ICE do show some evidence of a low level~~cloud base~~ peak (Collis et al. 2013).

### ~~3.3~~3.2 Phase composition and comparison with in situ observations

Due to the small sample size of observations from the single research flight on 18/02/2014, the observations from 18 of the Darwin HIWC flights have been used to allow for a more robust comparison of the model to the observations (Fig. 12 and 14). The majority of the flight time for these cases was in clouds with temperatures < -10 °C and vertical motions within the range of -2 to 2 m $s^{-1}$. Therefore, when comparing the model to the aircraft observations, the focus is on this subset of cloud conditions as there are limited observational samples outside of these ranges.

In the simulations, the relationship of IWC to vertical velocity changes with the temperature regime, as shown in Figure 12. For the warmest range of 0 to -5 °C the IWC reduces as the strength of the updraft increases from 1 m $s^{-1}$. For the two intermediate temperature regimes, -5 to -10 and -10 to -20 °C, the IWC is fairly constant with vertical velocities greater than 2 m $s^{-1}$, with the colder regime consisting of 1 g $m^{-3}$ more ice for a given vertical velocity. For the coldest regime analysed the IWC increases as the vertical velocity increases.

For the warmest temperature regime the decline of IWC with updraft speed is offset by the strong increase in LWC, with the fraction of condensate that is supercooled cloud water reaching 0.8 at 15 m $s^{-1}$ (not shown~~Fig.13~~). In this temperature regime there is no new ice being formed as heterogeneous freezing in the model does not occur until the temperature cools to -10 °C. Any ice in this regime has formed above and has been recirculated into these updrafts, and as the vertical velocity increases the saturation specific humidity increases faster than the supercooled water can be removed by deposition and riming resulting in the large LWC. The circulation of ice from high levels to those below was suggested by Black and Hallett (1999) to be a factor in the observed rapid glaciation of clouds in hurricanes. The no graupel and limited graupel cases do not show the same decline in IWC in the warmest temperature regime. For these cases the fraction of condensate that is supercooled water is lower so there is less competition for the available water vapour, which results in greater depositional ice growth. In these simulations the greater proportion of ice mass~~particles~~ with slower fall speeds leads to greater in-cloud residence times producing larger accumulated IWC than the other cases with two ice prognostics (see Fig. 13~~6~~). This shows that when graupel is included in the simulations and allowed to grow unrestricted, the removal of LWC by ice processes is less efficient in this temperature regime. The other simulation with different behaviour and larger IWC in this warmest regime is the case that includes rain heterogeneous freezing. In this simulation there is an additional source of ice and this results in greater IWC in strong updrafts due to the rain that is advected upwards freezing rather than remaining as liquid water as in the other simulations. The impact of this on the cloud liquid water is to increase the cloud water content in strong updrafts as shown in Figure 12~~3~~. This is due to the reduction in the riming of cloud water by graupel~~accretion of cloud water by rain~~ as compared to the accretion of cloud water by rain~~given the reduced rain water content~~.

The large IWC in the downdraft regions of the warmer temperature regime is where graupel is expected, which is often located behind and below the convective updrafts (Barnes and Houze 2014) where the suggestion is that~~the fast fall speeds of~~ these larger particles help to generate downdrafts through mass loading (Franklin et al. 2005; Jung et al. 2012). This argument is supported by analysis of the downdraft IWC that shows that the majority of the ice in the downdrafts is graupel. For example in the control simulation, 82% of the ice mass is graupel for the warmest regime downdraft of 5 m $s^{-1}$.

~~Figure 16 that shows that the simulations with the largest accumulated graupel mass tend to be the simulations with the largest IWC in the downdrafts.~~ The colder regime of -10 to -5 °C shows IWC invariable to vertical velocity. These colder temperatures will produce a greater difference in saturated vapour pressure and saturated vapour pressure over ice and, therefore, larger depositional growth rates via the Bergeron-Findeisen process than the warmest~~r~~ temperature regime. ~~There are few observations within the -10 – 0 °C regimes (Figure 12e), however, the observed IWC for vertical velocities between 0 and -2 m s$^{-1}$ shows broad agreement with the simulations with an average IWC of 0.5 g m$^{-3}$.~~

Compared to the warmer temperature regimes,~~For~~ the temperature regime of -20 to -10 °C shows~~there is~~ a small increase in IWC with vertical velocity (Fig. 12c) due to the effects of heterogeneous freezing (that occurs at temperatures < -10 °C) on increasing the mass of ice and further increases in the vapour pressure. In agreement with the observations, t~~T~~he simulations increase the IWC from -1 – 2 m s$^{-1}$, with ~~show fairly good agreement with the observations across the velocities -1 – 2 m s$^{-1}$, with~~ the mean modelled IWC increasing~~ranging~~ from 0.5 – 2 g m$^{-3}$. The observed IWC then drops off but increases again ~~to be equal to 2.4 g m$^{-3}$~~ for updrafts ≥~~of~~ 13~~5~~ m s$^{-1}$. The reduction in observed IWC seems ~~likely to be~~likely to be due to sampling, with few observations in strong updrafts.~~. For updrafts greater than 10 m s$^{-1}$ there is a large range of variability across the simulations and all are typically within one standard deviation of each other.~~

~~For the coldest temperature regime sampled by the aircraft, -30 to -20 °C, the observations show an increase in IWC as the strength of the downdraft intensifies to -3 m s$^{-1}$ (Fig. 12d), as what is simulated for all temperature regimes. The downdraft IWC of 0.2 – 1 g m$^{-3}$ is in reasonable agreement with the simulations and particularly for the simulation that has the additional ice prognostic variable, where the IWC does not monotonically increase with downdraft strength. Comparing the observed IWC for the two colder regimes shows a decrease in IWC at the colder temperatures, for example IWC is about 2 g m$^{-3}$ at 2 m s$^{-1}$ for the -20 – -10 °C regime and only 1 g m$^{-3}$ in the colder regime. The simulations capture this result and show that the reason may be due to the reduction in supercooled liquid water at the colder temperatures (Fig.14), suggesting that this is an important source for ice particle growth in this simulated case.~~ The spread in IWC across the simulations is typically not statistically significant, particularly for the stronger updrafts, however, the differences can be attributed to the effects that the changes have on producing and removing LWC, with different dynamics, turbulence and microphysics all displaying sensitivities to the amount and distribution of IWC within tropical clouds.

Across the four temperature regimes all of the simulations show an increase in cloud LWC with updraft strength (Figure 12e, f), with the LWC reducing as the temperature cools along with the fraction of condensate that is supercooled liquid water as shown in Figures 13 and 14. The strongest updrafts are associated with convective cores that will have minimal entrainment and consequently high supersaturations. Note that we include only cloud water in these figures, rather than cloud and rain, as it is only the cloud water that is used in the growth of ice via the Bergeron-Findeisen process and allowed to heterogeneously freeze in the model. Including rain water increases the LWC and the variance across the simulations for the warmer regimes but does not change the main conclusions regarding ice growth. Also note that the cloud water contents for the warmest temperature regime agree reasonably well with those presented in Table 3 of Heymsfield and Willis (2014). Between -10 and -5 $^{o}$C the fraction of condensate that is supercooled water reduces significantly compared to the warmer regime, however, the mass of cloud water stays the same. Hence the control on the amount of cloud water that occurs between -10 and 0 $^{o}$C is the updraft strength and not the temperature, due to heterogeneous freezing not occurring until the temperatures cool to -10 $^{o}$C and below. The simulations that use the generic ice PSD tend to have lower liquid water contents for a given vertical velocity, likely due to the increased accretion and riming growth due to the larger ice particle sizes compared to the explicit PSD (Fig. 4 and 147). This result continues to be seen for the colder temperature regimes shown in Figure 14.

Increasing the cloud droplet number concentration in the model only directly impacts the microphysical process of autoconversion between cloud droplets and rain, and reduces the precipitation efficiency. For this case the reduced autoconversion rate does not make a significant difference to the surface rainfall, since the ice processes dominate the rainfall production (see Fig. 3). However, the less efficient transfer of cloud water mass to rain does change the cloud structure with more LWC and a larger amount and fraction of condensate being supercooled water for the temperatures between -10 and -30 $^{o}$C, with the difference between the other simulations increasing with the strength of vertical motion (Fig.12). As cloud water is the only liquid water source used in the model for deposition growth via the Bergeron-Findeisen mechanism and that can freeze heterogeneously, this implies potentially greater growth rates for ice and stronger updrafts through enhanced latent heating; the so-called aerosol invigoration effect (Rosenfeld et al. 2008). While it is not clear from Figure 12 that this is the case, Figure 16 shows that the accumulated amount of aggregate mass is actually less in this simulation with enhanced droplet number concentration, however, this case generates the greatest mass of graupel. This shows that the larger mass of cloud water increases the riming by aggregates and thus the production of graupel, which results in a reduction in the total accumulated ice mass, possibly due to depositional growth of graupel not being included in the model.

The other simulation that produces more cloud water for updrafts > 5 m s$^{-1}$ in the coldest temperature regime is the simulation that includes ice splintering or the Hallet-Mossop process (Fig. 12f4). Looking at the accumulated ice crystal mass between the simulation that does and does not include an ice splintering parameterisation (Fig.13, qcf2 and qcf2hm), shows that while there tends to be less crystal mass at most heights when the H-M process is included, there are crystals present in updrafts up to 15 m s$^{-1}$, whereas in the qcf2 case there are no crystals present in updrafts > 4 m s$^{-1}$ (not shown). Similarly for the aggregates there is ice spread across a wider range of updrafts when the H-M process is included, particularly for the colder temperatures, resulting in a larger accumulated amount of snow and total ice (Fig. 136). The generation of a larger quantity of ice crystal mass in the H-M zone allows for a larger amount to be transported to the upper cloud levels by the convective updrafts where the crystals then grow through deposition, riming and aggregation producing a larger mass of snow. The increased latent heating in the H-M zone does produce a slightly larger 90$^{th}$ percentile cloud updraft velocity (Fig. 11). This increase in the number and/or strength of updrafts supports the transport of more liquid water in the case with the ice splintering parameterisation, which also helps to increase the IWC.

The in-cloud relative humidity is less variable as a function of updraft strength for the warmer temperature regimes in both the observations and the simulations (Fig. 15). The increase in RH as the vertical velocity increases for the colder temperature regimes is seen in the observations and simulations for the low updraft speeds, however, for the stronger updrafts the model either flattens off or continues to increase while the observations reduce the RH. This likely reflects the aircraft sampling and is seen in the IWC as well (Fig 12). Compared to the simulations, the higher RH for the temperature regime of -20 to -10 $^{o}$C in the observations for the updrafts greater than 10 m s$^{-1}$ coincides with less IWC in the observations and more in

~~the simulations. This result suggests that the model is too efficient in reducing supersaturation and growing ice particles through deposition. An additional experiment was performed to test the reduction in capacitance due to an axial ratio not equal to one (i.e. non-spherical particles). This reduction in the depositional growth rate did reduce the IWC (the total accumulated ice reduced by 5%) particularly in the strongest updrafts with the largest supersaturations, however, the RH did not appreciably increase (not shown). This is the opposite result found by Furtado et al. (2014) who found little effect on IWC and instead found a significant change in RH, probably reflecting the differing dynamical situations of the two studies, with their cases being steady state ice only clouds.~~

The observed mean mass-weighted~~characteristic~~ ice diameter ~~size (mean mass weighted diameter)~~ shown in Figure 14~~7~~ increases with warmer temperatures and shows a strong dependence on IWC, with the characteristic size decreasing with increasing IWC reflecting the dominance of smaller particles for higher IWC. This contrasts with the lack of dependence of mean ice particle size on IWC that has been observed in earlier flights over Darwin and Cayenne in 2010 – 2012 (Fridlind et al. 2015) but agrees with more recent findings by Leroy et al. (2015). These findings show similar results to those documented by Gayet et al. (2012), with high concentrations of ice crystals occurring in regions of ice water content > 1 g m$^{-3}$ sustained for at least 100 s at Darwin (Leroy et al. 2015) and > 0.3 g m$^{-3}$ in the over shooting convection in the midlatitudes in Western Europe (Gayet et al. 2012). Gayet et al. (2012) proposed that the high concentration of ice crystals that appeared as chain-like aggregates of frozen drops, could be generated by strong updrafts lofting supercooled droplets that freeze homogeneously. However, using updraft parcel model simulations, Ackerman et al. (2015) showed that this process produced a smaller median mass area equivalent diameter than is observed. They proposed a number of other possible microphysical pathways to explain the observations, including the Hallett-Mossop process and a large source of heterogeneous ice nuclei coupled with the shattering of water droplets when they freeze.

The modelled mean snow diameter increases with increasing temperature, reflecting the process of aggregation, however, the modelled snow PSD also increases the mean diameter with increasing IWC, with the rate of increase being similar in both the generic ice PSD and the explicit specified gamma size distribution. The mean diameter from the generic ice PSD tends to agree reasonably well with the observed size for IWC < 0.5 g m$^{-3}$, however, the sizes are significantly overestimated for IWC > 0.5 g m$^{-3}$. Given that the number concentration is dependent on the size of the particles, for a given IWC~~,~~ this implies that the generic ice PSD simulates larg~~small~~er concentrations of larger particles~~ for a given IWC~~ than the observations ~~as shown previously in Figure 9~~. This reflects the data that was used to develop the generic ice PSD coming largely from stratiform clouds with smaller IWC and larger ice particles. The explicit gamma PSD shows the opposite behaviour, underestimating the mean ice diameter for IWC < 0.5 g $m^{-3}$ and matching the observed size for higher IWC. To ~~be able to correctl~~more accurately~~y~~ represent the snow sizes in the model for this case requires a double moment microphysics scheme to be able to better capture the observed variability of the PSD, or~~ a bimodal PSD parameterisation or~~ the use of a wider data set that includes high IWC observations to generate a more applicable generic ice PSD parameterisation for modelling tropical convective cloud systems.

## 4 Conclusions

A set of 1 km horizontal grid length simulations has been analysed to evaluate the ability of the UM to simulate tropical convective cloud systems and to investigate the impacts of different dynamical, turbulent and microphysical representations on the cloud properties, including the phase composition~~ and ice water contents~~. The case study is ~~for~~ February 18 2014 where active monsoon conditions produced a mesoscale convective system in the Darwin area. ~~The simulations reproduce the observed deep westerly winds that are the source of moisture for the long-lived cloud system, however, the simulations are too warm and dry below the freezing level and too warm and moist above this level, particularly in the upper troposphere. The simulation with the differing dynamical core is the least representative of the observed sounding, with the most accurate being the simulation with an additional ice prognostic and heterogeneous rain freezing parameterisation.~~

Analysing 12 hours of observed and simulated radar reflectivity has shown that the simulations capture the intensification and decay of convective strength associated with the lifecycle of the MCS. ~~, with the timing of the deepest convection represented well.~~ However, convection occurs too early in the simulations, the radar detectable cloud tops heights are overestimated~~ by the simulations~~, as are the maximum reflectivities and areas above the freezing level with reflectivities greater than 30 dBZ. The observed maximum domain averaged precipitation rate coincides with the generation of significant anvil cloud, whereas the simulations generate the highest mean precipitation rate a few hours too early at the times of deepest convection. O~~Aircraft o~~bservations of maximum vertical velocity suggest that the new dynamical core simulation ~~-~~overestimates the strength of convection at the mature-decaying stage of the MCS. In this case the stronger updrafts contribute to the excessive reflectivities above the freezing level, but this was apparent in all of the simulations albeit to a lesser degree, suggesting that both the updraft dynamics and the particle sizes are responsible for this error.~~. These strong convective updrafts will loft condensate, including large particles, into the upper troposphere where their subsequent freezing will release latent heat that will further drive the simulated updrafts.~~

~~In the observed reflectivity distribution there is evidence of the lofting of large particles up to 12 km, which is captured by a number of the simulations although the heights are above 15 km and the reflectivities larger than those observed by up to 20 dBZ.~~

The simulated reflectivity CFADs show more of a convective type profile compared to the observations, with broader distributions and a greater occurrence of high reflectivity outliers. This~~that~~ suggests a larger number of convective cells in the simulations, as was apparent in the plan views of OLR and 2.5 km radar reflectivity, which has been ~~.~~seen in tropical convective-scale model intercomparison studies (e.g. Varble et al. 2014a). The simulation with the differing turbulence parameterisation showed the best agreement with the observed maximum reflectivity at the later times of 23 – 24 UTC. The change to the 3D Smagorinsky scheme induces greater mixing ~~and more dilute convective plumes~~ resulting in a reduction of the maximum vertical velocities and reflectivities during the mature-decaying MCS stages. This same reduction in the vertical velocity and reflectivity up to 8 km was also found with a change to the microphysics formulation with the addition of a rain heterogeneous freezing parameterisation. At 17 – 18 UTC at the time of deepest convection, all simulations showed a similar error in maximum reflectivity regardless of dynamics or turbulence formulation due to the larger and less variable maximum updrafts across all of the simulations at these times.~~, and in fact the 3D Smagorinsky scheme produced the fastest 90th percentile updraft speed.~~

The largest sensitivities in the maximum updraft velocities are generally produced by changes to the dynamical and turbulence formulations in the model. However, the spread across the simulations for the mean and percentiles of updraft velocity show the greatest sensitivity coming from changes to the microphysical parameters and processes. Changing the microphysics affects the dynamics by altering the vertical distribution of latent heating~~, which drives the vertical motions~~. The horizontal mass divergence ~~and convective updraft buoyancy~~ was shown to be most sensitive to the turbulence parameterisation in the mixed-phase regions of the updrafts, where the greater mixing generated larger mass divergence, indicative of greater entrainment and a greater detrainment of mass at these heights. The upper ice-only regions of the convective updrafts showed that the control on updraft buoyancy was the sizes of the ice particles. Simulations with swith smaller particles have fewerreducing occurrences of positivelyupdraft buoyancy convective updrafts, y and limiting the cloud top heights, reflecting the importance of the microphysical processes on the convective dynamics.

The simulations that use an explicit ice PSD rather than the generic PSD parameterisation produce greater occurrences of larger reflectivities that more closely resemble the observations, although the modal reflectivity is overestimated. The reflectivity distributions as a function of height do not show the same slope with altitude when comparing the observations to the simulations using the generic ice PSD. Given that at the heights of 6 – 9 km the domain is almost completely covered by hydrometeors, this suggests that for the majority of occurrences the temperature dependency in the generic ice PSD and the implicit representation of aggregation is too weak. This can also be seen in the comparison of the particle mean diameters with the in situ observations where the explicit PSD for an IWC of 0.5 g $m^{-3}$ increases by about 2.6 times from the coolest to the warmest regime, while the generic ice PSD increases by 1.6 and the observations show more than a tripling in mean size. The beneficial impact of including a rain heterogeneous freezing parameterisation was shown through the reduction of large raindrops being advected above the freezing level, which was not observed by the radar or aircraft during the matures stage of the MCS and supports previous observations that show that most drops in oceanic convection freeze between -6 and -18 ºC (Stith et al. 2002). The simulation without graupel also overestimates the reflectivities at the melting level demonstrating that it is not only graupel that causes excessively large reflectivities but also snow in simulations that use a single moment microphysics scheme.

Analysing the relationship between phase composition and vertical velocity for 4 different temperature regimes shows that the LWC increases with increasing updraft strength, and as the temperature cools the LWC reduces along with the fraction of condensate that is supercooled liquid water. With increasing ascent the rate that the saturation specific humidity is lowered is increasingly faster than the rate that the liquid water can be reduced by deposition and riming of ice, resulting in an increase of LWC with vertical velocity. For the warmest temperature regimes the simulations with no or restricted graupel growth produced the greatest amount of IWC and lowest LWC for vertical velocities greater than 7 m $s^{-1}$. Ice in

~~these regimes with temperatures > –10 ºC has formed above and has been recirculated into these updrafts. The perturbed graupel cases have a larger amount of mass contained in the slower falling snow particles and this results in a more efficiency removal of LWC through increased in-cloud residence time and an increase in the accumulated ice water content.~~

Analysing the relationship between phase composition and vertical velocity for 4 different temperature regimes showed that the ~~The simulations show that the growth of liquid drops is more sensitive to the vertical velocity than the growth of ice particles, as has been documented previously (Korolev 2008). For the colder temperature regimes the simulations that use the explicit ice PSD rather than the generic ice PSD parameterisation tend to have more LWC, which is probably due to the reduced accretion and riming rates associated with the smaller particles. The three simulations that tended to produce more LWC for a given updraft strength for the colder regimes are the simulations with an increased cloud droplet number concentration, inclusion of an ice splintering parameterisation and inclusion of a heterogeneous rain freezing parameterisation. Increasing the cloud droplet number concentration reduces the precipitation efficiency of warm rain processes and generates more cloud water and a greater fraction of condensate being supercooled liquid water for temperatures between -10 and -30 ºC. In the model cloud water is the only liquid water used for depositional growth via the Bergeron-Findeisen mechanism and heterogeneous freezing, and the increased cloud water in this simulation produces the largest accumulation of graupel. Including a parameterisation of the secondary ice production Hallett-Mossop process that increases the deposition rate generates a larger quantity of ice, which through the increased latent heating supports the transport of more cloud liquid water and allows ice crystals and aggregates to be present across a wider range of updraft speeds. The other simulation with different behaviour and larger cloud LWC is the case that includes rain heterogeneous freezing. The impact of including this process in the model is to increase the cloud water content in strong updrafts due to the reduction in the accretion of cloud water by rain given the reduced rain water content.~~phase composition in the modelled convective updrafts is controlled by:

1. The size of the ice particles, with larger particles growing more efficiently through riming, producing larger IWC.

2. The efficiency of the warm rain process, with greater cloud water contents being available to support larger ice growth rates.

3. Exclusion or limitation of graupel growth, with more mass contained in slower falling snow particles resulting in an increase of in-cloud residence times and more efficient removal of LWC.

The evaluation of a tropical mesoscale convective system in this study has documented a number of model shortcomings and developments that improve the model performance:

1. Excessive areas with high reflectivities improve with reduced ice sizes, inclusion of a heterogeneous freezing rain parameterisation, an additional ice prognostic variable and increased turbulent mixing through the use of the 3D Smagorinsky turbulence scheme.
2. Too much rain above the freezing level is reduced with the inclusion of a heterogeneous rain freezing parameterisation.
3. T~~oo little entrainment with t~~oo little stratiform cloud and rain area is increased with increased turbulent mixing~~ and smaller ice sizes~~.
4. ~~Too efficient depositional growth of ice is improved with a reduction in depositional capacitance that includes the effects of non-spherical ice particles.~~

While the listed model changes do improve aspects of the simulations, none of these produce a simulation that closely matches all of the observations. This study has shown the need to include a better representation of the observed ~~bimodal~~ size distribution, which could be achieved through the use of a double moment microphysics scheme. Being able to predict both the number concentration and mass would allow the model to better represent the observed variability of the PSD, which would impact the model's representation of the ice water contents and reflectivities, as well as the convective dynamics through the effects of latent heating and water loading on buoyancy.

## Acknowledgements

This research has received funding from the Federal Aviation Administration (FAA), Aviation Research Division, and Aviation Weather Division, under agreement CON-I-2901 with the Australian Bureau of Meteorology. The research was also conducted as part of the European Union's Seventh Framework Program in research, technological development and demonstration under grant agreement n°ACP2-GA-2012-314314, and the European Aviation Safety Agency (EASA) Research Program under service contract n° EASA.2013.FC27.

Funding to support the development and testing of the isokinetic bulk TWC probe was provided by the FAA, NASA Aviation Safety Program, Environment Canada, and the National Research Council of Canada. Funding for the Darwin flight project was provided by the EU Seventh Framework Program agreement and EASA contract noted above, the FAA, the NASA Aviation Safety Program, the Boeing Co., Environment Canada, and Transport Canada. We acknowledge use of the MONSooN system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, which is a strategic partnership between the Met Office and the Natural Environment Research Council. We would like to express our thanks to Stuart Webster and Adrian Hill for providing the control model configuration, and to Paul Field for suggesting the analysis presented in Figure 9. The satellite data were provided by the NASA Langley group led by Pat Minnis. The RASTA cloud radar vertical velocity retrieval was generously provided by Julien Delanoë. We thank two anonymous reviewers for comments and suggestions that improved the manuscript.

~~The RASTA vertical velocity retrieval was generously provided by Julien Delanoë.~~

## References

Abel, S. and I.A. Boutle, 2012: An improved representation of the rain drop size distribution for single-moment microphysics schemes, *Q. J. Roy. Meteor. Soc.*, **138**, 2151-2162

Abel, S. and B.J. Shipway, 2007: A comparison of cloud-resolving model simulations of trade wind cumulus with aircraft observations taken during RICO. *Q. J. R. Meteorol. Soc.*, **133**, 781 – 794

Ackerman, A.S., A.M. Fridlind, A. Grandin, F. Dezitter, M. Weber, J.W. Strapp, A.V. Korolev, 2015: High ice water content at low radar reflectivity near deep convection. Part 2: Evaluation of microphysical pathways in updraft parcel simulations. *Atmos. Chem. Phys.*, **15**, 11729 - 11751

Bange, J., Esposito, M., Lenschow, D. H., Brown, P. R. A., Dreiling, V., Giez, A., Mahrt, L., Malinowski, S. P., Rodi, A. R., Shaw, R. A., Siebert, H., Smit, H. and Zöger, M. (2013) Measurement of Aircraft State and Thermodynamic and Dynamic Variables, in Airborne Measurements for Environmental Research: Methods and Instruments (eds M. Wendisch and J.-L. Brenguier), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. doi: 10.1002/9783527653218.ch2

Barnes, H.C and R.A. Houze Jr., 2014: Precipitation hydrometeor type relative to mesoscale airflow in mature oceanic deep convection of the Madden-Julian Oscillation. *J. Geophys. Res. Atmos.*, doi: 10.1002/2014JD022241

Baumgardner, D., et al., 2001. The cloud, aerosol and precipitation spectrometer (CAPS): a new instrument for cloud investigations. *Atmos. Res.*, **59–60**, 251–264

Black, R. and J. Hallett, 1999: Observations of the distribution of ice in hurricanes. *J. Atmos. Sci.*, **43**, 802 – 822

Bigg, E.K., 1953: The supercooling of water. *Proc. Phys. Soc. London*, **B66**, 688-694

Blossey, P.N., C.S. Bretherton, J. Cetrone and M. Kharoutdinov, 2007: Cloud-resolving model simulations of KWAJEX: Model sensitivities and comparisons with satellite and radar observations. *J. Atmos. Sci.*, **64**, 1488 – 1508

Boutle, I.A. J.E.J. Eyre and A.P. Lock, 2014: Seamless stratocumulus simulation across the turbulent gray zone. *Mon. Wea. Rev.*, **142**, 1655 - 1668

Bretherton, C.S., P.N. Blossey and M. Khairoutdinov, 2005: An energy-balance analysis of deep convective self-aggregation above uniform SST. *J. Atmos. Sci.*, **62**, 4273 – 4292

Cardwell, J.R., T.W. Choularton, D. Wilson and R. Kershaw, 2002: Use of an explicit model of the microphysics of precipitating stratiform cloud to test a bulk microphysics scheme. *Q. J. R. Meteorol. Soc.*, **128**, 573 – 592

Clark, A. J., W. A. Gallus, and T.-C. Chen, 2007: Comparison of the diurnal precipitation cycle in convection-resolving and non-convection-resolving mesoscale models, *Mon. Wea. Rev.*, **135**, 3456–3473

Collis, S., A. Protat, P.T. May and C. Williams, 2013: Statistics of storm updraft velocities from TWP-ICE including verification with profiling measurements. *J. App. Meteor.*, **52**, 1909 – 1922

Cox, G.P., 1988: Modelling precipitation in frontal rainbands. *Q. J. R. Meteorol. Soc.*, **114**, 115 – 127

Davies, T., M.J.P. Cullen, A.J. Malcolm, M.H. Mawson, A. Staniforth, A.A. White and N. Wood, 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.*, **131**, 1759 – 1782

Davison, C. R., J.D. MacLeod, and J.W. Strapp, 2009: Naturally Aspirating Isokinetic Total Water Content Probe: Evaporator Design and Testing. *1st AIAA Atmospheric and Space Environments*, June 25, 2009, San Antonio, Texas, AIAA-2009-3861.

Del Genio, A.D. and J. Wu, 2010: The role of entrainment in the diurnal cycle of continental convection. *J. Clim.*, **23**, 2722-2738

Edwards, J.M. and A. Slingo, 1996: Studies with a new flexible radiation code. I: Choosing a configuration for a large-scale model. *Q. J. R. Meteorol. Soc.*, **122**, 689 – 720

~~Fein, J.S. and P.L. Stephens, 1987: *Monsoons.* Wiley, 632pp.~~

Ferrier, B.S., 1994: A double-moment multiple-phase four-class bulk ice scheme. Part I: Description. *J. Atmos. Sci.*, **51**, 249 – 280

Field, P.R., A.J. Heymsfield and A. Bansemer, 2007: Snow size distribution parameterisation for midlatitude and tropical ice clouds. *J. Atmos. Sci.*, **64**, 4346 – 4365

Fletcher, N.H., 1962: *The Physics of Rain Clouds.* Cambridge University Press, 386 pp.

Franklin, C.N., G.J. Holland and P.T. May, 2005: Sensitivity of tropical cyclone rainbands to ice-phase microphysics. *Mon. Weather Rev.*, **133**, 2473 – 2493

Franklin, C.N., Z. Sun, D. Bi, M. Dix, H. Yan and A. Bodas-Salcedo, 2013: Evaluation of clouds in ACCESS using the satellite simulator package COSP: Global, seasonal and regional cloud properties. *J. Geophys. Res. Atmos.*, **118**, 732 – 748

Fridlind, A.M., A.S. Ackerman, J.-P. Chaboureau, J. Fan, W.W. Grabowski, A.A. Hill, T.R. Jones, M.M. Khaiyer, G. Liu, P. Minnis, H. Morrison, L. Nguyen, S. Park, J.C. Petch, J.-P. Pinty, C. Schumacher, B.J. Shipway, A.C. Varble, X. Wu, S. Xie and M. Zhang, 2012: A comparison of TWP-ICE observational data with cloud-resolving model results. *J. Geophys. Res. Atmos.*, **117,** D05204, doi:10.1029/2011JD016595

Fridlind, A.M., A.S. Ackerman, A. Gandin, F. Dezitter, M. Weber, J.W. Strapp, A. V. Korolev and C.R. Williams, 2015: High ice water content at low radar reflectivity near deep convection – Part 1: Consistency of in situ and remote-sensing observations with stratiform rain column simulations. *Atmos. Chem. Phys. Discuss.*, **15**, 16505 - 16550

Furtado, K., P.R. Field, R. Cotton and A.J. Baran 2014: The sensitivity of simulated high clouds to ice crystal fall speed, shape and size distribution. *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.2457

Gayet, J.-F., G. Mioche, L. Bugliaro, A. Protat, A. manikin, M. Wirth, A. Dornbrack, V. Shcherbakov, B. Mayer, A. Garnier and C. Gourbeyre, 2012: On the observation of unusual high concentration of small chain-like aggregate ice crystals and large ice water contents near the top of a deep convective cloud during the CIRCLE-2 experiment. *Atmos. Chem. Phys.*, **12**, 727 - 744

Gregory, D. and P.R. Rowntree, 1990: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea. Rev.*, **118**, 1483-1506

Hallett, J. and S.C. Mossop, 1974: Production of secondary ice particles during the riming process. *Nature*, **249**, 26 – 28

Hanley, K.E., R.S. Plant, T.H.M. Stein, R.J. Hogan, J.C. Nicol, H.W. Lean, C. Halliwell and P.A. Clark, 2014: Mixing-length controls on high-resolution simulations of convective storms. *Q. J. R. Meteorol. Soc.*,doi:10.1002/qj.2356

Heymsfield, A. J., A. Protat, R. T. Austin, D. Bouniol, J. Delanoë, R. Hogan, H. Okamoto, K. Sato, G.-J. van Zadelhoff, D. Donovan, and Z. Wang, 2008: Testing and evaluation of ice water content retrieval methods using radar and ancillary measurements. J. Appl. Meteor. Climate, **47**, 135-163

Heymsfield, A.J., A. Bansemer, G. Heymsfield and A.O. Fierro, 2009: Microphysics of maritime tropical convective updrafts at temperatures from -20° to -60°. *J. Atmos. Sci.*, **66**, 3530 – 3562

Heymsfield, A.J. and P. Willis, 2014: Cloud conditions favouring secondary ice particle production in tropical maritime convection. *J. Atmos. Sci.*, **71**, 4500 - 4526

Heymsfield, G.M., L. Tian, A.J. Heymsfield, L. Li and S. Guimond, 2010: Characteristics of deep tropical and subtropical convection from nadir-viewing high-altitude airborne Doppler radar. *J. Atmos. Sci.*, **67**, 285 – 308

Hogan, R.J., P.R. Field, A.J. Illingworth, R.J. Cotton and T.W. Choularton, 2002: Properties of embedded convection in warm-frontal mixed-phase cloud from aircraft and polarimetric radar. *Q. J. R. Meteorol. Soc.*, **128**, 451 – 476

Hogan, R.J., M.P. Mittermaier and A.J. Illingworth, 2006: The retrieval of ice water content from radar reflectivity factor and temperature and its use in evaluating a mesoscale model. *J. Appl. Meteorol.*, **45**, 301 - 317

Houze, R.A., P.V. Hobbs, P.H. Herzegh and D.B. Parsons, 1979: Size distributions of precipitation particles in frontal clouds. *J. Atmos. Sci.*, **36**, 156 – 162

Jung, S.-A., D.-I. Lee, B. Jou and H. Uyeda, 2012: Microphysical properties of maritime squall line observed on June 2, 2008 in Taiwan. *Journal of the Meteorological Society of Japan*, **90**, 83~~1~~3-8~~2~~50

Keenan, T.D., K. Glasson, F. Cummings, T.S. Bird, J. Keeler and J. Lutz, 1998: The BMRC/NCAR C-band polarimetric (C-POL) radar system. *J. Atmos. Oceanic Technol.*, **15**, 871 - 886

~~Korolev, A.V., 2008: Rates of phase transformations in mixed-phase clouds. *Q. J. R. Meteorol. Soc.*, **134**, 595 – 608~~

Lang, S.E., W.-K. Tao, X. Zeng and Y. Li, 2011: Reducing the biases in simulated radar reflectivities from a bulk microphysics scheme: Tropical convective systems. *J. Atmos. Sci.*, **68**, 2306 – 2320

Lawson, R. P., L. J. Angus, and A. J. Heymsfield, 1998: Cloud particle measurements in thunderstorm anvils and possible threat to aviation, J. Aircraft, 35(1), 113-121

Lawson, R. P., D. O'Connor, P. Zmarzly, K. Weaver, B. A. Baker, Q. Mo, and H. Jonsson, 2006: The 2D-S (Stereo) probe: Design and preliminary tests of a new airborne, high speed, high-resolution particle imaging probe, *J. of Atmos. Oceanic Technol.*, **23**, 1462-1477

Leroy, D., E. Fontaine, A. Schwarzenboeck, J.W. Strapp, L. Lilie, J. Delanoë, A. Protat, F. Dezitter and A. Grandin, 2015: HAIC/HIWC field campaign-specific findings on PSD microphysics in high IWC regions from in situ measurements: Median mass diameters, particle size distribution characteristics and ice crystal shapes. Tech. Rep. 2015-01-2087, SAE International, Warrendale, PA, USA, doi:10.4271/2015-01-2087

Liu, G. and J. Curry, 1999: Remote sensing of ice water characteristics in tropical clouds using aircraft microwave measurements. *J. App. Meteor.*, **37**, 337 – 355

Lock, A.P., A.R. Brown, M.R. Bush, G.M. Martin and R.N.B. Smith, 2000: A new boundary-layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Weather Rev.*, **128**, 3187 – 3199

Marshall, J.S. and W.M.K. Palmer, 1948: The distribution of raindrops with size. *Journal of Meteorology*, **5**, 165 – 166

Mason, J. G., J. W. Strapp, and P. Chow, 2006: The ice particle threat to engines in flight. 44th AIAA Aerospace Sciences Meeting, Reno, Nevada, 9-12 January 2006, AIAA-2006-206. Available online at http://arc.aiaa.org/doi/abs/10.2514/6.2006-206

May, P.T. and T. Lane, 2009: A method for using weather radar data to test cloud resolving models. *Meteorological Applications*, **16**, 425 – 432

May, P.T. and D.K. Rajopadhyaya, 1999: Vertical velocity characteristics of deep convection over Darwin, Australia. *Mon. Weather Rev.*, **127**, 1056 – 1071

McBeath, K., P.R. Field and R.J. Cotton, 2014: Using operational weather radar to assess high-resolution numerical weather prediction over the British Isles for a cold air outbreak. *Q. J. R. Meteorol. Soc.*, **140**, 225 – 239

Minnis, P. and W.L. Smith Jr., 1998: Cloud and radiative fields derivedfrom GOES-8 during SUCCESS and the ARM-UAV spring 1996 flight series, *Geophys. Res. Lett.*, **25,** 1113–1116.

Minnis, P. et al. 2008: Cloud detection in non-polar regions for CERES using TRMM VIRS and Terra and Aqua MODIS data, IEEE Trans. *Geosci. Remote Sens*., **46**, 3857–3884.

Minnis, P., et al. 2011: CERES Edition 2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data–Part I: Algorithms, *IEEE Trans. Geosci. Remote Sens*., **11**, 4374–4400, doi:10.1109/TGRS.2011.2144601.

Mitchell, D.L., 1996: Use of mass- and are-dimensional power laws for determining precipitation particle terminal velocities. *J. Atmos. Sci.*, **53**, 1710 – 1722

Morrison, H., G. Thompson and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Weather Rev.*, **137**, 991 – 1007

Protat, A., and C. R. Williams, 2011: The Accuracy of Radar Estimates of Ice Terminal Fall Speed from Vertically Pointing Doppler Radar Measurements. Journal of Applied Meteorology and Climatology, **50**, 2120–2138

Protat, A., and I. Zawadzki, 1999: A Variational Method for Real-Time Retrieval of Three-Dimensional Wind Field from Multiple-Doppler Bistatic Radar Network Data. Journal of Atmospheric and Oceanic Technology, **16**, 432–449

Puri, K. et al, 2013: Implementation of the initial ACCESS numerical weather prediction system. *Aust. Meteorol. Oceanogr. J.*, **63**, 265-284

Rosenfeld, D. et al. 2008: Flood or drought: How do aerosols affect precipitation? *Science* **321**(5894):1309–1313

Shige, S., Y.N. Takayabu, S. Kida, W.-K. Tao, X. Zeng, C. Yokoyama and T. L'Ecuyer, 2009: Spectral retrieval of latent heating profiles from TRMM PR data. Part IV: Comparison of lookup tables from two- and there-dimensional cloud-resolving model simulations. *J. Clim.*, **22**, 5577-5594

Smagorinsky, J., 1963: General circulation experiments with the primitive equations. I: The basic experiment. *Mon. Weather Rev.*, **91**, 99 – 164

Smith, RN.B., 1990: A scheme for predicting layer clouds and their water contents in a general circulation model. *Q. J. R. Meteorol. Soc.*, **116**, 435 – 460

Stith, J.L., J.E. Dye, A. Bansemer and A.J. Heymsfield, 2002: Microphysical observations of tropical clouds. *J. App. Meteor.*, **41**, 97 – 117

Stith, J.L., A. Hagerty, A.J. Heymsfield and C.A. Grainger, 2004: Microphysical characteristics of tropical updrafts in clean conditions. *J. App. Meteor.*, **43**, 779 – 794

Strapp, J. W., G. A. Isaac. A. Korolev, T. Ratvasky, R. Potts, P. May, A. Protat, P. Minnis, A. Ackerman, A. Fridlind, J. Haggerty, and J. Riley, 2015: The High Ice Water Content (HIWC) Study of deep convective clouds: Science and technical plan. FAA Rep. DOT/FAA/TC-14/31, in press

Tao, W.-K., J.R. Scala, B. Ferrier and J. Simpson, 1995: The effect of melting processes on the development of a tropical and midlatitude squall line. *J. Atmos. Sci.*, **52**, 1934 – 1948

VanWeverberg, K., A.M. Vogelmann, W. Lin, E.P. Luke, A. Cialella, P. Minnis, M. Khaiyer, E.R. Boer and M.P. Jensen, 2013: The role of cloud microphysics parameterisation in the simulation of mesoscale convective system clouds and precipitation in the Tropical Western Pacific. *J. Atmos. Sci.*, **70**, 1104 – 1181

Varble, A., A.M. Fridlind, E.J. Zipser, A.S. Ackerman, J.-P. Chaboureau, J. Fan, A. Hill, S.A. McFarlane, J.-P. Pinty and B. Shipway, 2011: Evaluation of cloud-resolving model intercomparison simulations using TWP-ICE observations: Precipitation and cloud structure. *J. Geophys. Res. Atmos.,* **116,** doi:10.1029/2010JD015180

Varble, A., E.J. Zipser, A.M. Fridlind, P. Zhu, A.S. Ackerman, J.-P. Chaboureau, S. Collis, J. Fan, A. Hill and B. Shipway, 2014: Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations. Part I: Deep convective updraft properties. *J. Geophys. Res. Atmos.*, **119,** 13891 – 13918

Varble, A., E.J. Zipser, A.M. Fridlind, P. Zhu, A.S. Ackerman, J.-P. Chaboureau, J. Fan, A. Hill, B. Shipway and C. Williams, 2014: Evaluation of cloud-resolving and limited area model intercomparison simulations using TWP-ICE observations. Part 2: Precipitation microphysics. *J. Geophys. Res. Atmos.*, **119,** doi:10.1002/2013JD021372

Waliser, D.E. et al., 2009: Cloud ice: A climate model challenge with signs and expectations of progress. *J. Geophys. Res*., **114**, D00A21, doi:10.1029/2008JD010015

Walters, D. N., Brooks, M. E., Boutle, I. A., Melvin, T. R. O., Stratton, R. A., Bushell, A. C., Copsey, D., Earnshaw, P. E., Gross, M. S., Hardiman, S. C., Harris, C. M., Heming, J. T., Klingaman, N. P., Levine, R. C., Manners, J., Martin, G. M., Milton, S. F., Mittermaier, M. P., Morcrette, C. J., Riddick, T. C., Roberts, M. J., Selwood, P. M., Tennant,W.J., Vidale, P.-L.,Wilkinson, J. M., Wood, N., Woolnough, S. J., and Xavier, P. K.: The Met Office Unified Model Global Atmosphere 6.0 and JULES Global Land 6.0 configurations, in preparation, 2015.

Weusthoff, T., F. Ament, M. Arpagaus and M.W. Rotach, 2010: Assessing the benefits of convection-permitting models by neighbourhood verification: Examples from MAP D-PHASE. *Mon. Wea. Rev*., **138**, 3418–3433

Wilkinson, J.M., 2013: *The Large-Scale Precipitation Parameterisation Scheme,* Unified Model Documentation Paper 26, Met Office, Exeter, UK. http://collab.metoffice.gov.uk/twiki/pub/Support/Umdp/026_84.pdf

Wilson, D.R. and S.P. Ballard, 1999: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.*, **125**, 1607 – 1636

Wilson, D.R., A.C. Bushell, A.M. Kerr-Munslow, D.P. Jeremy and C.J. Morcrette, 2008: PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Q. J. R. Meteorol. Soc.*, **134**, 2093 – 2107

Wisner, C., H.D. Orville and C. Myers, 1972: A numerical model of a hail bearing cloud. *J. Atmos. Sci.*, **29**, 1160 – 1181

Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M. and Thuburn, J., 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. Q.J.R. Meteorol. Soc., 140: 1505–1520. doi:10.1002/qj.2235

Yuter, S.E. and R.A. Houze Jr., 1995: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part III: Vertical mass transport, mass divergence, and synthesis. *Mon. Wea. Rev.*, **123**, 1964 – 1983

Zipser, E.J., D.J. Cecil, C. Liu, S.W. Nesbitt and D.P. Yorty, 2006: Where are the most intense thunderstorms on Earth? *Bull. Amer. Meteor. Soc.*, **87**, 1057 - 1071

Table 1. Parameters used to define the mass-diameter relationships (1) and particle size distributions (2), where $f(T)$ is given by (3).

| Parameter | Units | Rain | Aggregates | Crystals | Graupel |
|---|---|---|---|---|---|
| $a$ | kg $m^{-b}$ | 523.56 | 2.3 x $10^{-2}$ | 2.3 x $10^{-2}$ | 261.8 |
| $b$ | | 3.0 | 2.0 | 2.0 | 3.0 |
| $N_0$ | $m^{-4}$ | $0.22\lambda^{2.2}$ | 2 x $10^6$ $f(T)$ | 40 x $10^6$ $f(T)$ | 5 x $10^{25}\lambda^{-4}$ |
| μ | | 0 | 0 | 0 | 2.5 |

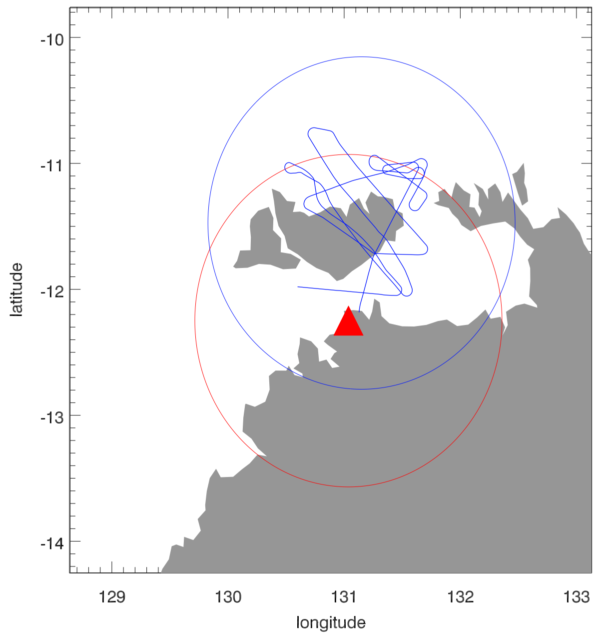
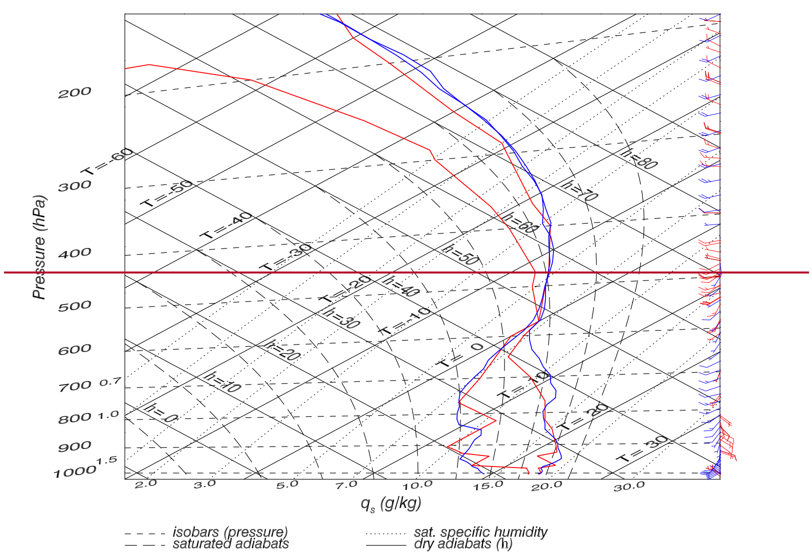



Figure 1. 1 km simulation domain with the radar location denoted by the red triangle and the 150 km range of the radar shown by the red circle. The aircraft flight track is shown by the blue line with the domain used in the aircraft comparison given by the blue circle.



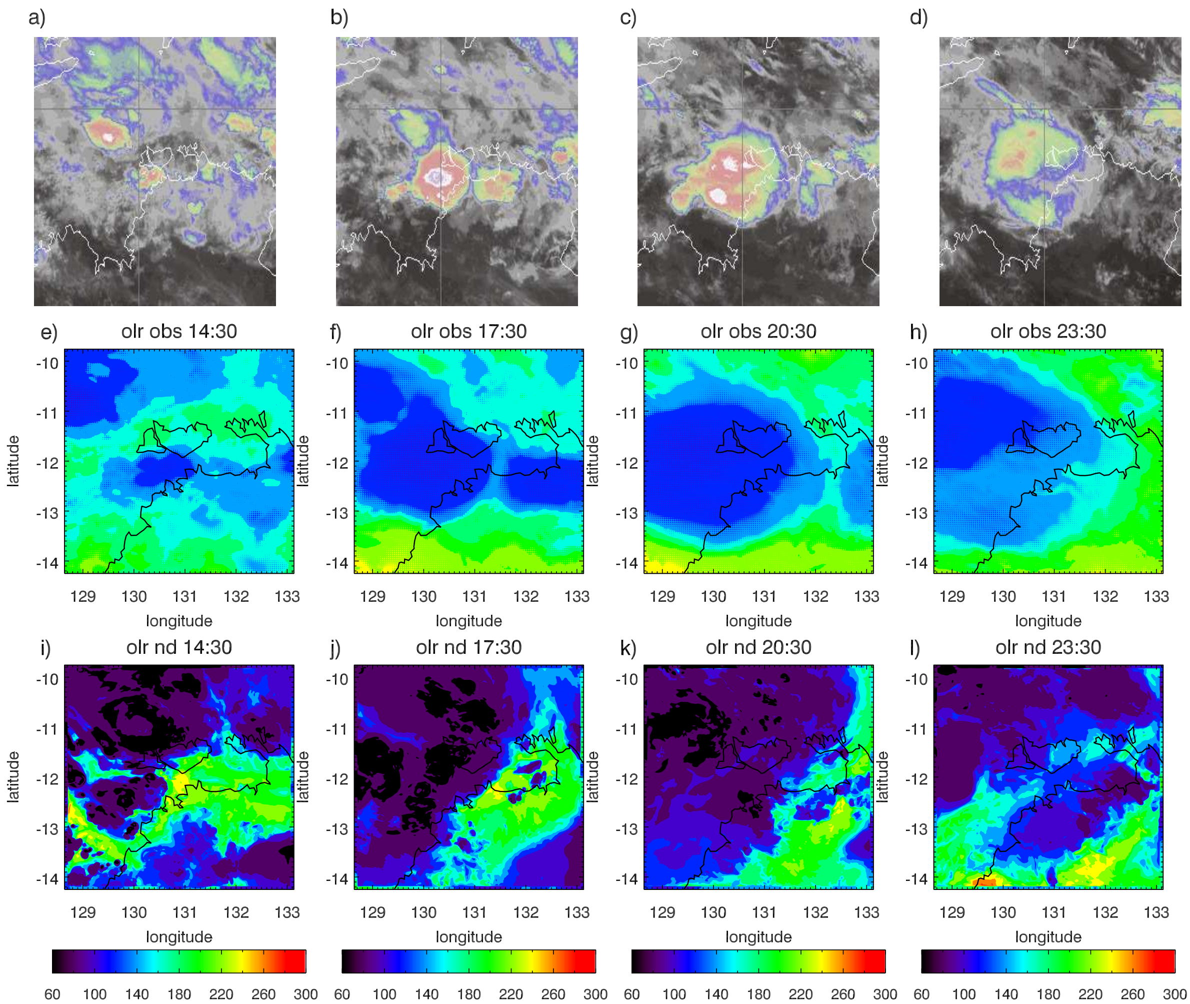
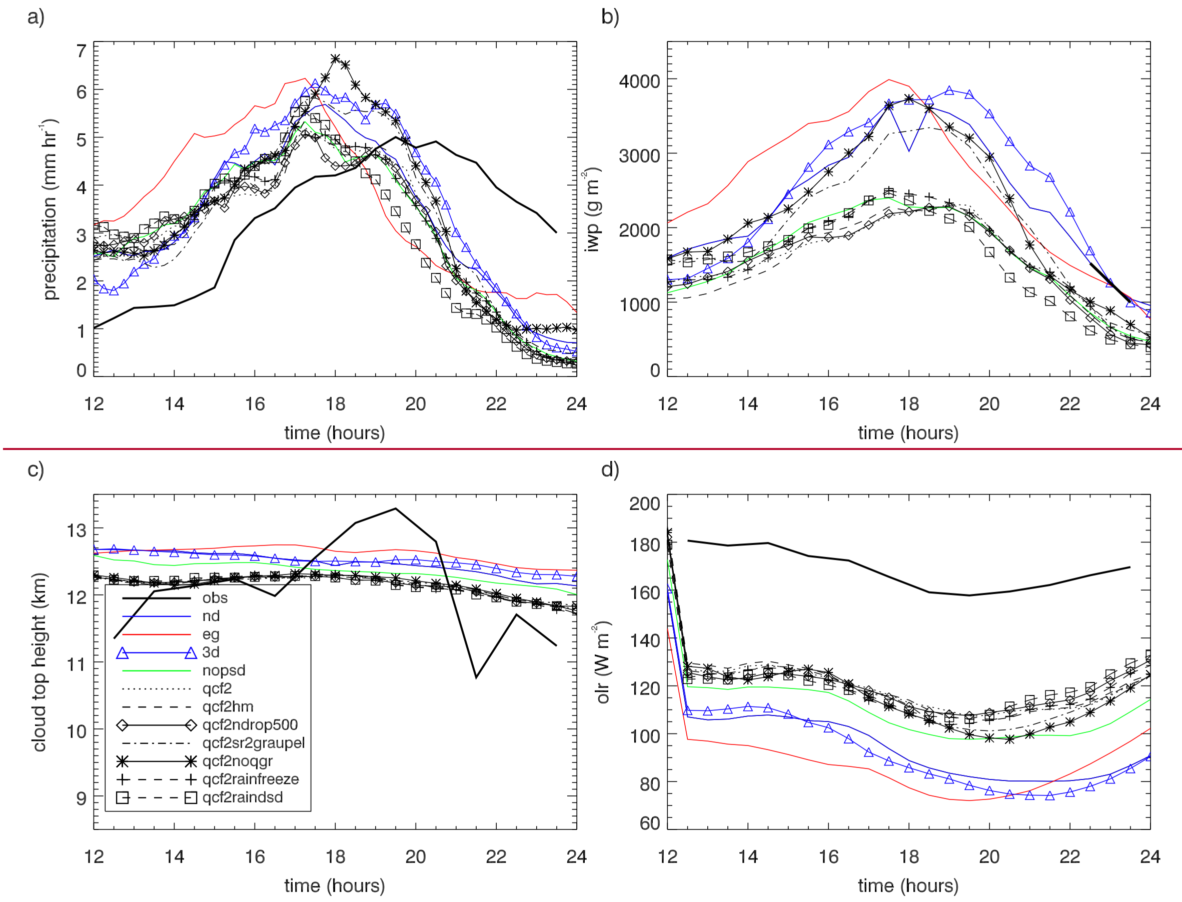




Figure 2. ~~Thermodynamic diagram showing the observed (red) and modelled (blue) temperature, dew point temperature and winds for Darwin at 23 UTC 18/02/2014. A short/long wind barb represents 5/10 knots.~~ Top row: time series of enhanced infrared satellite

imagery over the Darwin region on 18/02/2014 a) 14:30, b) 17:30, c) 20:30 and d) 23:30 UTC. Middle row: time series of observed

outgoing longwave radiation centred on the Darwin radar, where the pixel level satellite data has been interpolated onto the 1 km model grid. Last row: as above, but for the modelled outgoing longwave radiation from the control experiment labelled nd.


a)
b)
c)
d)
precipitation (mm hr⁻¹)
iwp (g m⁻²)
cloud top height (km)
olr (W m⁻²)
time (hours)
obs
nd
eg
3d
nopsd
qcf2
qcf2hm
qcf2ndrop500
qcf2sr2graupel
qcf2noqgr
qcf2rainfreeze
qcf2raindsd

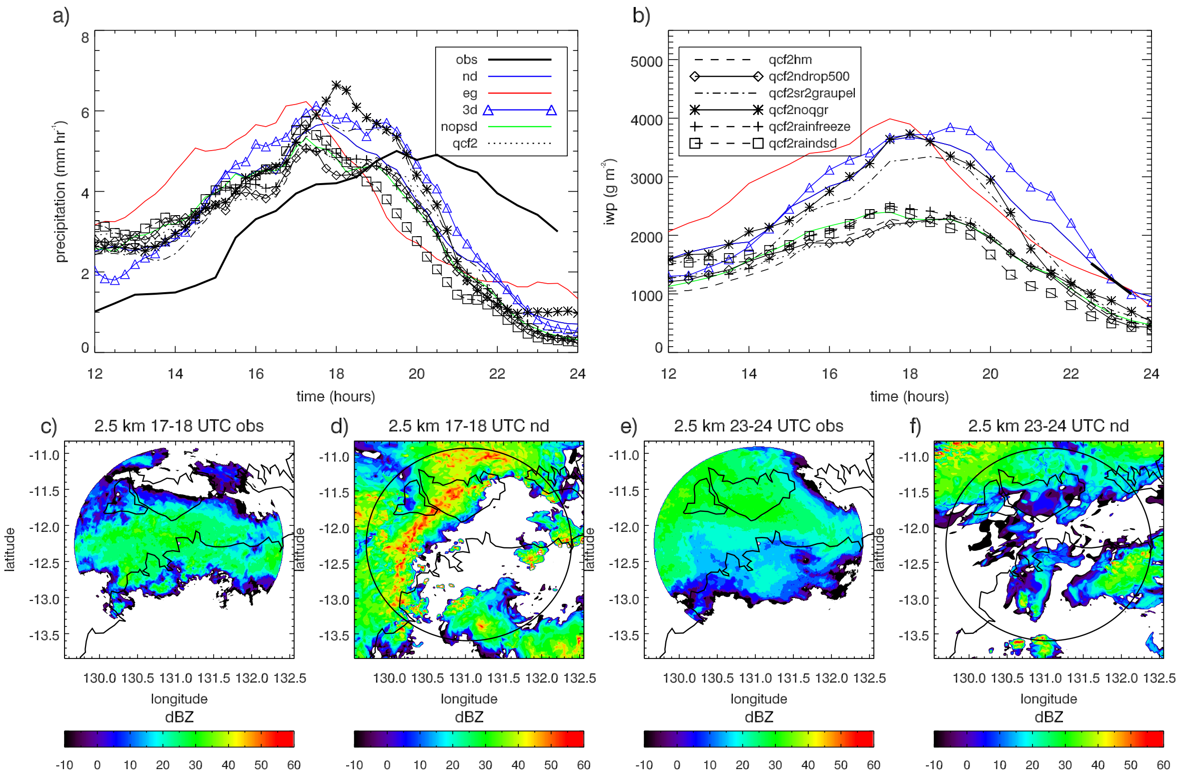
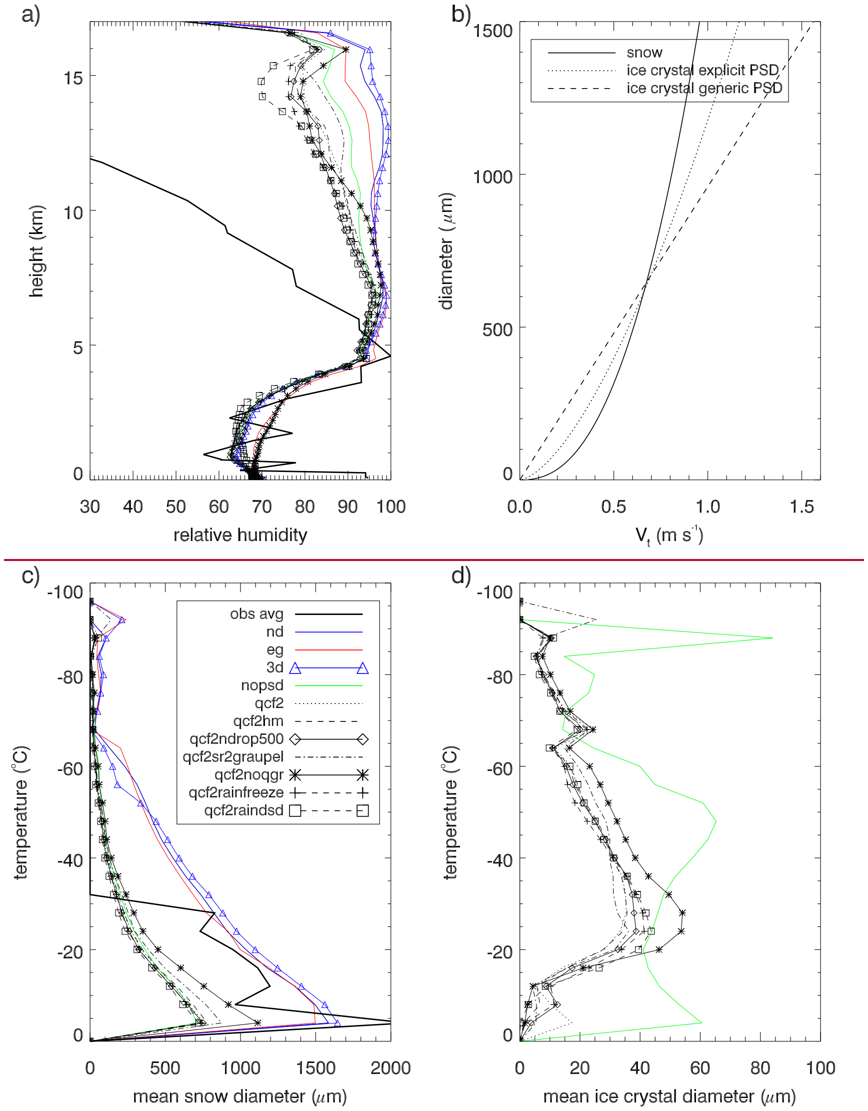

a)
b)
precipitation (mm hr⁻¹)
iwp (g m⁻²)
time (hours)
obs
nd
eg
3d
nopsd
qcf2
qcf2hm
qcf2ndrop500
qcf2sr2graupel
qcf2noqgr
qcf2rainfreeze
qcf2raindsd
c) 2.5 km 17-18 UTC obs
d) 2.5 km 17-18 UTC nd
e) 2.5 km 23-24 UTC obs
f) 2.5 km 23-24 UTC nd
latitude
longitude
dBZ

Figure 3. Time series of domain mean a) precipitation (mm hr$^{-1}$) and, b) ice water path (g m$^{-3}$), c) cloud top height (km) and d) outgoing longwave radiation (W m$^{-2}$). The observations are from the CPOL radar in a) and the satellite retrievals b),in the other panels (note that the observed IWP is only plotted from 22:30 – 23:30). The time period spans 12 – 24 UTC on 18/02/2014. c) 2.5 km observed radar reflectivity averaged over 17 – 18 UTC, d) as in c) except for the modelled reflectivity from the control simulation (nd), e) as in c) except for 23 – 24 UTC, d) as in d) except for 23 – 24 UTC.

1 
a)
b)
c)
d)
height (km)
relative humidity
diameter (μm)
$V_t$ (m s$^{-1}$)
snow
ice crystal explicit PSD
ice crystal generic PSD
temperature (°C)
mean snow diameter (μm)
mean ice crystal diameter (μm)
obs avg
nd
eg
3d
nopsd
qcf2
qcf2hm
qcf2ndrop500
qcf2sr2graupel
qcf2noqgr
qcf2rainfreeze
qcf2raindsd

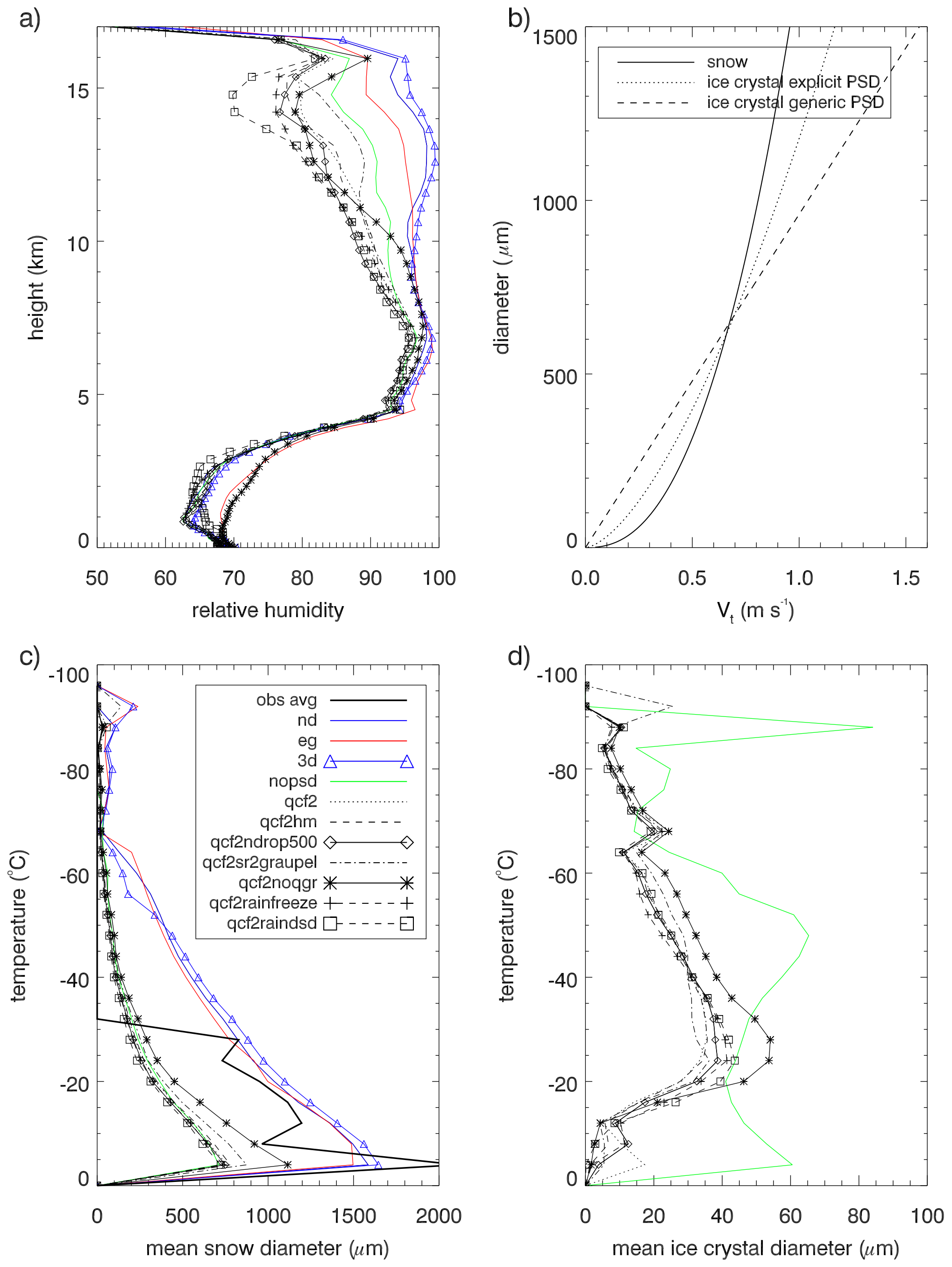
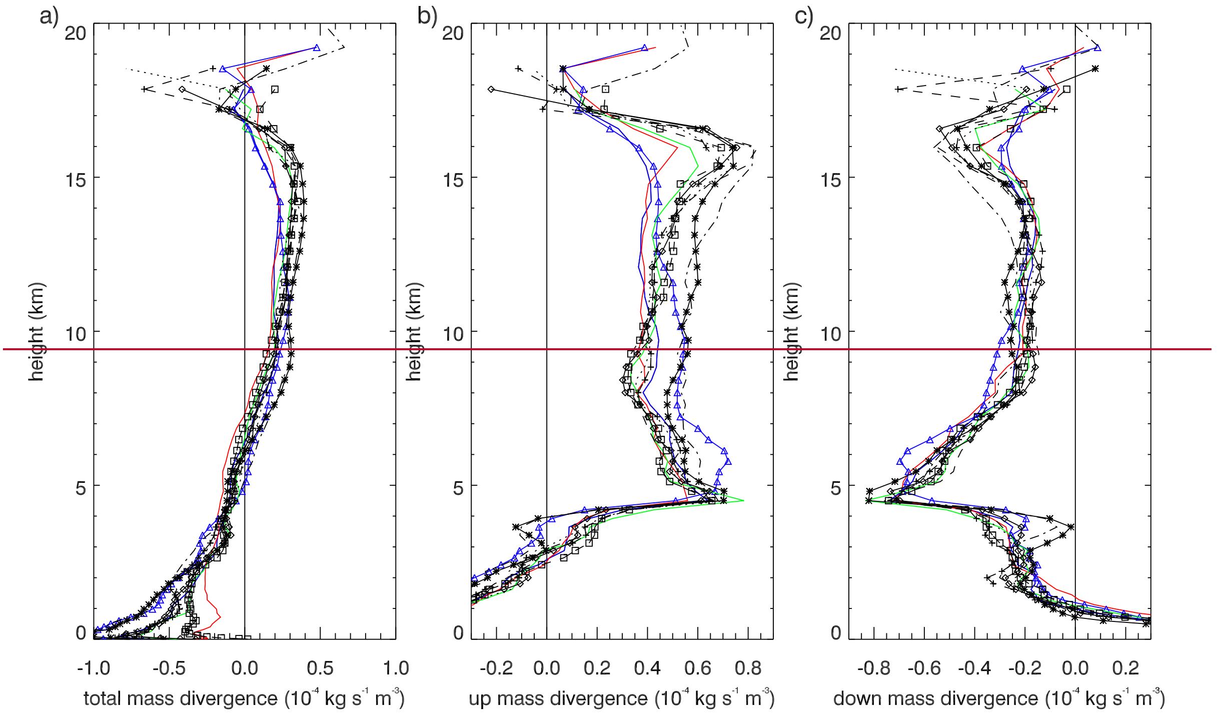



Figure 4. a)~~ Observed relative humidity at Darwin on 18/02/2014 at 23 UTC in solid black line. ~~Simulated relative humidity is for the area encompassed by the 150 km radius centred on the Darwin radar on 18/02/2014 from 23 – 24 UTC. b) Ice fall speeds (m s$^{-1}$) as a function of diameter (µm) for the snow category and the ice crystals used in the simulations with the explicit and generic PSD, see text for details. c) Mean mass-weighted snow diameter (µm) as a function of temperature ($^{o}$C) where the observations are from the aircraft and have been

averaged to be representative of a 1 $km^2$ grid cell. d) As for c) except for the mean mass-weighted ice crystal diameter (µm).



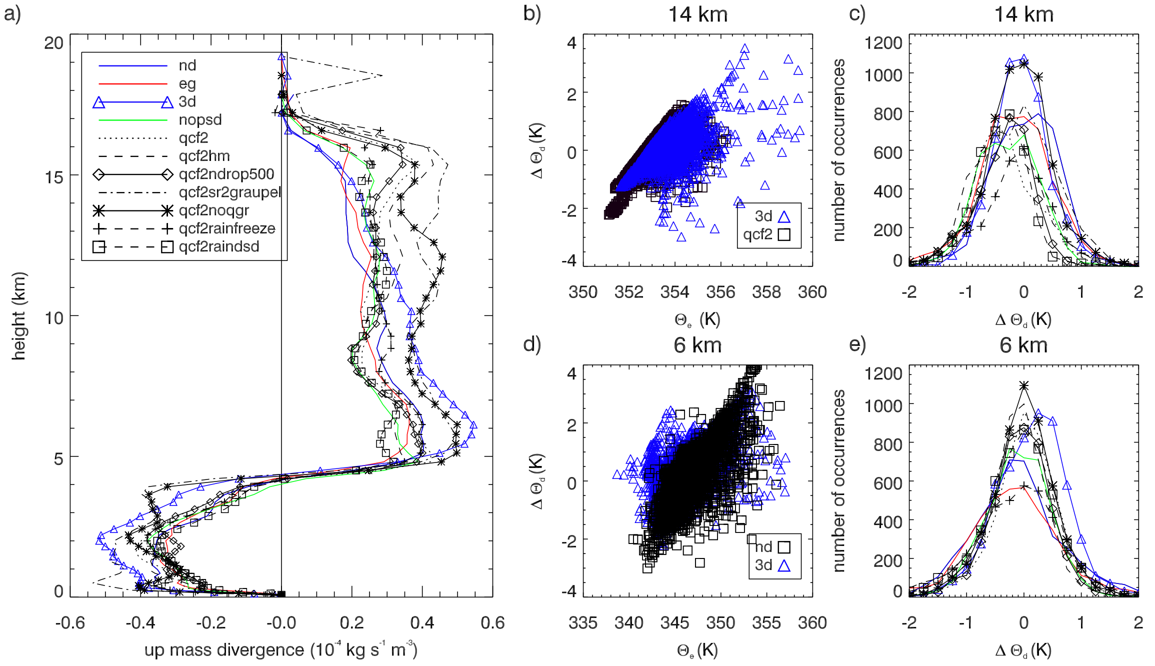
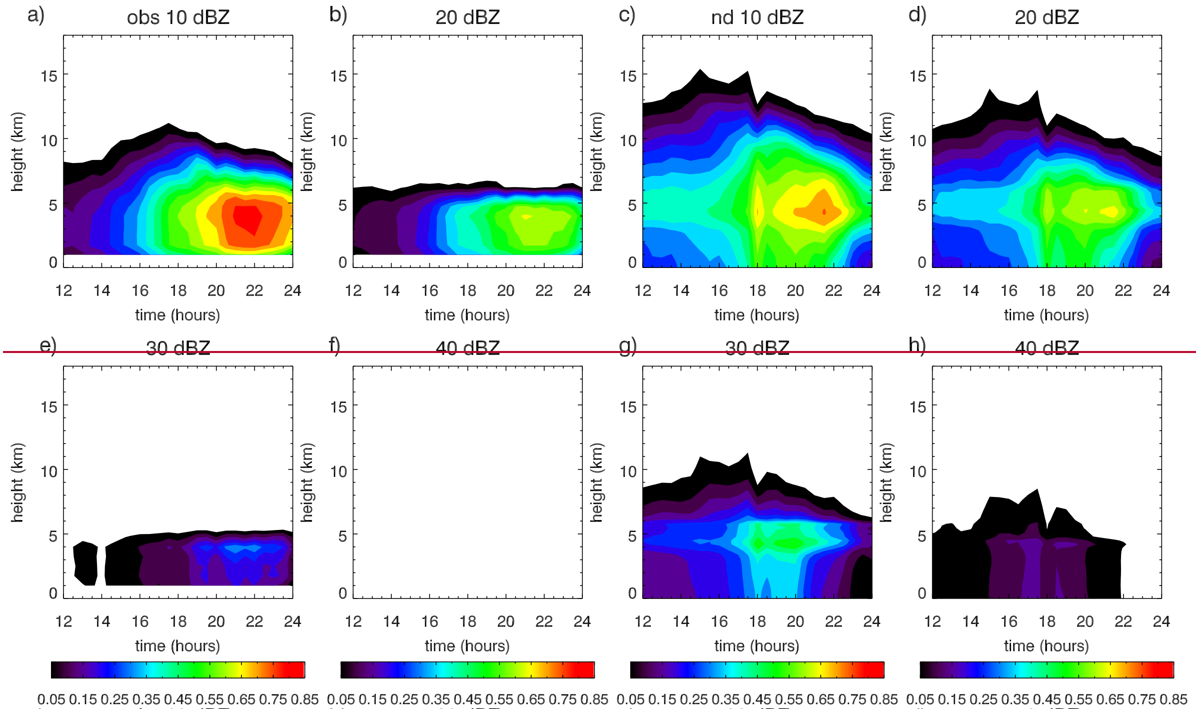




Figure 5. a) Vertical profile of convective updraft (> 1 m s$^{-1}$) mean horizontal mass divergence ($10^{-4}$ kg s$^{-1}$ m$^{-3}$) at 18 UTC. b) scatterplot of $\theta_e$ against $\Delta\theta_d$ at 14 km for two simulations that change the turbulent mixing (3d) and add an additional ice prognostic variable and have smaller ice sizes (qcf2). c) Histogram of $\Delta\theta_d$ at 14 km. d) As in b) except for 6km and comparing the control (nd) and the 3d simulations, and e) as in c) except for 6 km. See text for details. ~~, b) mean for the upwards vertical velocity, and c) for the downwards~~

~~vertical velocity. The legend for the simulations is as in Figure 3.~~


a) obs 10 dBZ
b) 20 dBZ
c) nd 10 dBZ
d) 20 dBZ
e) 30 dBZ
f) 40 dBZ
g) 30 dBZ
h) 40 dBZ
height (km)
time (hours)
0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85

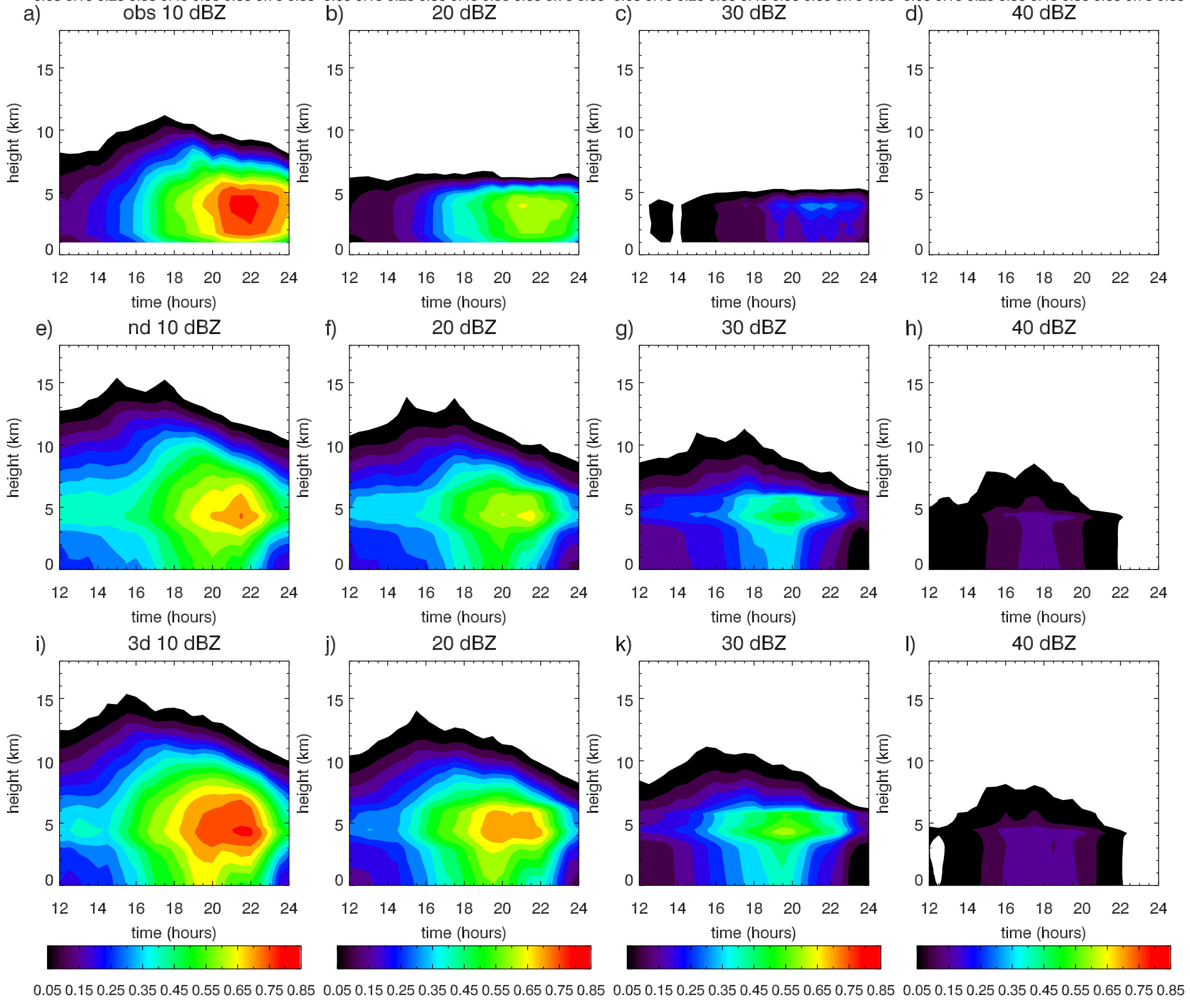

a) obs 10 dBZ
b) 20 dBZ
c) 30 dBZ
d) 40 dBZ
e) nd 10 dBZ
f) 20 dBZ
g) 30 dBZ
h) 40 dBZ
i) 3d 10 dBZ
j) 20 dBZ
k) 30 dBZ
l) 40 dBZ
height (km)
time (hours)
0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85

Figure 6. The observed (topleft 4 panelslsls), and simulated by the control model (middleright 4 panels) and simulated with a change to the turbulent mixing (lower panel) fraction of radar detected area covered by reflectivities greater than a,e,ie) 10, b,f,jd) 20, ce,g,k) 30 and df,h,l) 40 dBZ for 12 – 24 UTC on 18/02/2014.

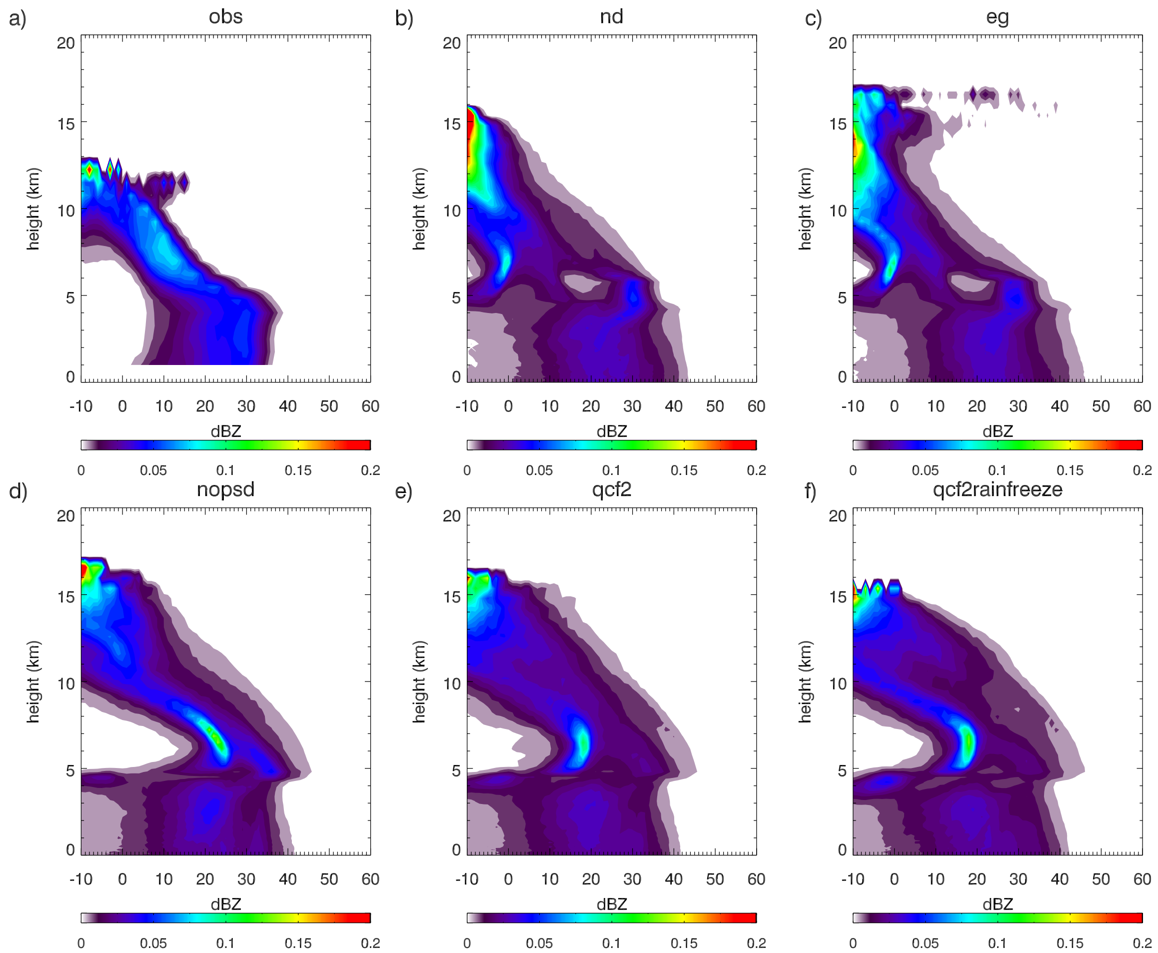



Figure 7. Contoured frequency with altitude diagrams of radar reflectivity for the region within 150 km of the radar for the times 23 – 24 UTC. a) Observations, b) control simulation, c) ENDGame dynamical core simulation, c) no use of the generic ice PSD parameterisation, d) additional ice prognostic and e) inclusion of heteorogeneous ice freezing parameterisation. See text for details on different simulations.

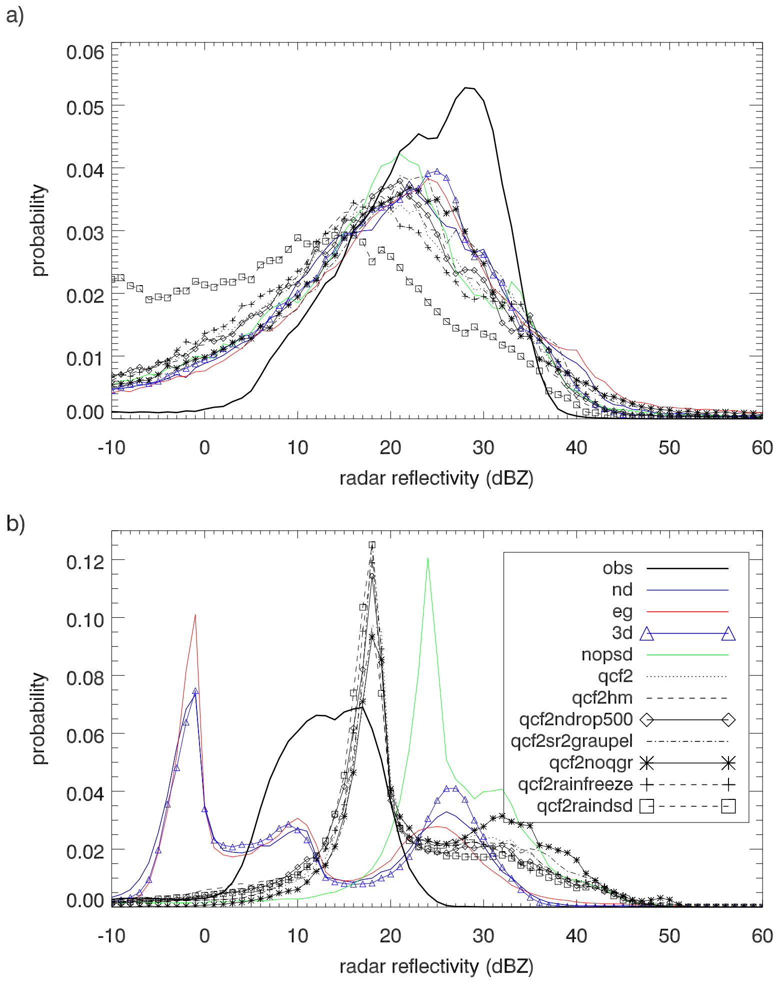



Figure 8. Radar reflectivity probability density functions for two heights, a) 2.5 and b) 6 km.

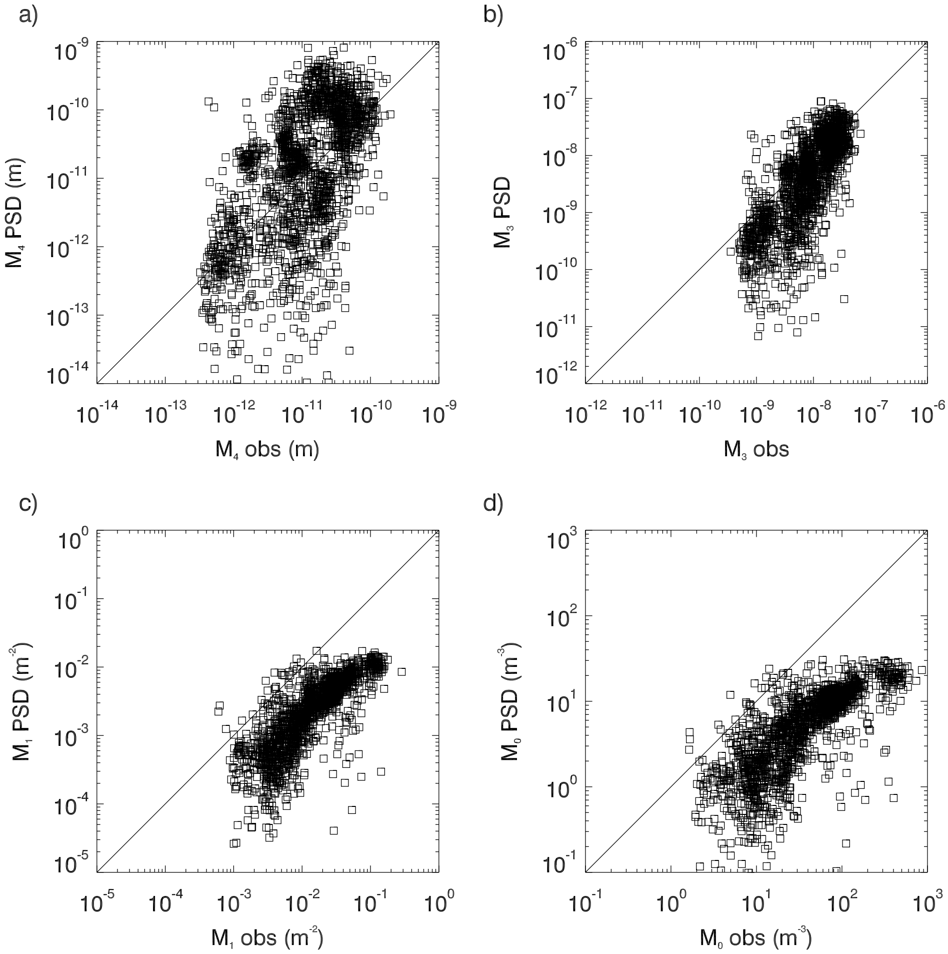
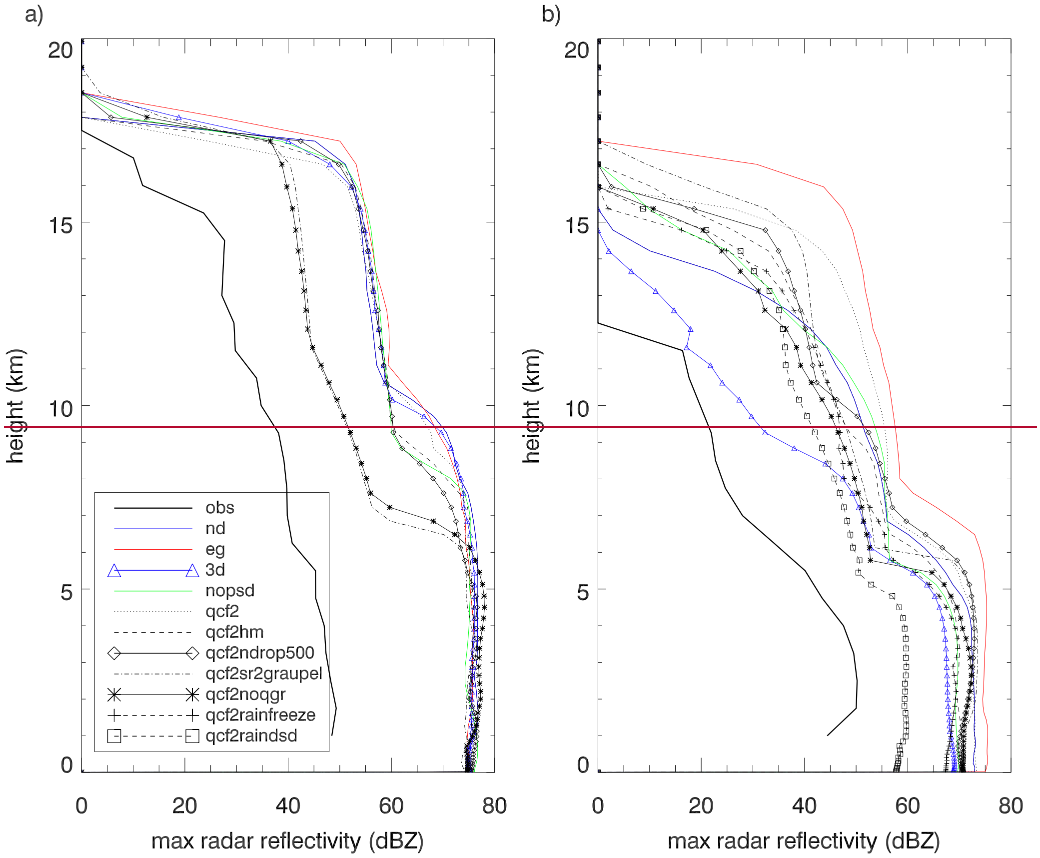



Figure 9. Moments ($4^{th}$, $3^{rd}$; $1^{st}$ and $0^{th}$) of the observed particle size distribution by the aircraft (for particles with diameters > 100 µm) and predicted using the PSD parameterisation with the observed ice water content (> $10^{-3}$ g $m^{-3}$), temperature and mass-diameter relationship.

1



a)
b)
height (km)
max radar reflectivity (dBZ)
obs
nd
eg
3d
nopsd
qcf2
qcf2hm
qcf2ndrop500
qcf2sr2graupel
qcf2noqgr
qcf2rainfreeze
qcf2raindsd


1

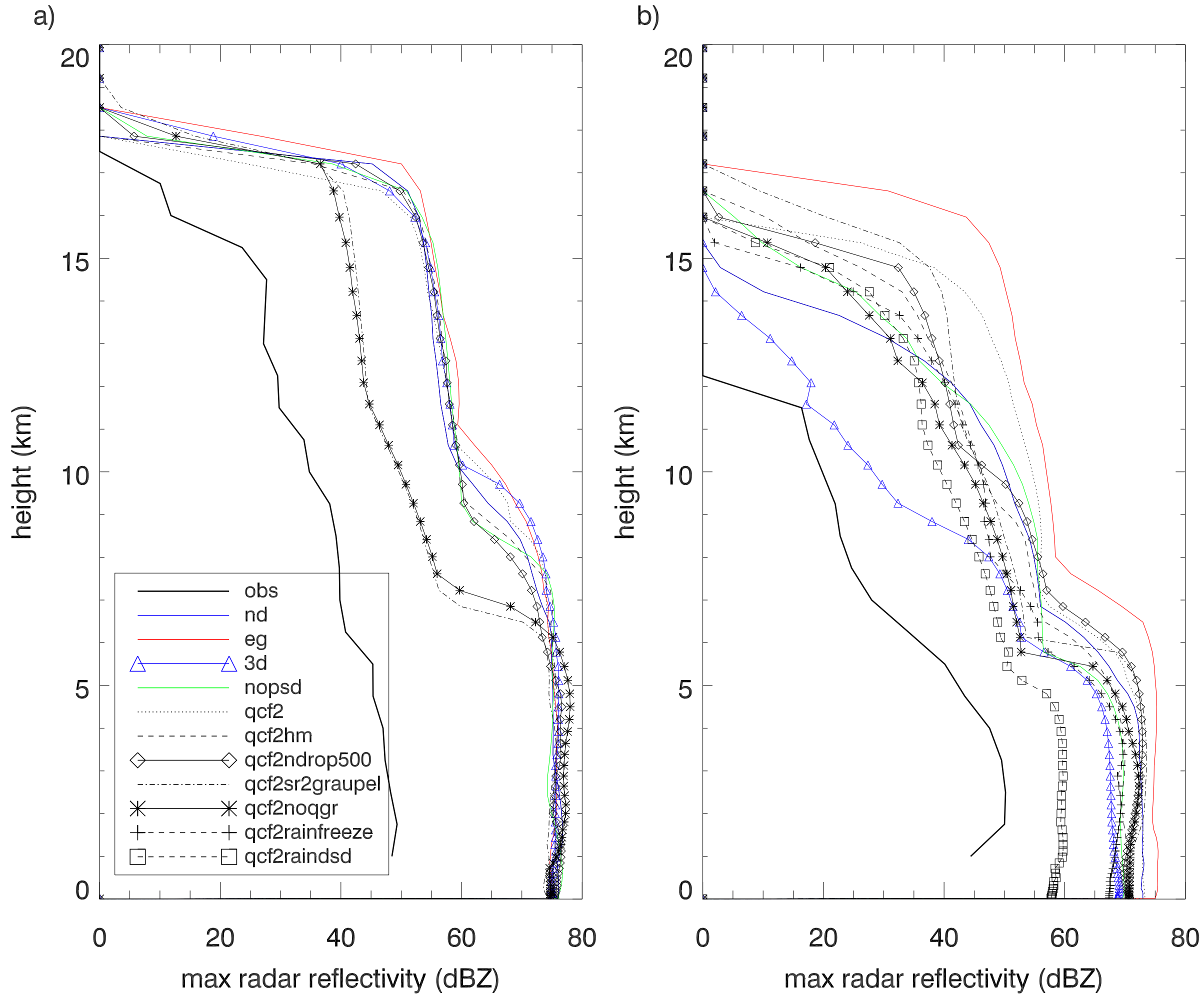
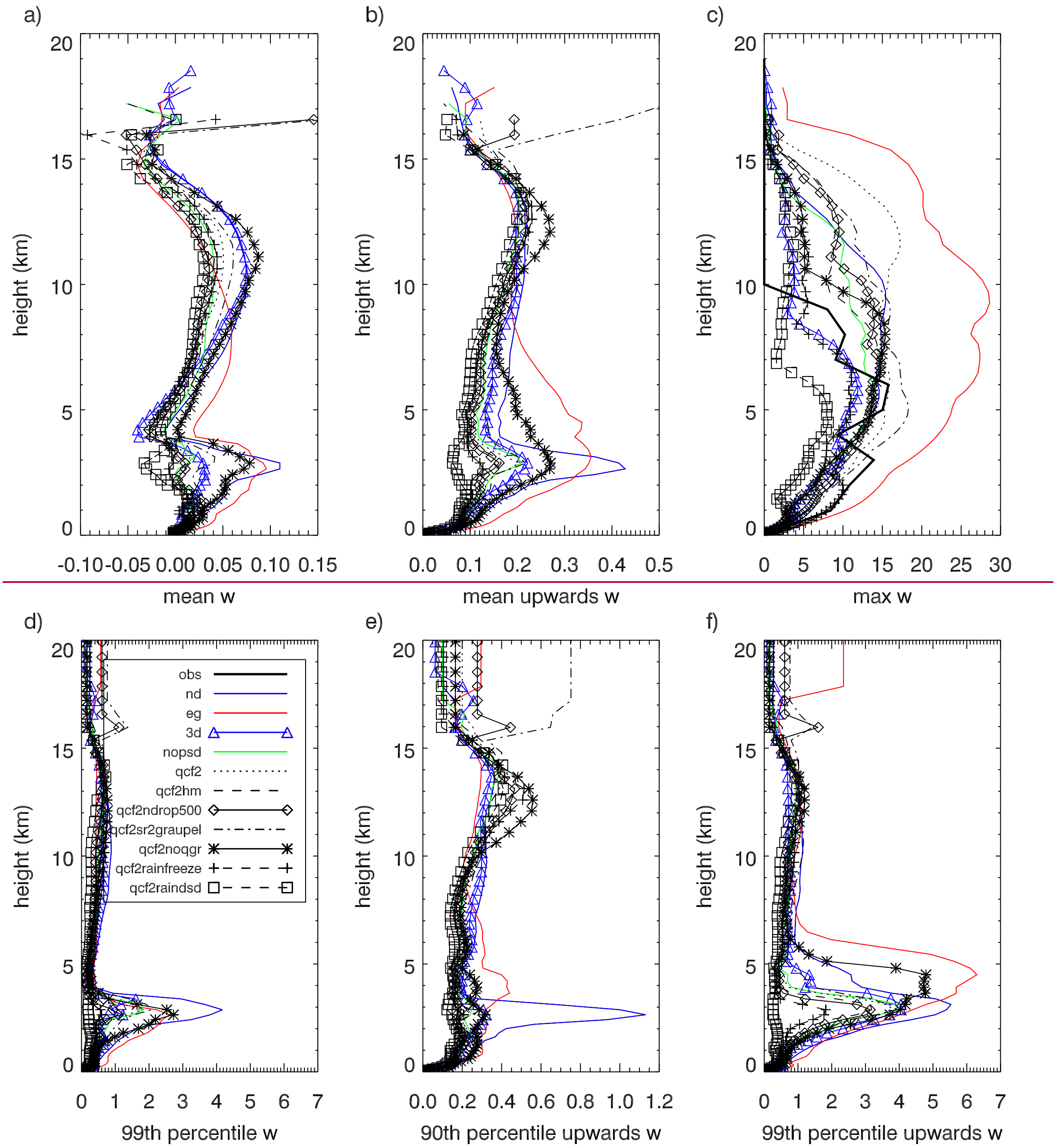



Figure 10. Profiles of maximum radar reflectivity for the times a) 17 – 18 UTC and b) 23 – 24 UTC.


a)
b)
c)
d)
e)
f)
height (km)
mean w
mean upwards w
max w
99th percentile w
90th percentile upwards w
99th percentile upwards w
obs
nd
eg
3d
nopsd
qcf2
qcf2hm
qcf2ndrop500
qcf2sr2graupel
qcf2noqgr
qcf2rainfreeze
qcf2raindsd

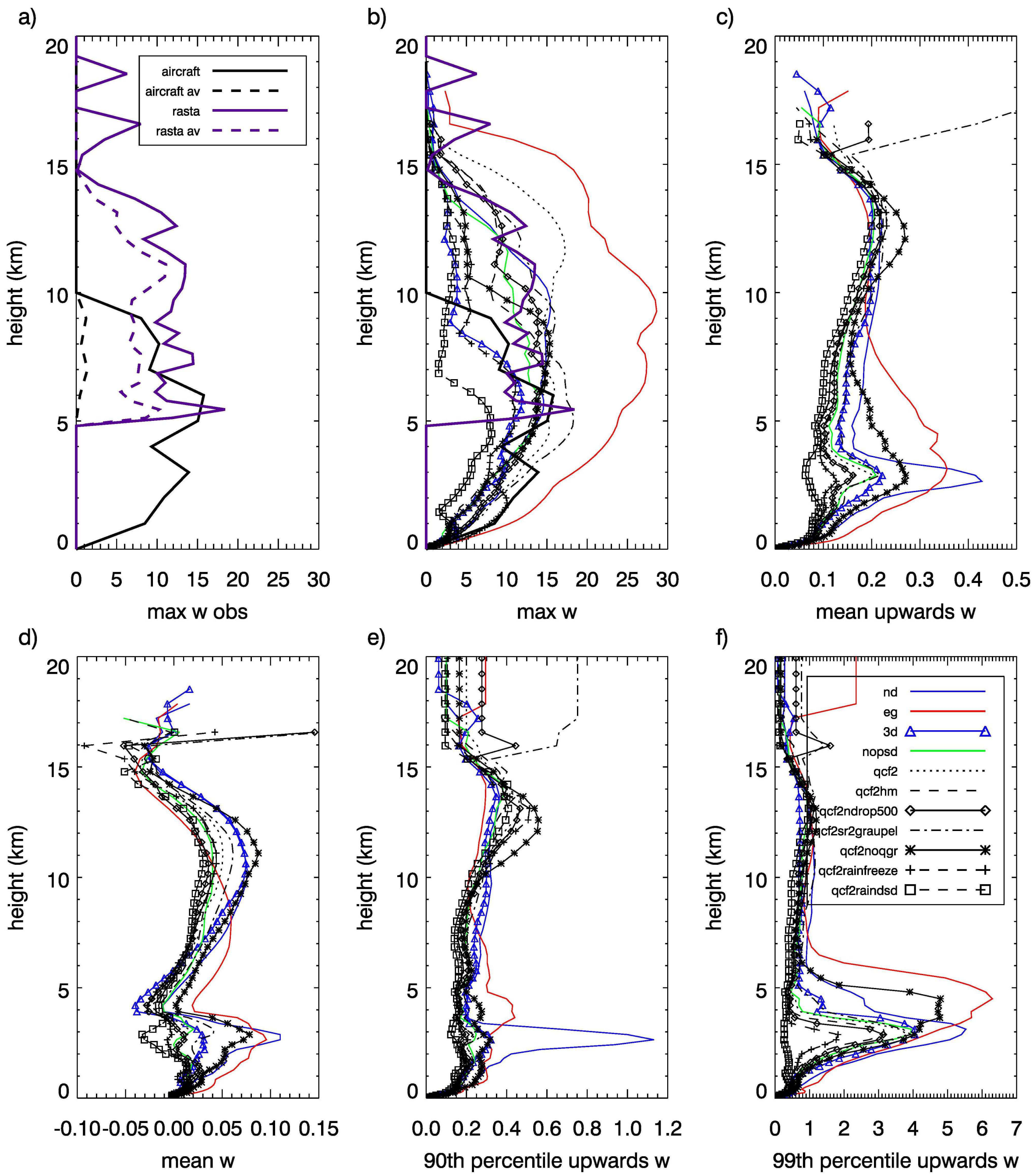
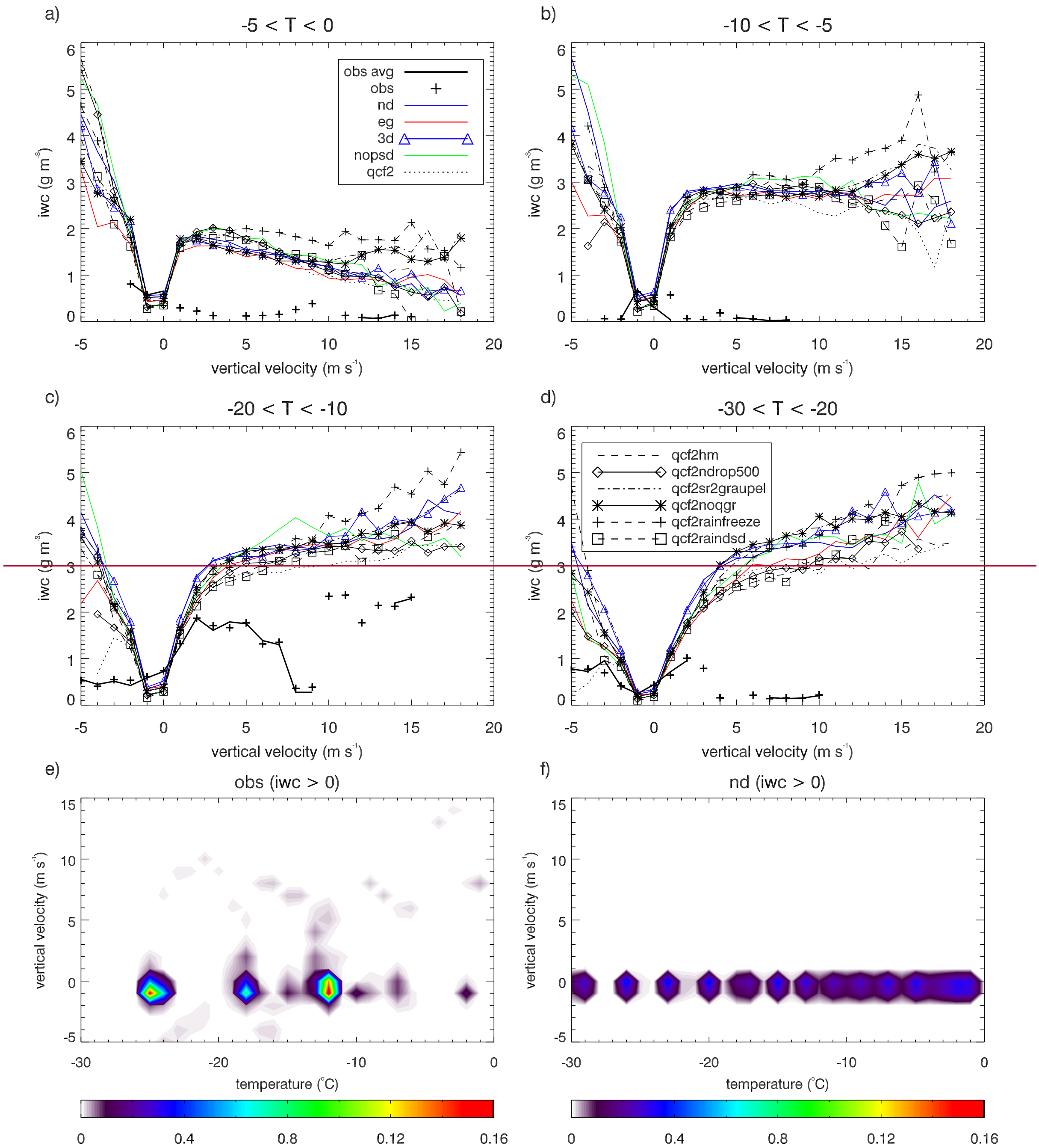



Figure 11. a) Maximum vertical velocity observed by the aircraft and derived from RASTA (Radar SysTem Airborne) for the times 23 – 24 UTC. Solid lines are using the highest resolution observations, dashed lines are using the observations averaged to the 1 km resolution. Modelled i~~I~~n-cloud vertical velocity statistics (m s$^{-1}$) over the radar domain for the times 23 – 24 UTC:~~.~~ b~~a~~) maximum~~Mean~~, c~~b~~) updraft mean, d~~c~~) mean~~maximum~~, e~~d~~) updraft 90~~9~~th percentile, and f~~e~~) updraft 99~~0~~th percentile~~ and f) updraft 99th percentile~~.

1


a)
-5 < T < 0
b)
-10 < T < -5
c)
-20 < T < -10
d)
-30 < T < -20
obs avg
obs
nd
eg
3d
nopsd
qcf2
qcf2hm
qcf2ndrop500
qcf2sr2graupel
qcf2noqgr
qcf2rainfreeze
qcf2raindsd
iwc (g m-3)
vertical velocity (m s-1)
e)
obs (iwc > 0)
f)
nd (iwc > 0)
temperature (°C)
0
0.4
0.8
0.12
0.16

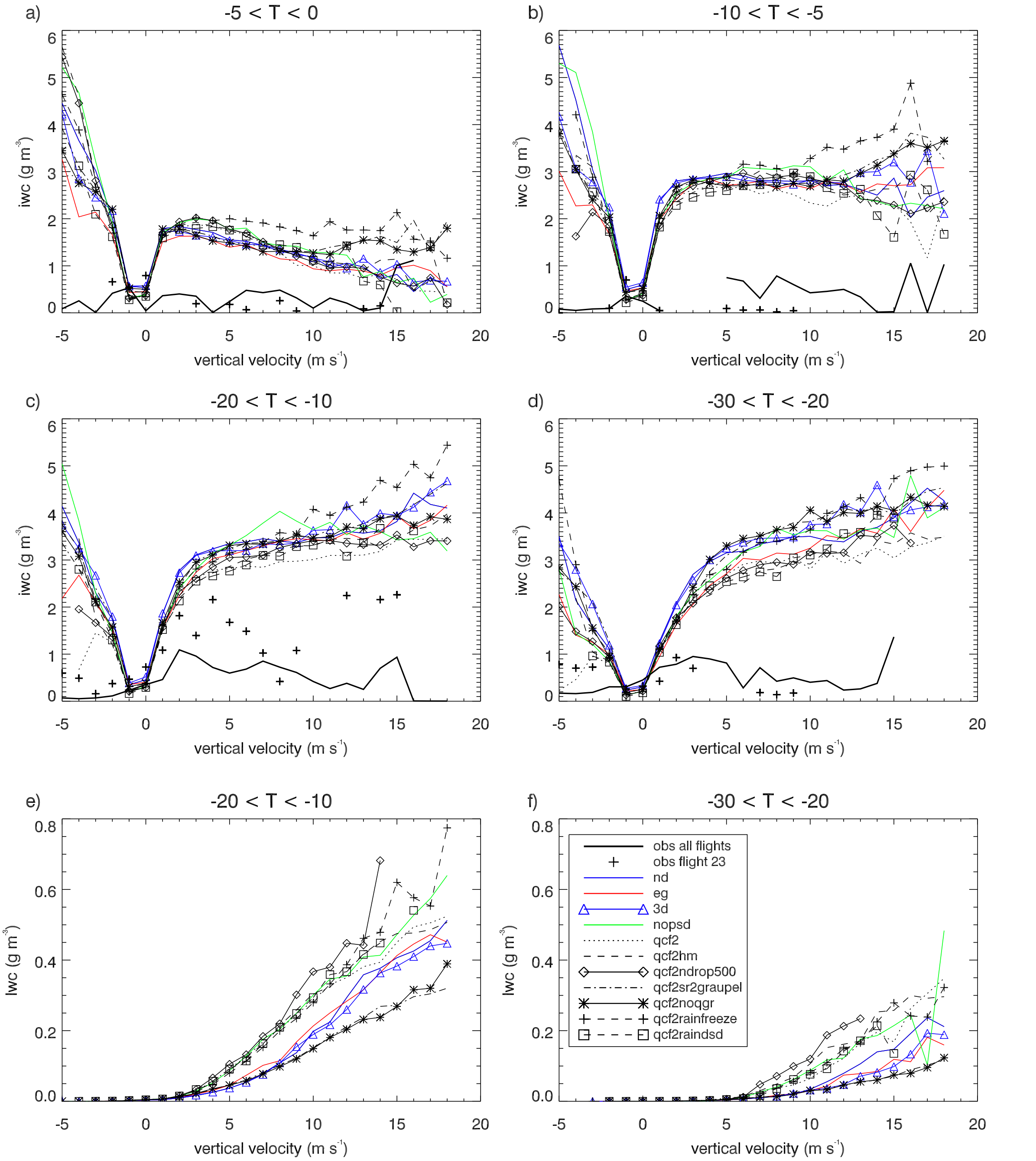



Figure 12. Ice water content (g m$^{-3}$) as a function of vertical velocity (m s$^{-1}$) for four temperature regimes: a) -5 – 0; b) -10 – -5; c) -20 – -10, and; d) -30 – -20 $^{o}$C. e) and f) show liquid water content (g m$^{-3}$) as a function of vertical velocity for the two coldest regimes: e) -20 – -10, and; f) -30 – -20 $^{o}$C. the joint probability density functions of vertical velocity and temperature for the observations and the control 1 km simulation for regions with IWC > 0.

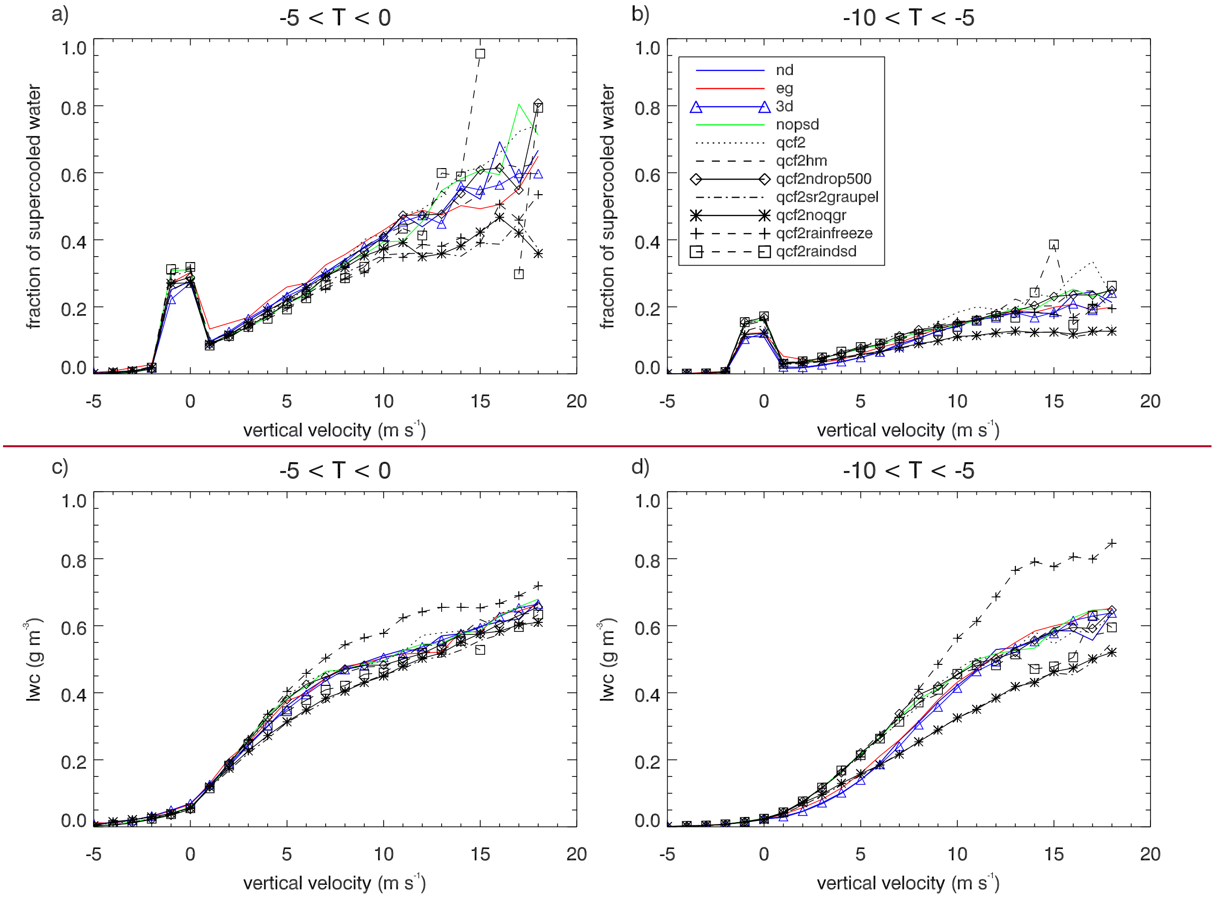



Figure 13. a), b) Fraction of condensate that is supercooled cloud liquid water, c), d) liquid water content (g $m^{-3}$) for the two warmer temperature regimes a), c) -5 – 0 and b), d) -10 – -5 ºC.

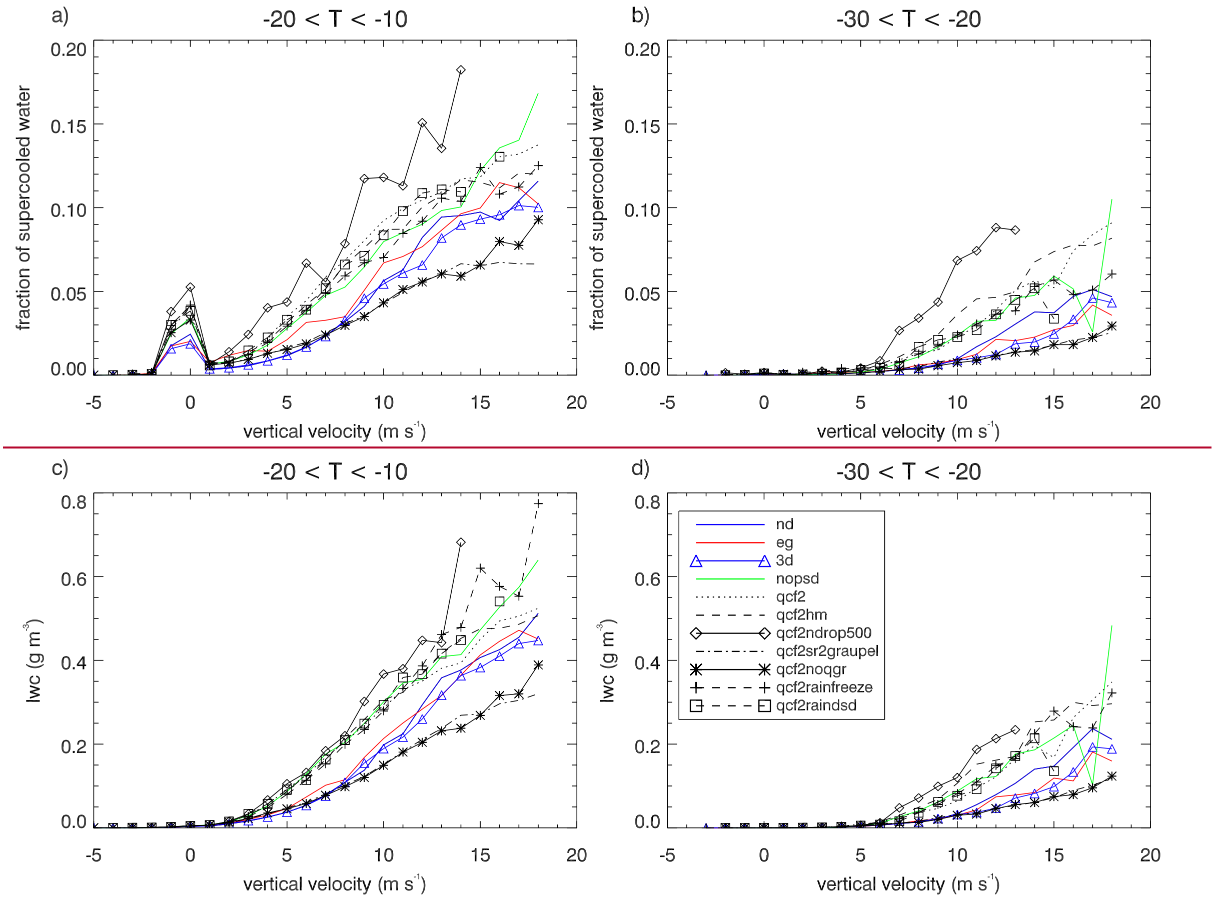
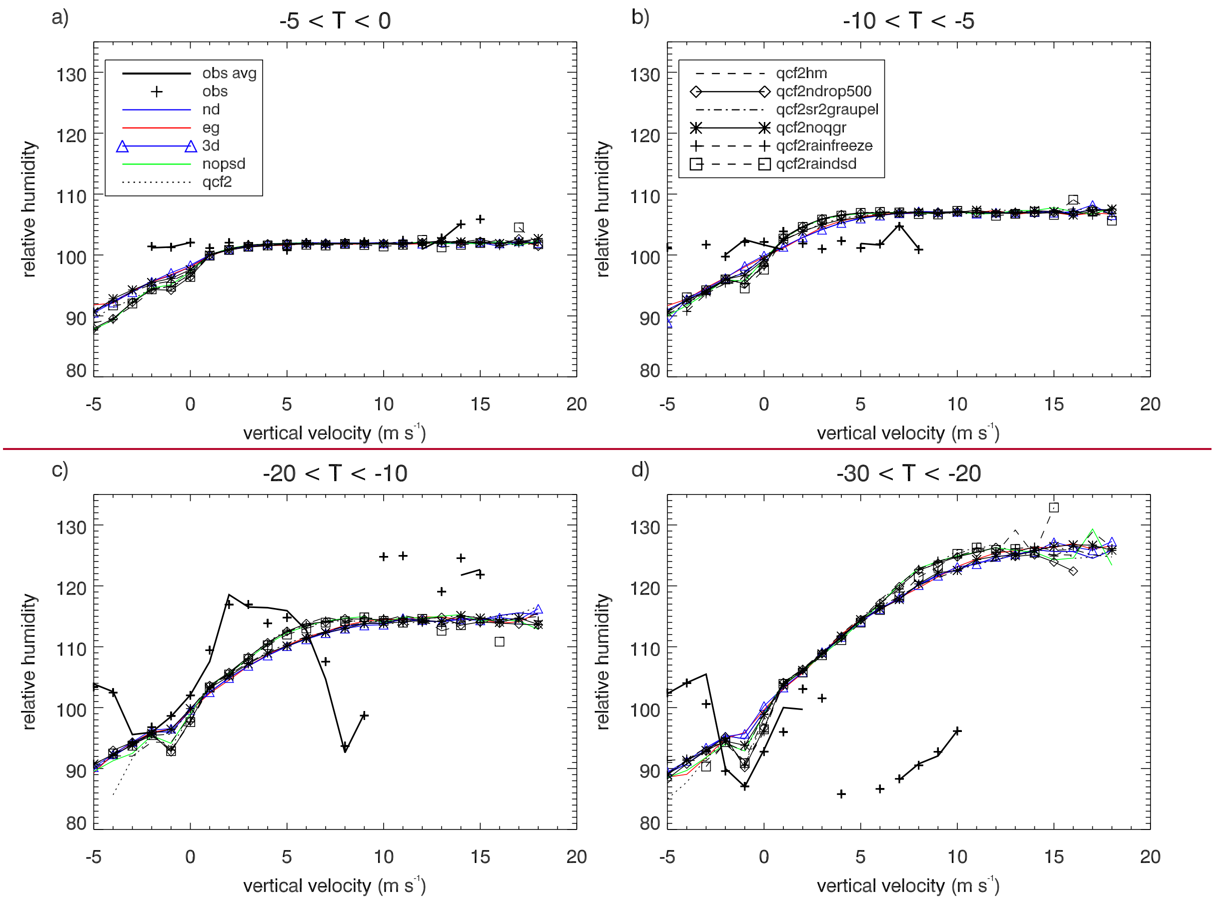



Figure 14. As in Figure 9 except for the two colder temperature regimes: a), c) -20 – -10 and b), d) -30 – -20 ºC.




~~Figure 15. As in Figure 8 accept for relative humidity as a function of vertical velocity for the four temperature regimes.~~

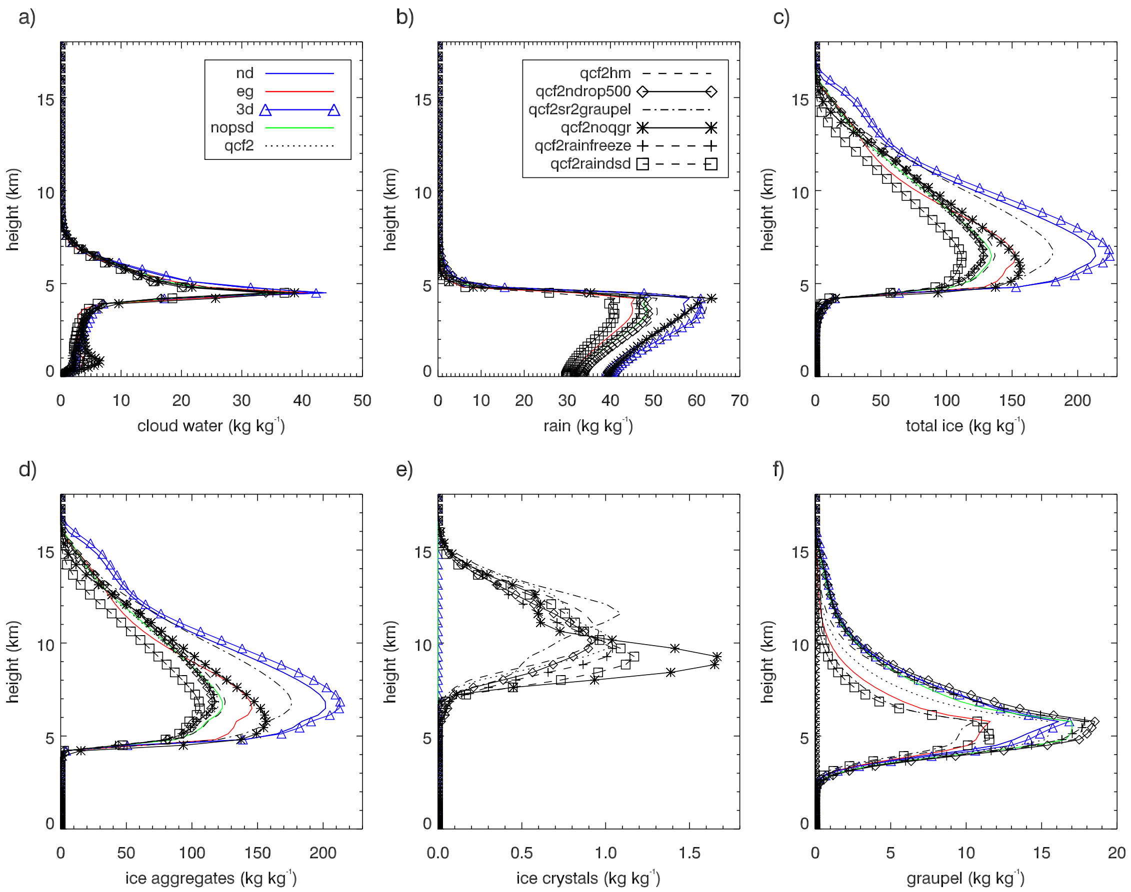
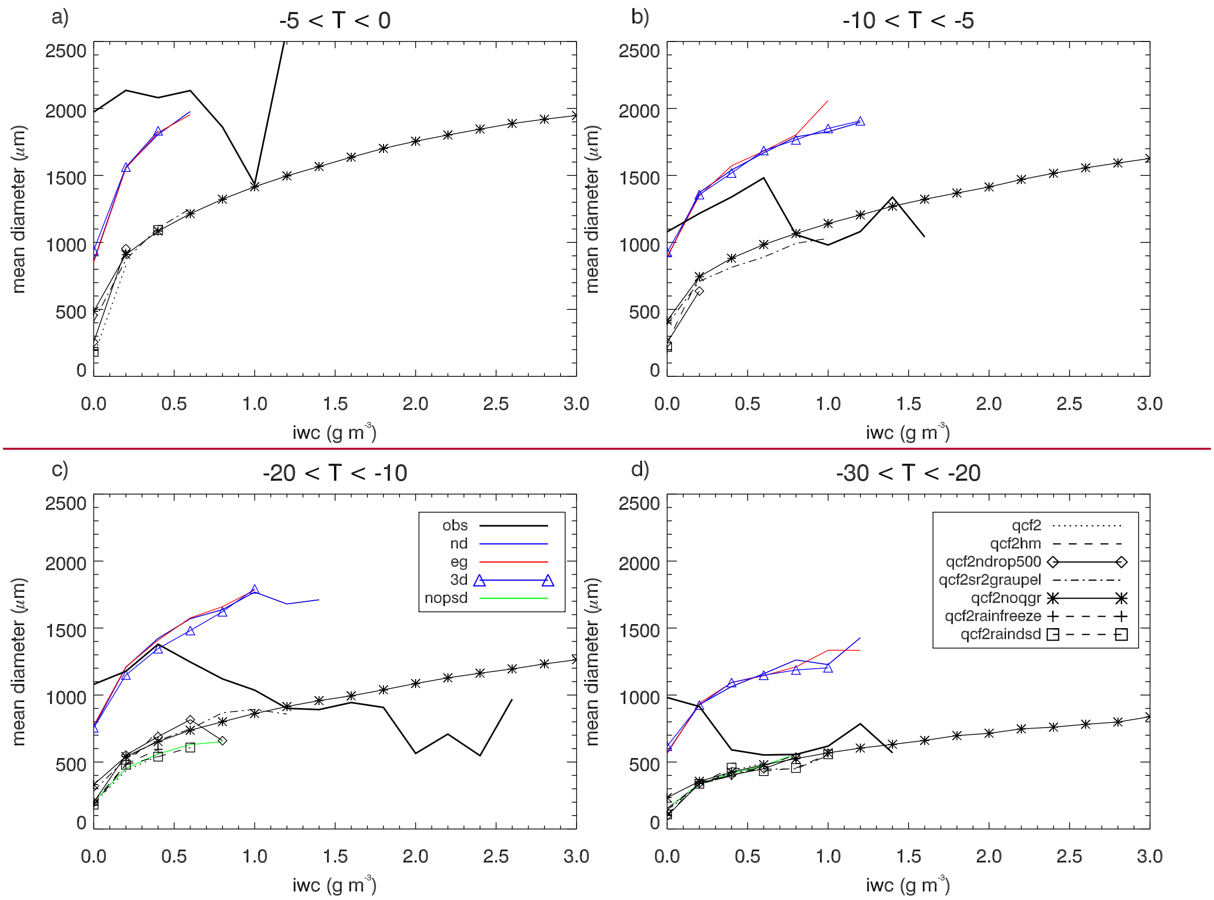



Figure 136. For the aircraft analysis region (150 km radius from the mean aircraft track), the total accumulated water contents (kg kg$^{-1}$) over the domain from 23 – 24 UTC. a) Cloud liquid water, b) rain water, c) total ice, d) ice aggregates/snow, e) ice crystals and f) graupel.


a)
-5 < T < 0
b)
-10 < T < -5
c)
-20 < T < -10
d)
-30 < T < -20
mean diameter (µm)
iwc (g m-3)
obs
nd
eg
3d
nopsd
qcf2
qcf2hm
qcf2ndrop500
qcf2sr2graupel
qcf2noqgr
qcf2rainfreeze
qcf2raindsd

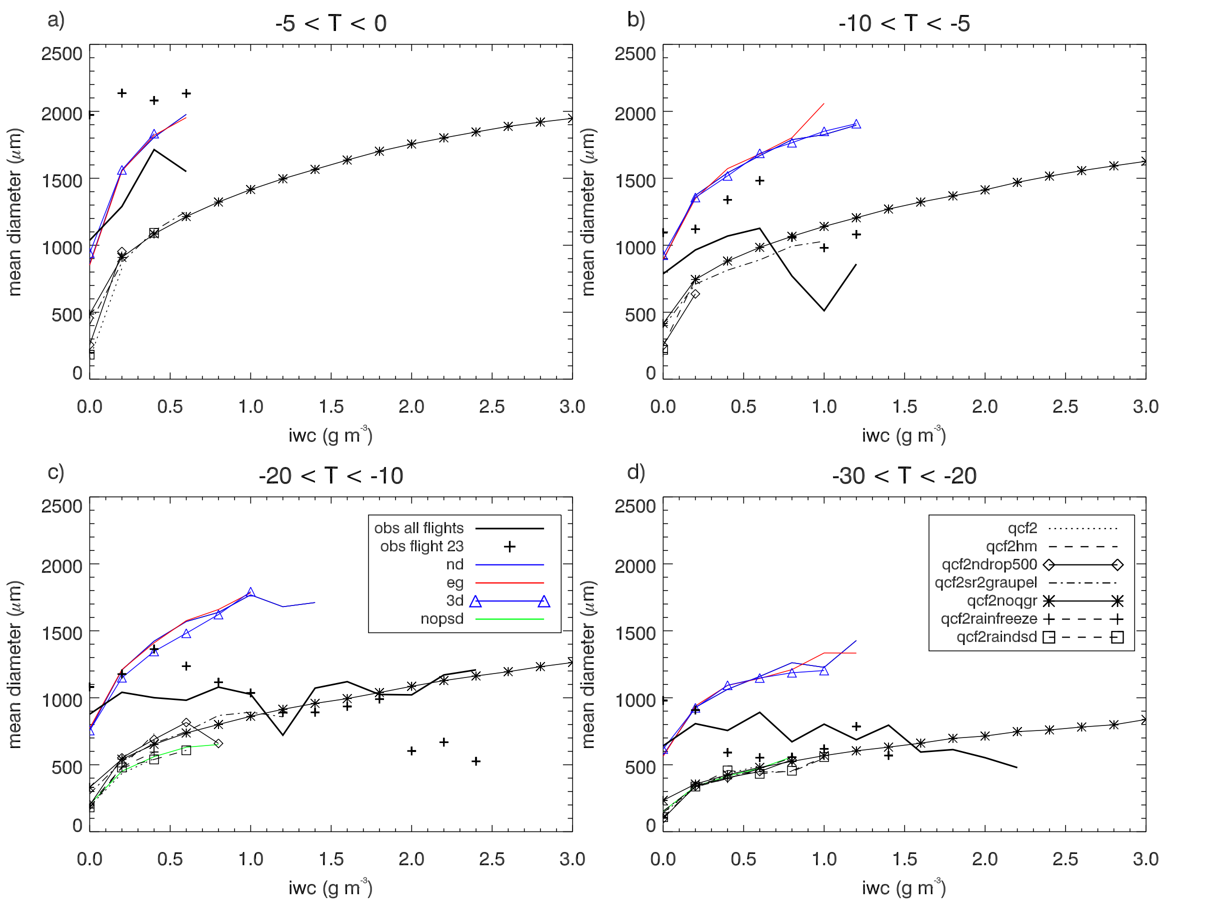



Figure 147. Mean mass-weighted ice particle size (µm) as a function of ice water content (g m$^{-3}$) for four temperature regimes: a) -5 – 0, b) -10 – -5, c) -20 – -10, and; d) -30 – -20 $^{o}$C.