Response to Referees

Overall Response: We would like to thank the reviewers for their detailed comments and suggestions, which are helpful in improving the manuscript. In this response, we have addressed all the reviewers' suggestions and comments, and have heeded nearly all of the suggestions from the reviewers and incorporate the changes in the resubmitted manuscript. More details are provided in this response following the reviewers' specific comments.

Response to Referee #1, Grant Allen:

Summary:

This study is, in essence, a validation of prognostic aerosol capability for WRF-CHEM with an interesting and encouraging further validation of interactive aerosol coupled with a dual-moment cloud microphysics (Morrison) scheme. This exercise makes good (but as yet incomplete) use of an enormous wealth of data collected during the VOCALS-Rex field campaign conducted over the South East Pacific in late 2008. Marine boundary layer thermodynamic quantities, aerosol number, aerosol mass and composition, and cloud bulk properties are all analysed and intercompared statistically over meteorologically and compositionally distinct spatial domains within the wider South East Pacific. The study represents an enormous data analysis exercise, for which the authors should be commended. However, in one or two instances, I think the authors have jumped a little too quickly to confidently stated conclusions that cannot be fully justified (especially in terms of boundary layer structure) – these will be detailed in the general comments below. Also, and perhaps most importantly, it seems that the failure to run the model at sufficiently high spatial resolution so as to capture convective dynamic processes on scales relevant to MBL SCu, or to otherwise include available sub-grid convective parameterisations, may potentially lead to large errors that could have been avoided. Since we are looking here at aerosol-cloud interactions, convective dynamics are critical but not represented as well as they need to be for this analysis to provide useful additional insight for modellers as the study stands at present. In addition, I feel that an opportunity has been missed to make good use of studies already published in the VOCALS special issue (i.e. Allen et al., 2011) and also the BAe-146 dataset. I have chosen to lift anonymity here so that I can help the authors if they would like me to (without asking for co-author recognition).

The work presented here is very important to the atmospheric modelling community given the growing uptake and success of the WRF model and it is timely given the recent improvements to WRF-CHEM. The study also provides a useful distillation of measurement datasets from the VOCALS-Rex campaign, which would be useful to a wider audience (e.g. those interested in composition and dynamics in general). As such, the study is applicable for publication in ACP and will ultimately represent an important and extensive contribution to the VOCALS special issue.

The paper is well presented and well written, the figures and tables are of excellent quality and the study has been expertly executed in the main. It is this reviewer's recommendation that this study be published in ACP subject to some important revisions, which will now be detailed below. These revisions are hopefully minor relative to the sheer amount of excellent work that must have already gone into this study. This study is so important and already so well executed, that I think it is worth the extra effort to make sure it is the very best that it can be and I hope the authors recognise the effort taken in this review to that constructive end. As someone who is not familiar with the intricacies of WRF-CHEM, I have asked for further advice and comments from colleagues working with the WRF model – these are also included below.

General Response: We thank the referee, Dr. Grant Allen, for the constructive comments and suggestions. By addressing them during the revision we believe the manuscript has been greatly improved. Below are our detailed responses to the specific comments.

General Comments:

1/ Co-authors: Have the authors carefully checked the status of data protocols on all the datasets used in their study? Good use has been made of a large proportion of the VOCALS-Rex dataset but co-authorship may need to be offered to instrument PIs etc whose data are used here unless data protocols now specifically preclude it. I see several names (several spelled wrong) in the acknowledgements but care needs to be taken that this is sufficient for those (and other) scientists. For this reviewer's part, I am happy that the GOES-10 dataset I spent several weeks creating is being used and I do not require co-authorship, though it would have been proper to have been given the choice. Issues with properly recognising work contributions are very important. The long author list of Allen et al.,2011 is testament to such requirement and I can confirm that getting that author list right was necessarily difficult. This is a growing issue amongst the modelling community, who use measured data as if they were created from thin air (no pun intended). Quality assured measurement datasets are the result of a lot of hard and careful work by many people and future good quality datasets rely on funding that is justified by publications of work which use it and recognise it. The converse is also true – model output should not be used by those primarily working with instruments without recognition of the people who generated such products. Most importantly, co-authorship is important to ensure proper use of the most accurate datasets. I strongly recommend that the authors and the editor ensure that this issue has been paid due diligence and add co-authors if necessary.

Response 1.1.0: Thanks for bringing this up. We certainly did not mean to under-appreciate the contributions of scientists, technicians, and crews in the field who had made the observations available to the community. As suggested, we have contacted individual PIs and offered them co-authorship. Those who accepted have been added to the author list, and some of them declined. In the acknowledgements, specific statements provided by some PIs have been added. Proper citations to some observational studies have also been added in the observational data section (Sect. 2.2).

2/ Use of Allen et al., 2011: I feel a little self-important having to promote my own work here but I do feel that there is a wealth of observational statistics and analysis already available in the VOCALS special issue that would have been directly applicable to this study that it would have made sense to make use of or at least discuss. The authors are entirely entitled to repeat my analysis, tailoring it to their needs but I think the authors would have been doing a great service to VOCALS if they had better used (or even more thoroughly intercompared) the existing distillation of data in Allen et al., 2011, which was written for the very purpose of informing studies such as these.

Response 1.2.0: We have taken the recommendation to do a closer cross-check of the results from Allen et al. (2011) and those from this manuscript in addition to the comparisons we had with Bretherton et al. (2010), Hawkins et al. (2010), etc. Specific details are provided below to address the detailed comments. In the following we describe in more detail how we tailored the use of the measurement data driven by the goal of our own project research.

Specifically, Figure 3 in this paper is an incomplete copy of panel b in Figure 11 of Allen et al., 2010. This study uses cloud droplet data (Fig.3) from the C-130 only, where Fig. 11 of Allen et al, weights all data from the G1, BAe-146 and C-130 aircraft, presenting a more complete statistical picture. Figure 11 of Allen et al., could have been happily reused in this paper and I recommend that it still is.

Response 1.2.1: We selected the C-130 and RB as our main measurement platforms for this study due to two main reasons: 1) both the C-130 and RB provided measurements extended to the remote ocean region (~88°W) which are necessary for our purpose of contrasting different aerosol and cloud characteristics over polluted and clean environments; and 2) the RB provided measurements of energy fluxes, DMS ocean-to-air transfer velocity, and other valuables variables at the surface level that are important to our model assessment, and those measurements are complementary to measurements on the flight platforms. The selection of the RB as the main platform leads to the division of the model domain into two regions (remote and costal regions), since the RB sampled more intensively around ~85°W and ~75°W. BAe-146, however, did not fly beyond west of 81°W (with an average sampling longitude of 79°W over the remote region). Their measurements only covered a small portion of what we have defined as the remote region in our domain, and thus are less suitable for studying the coastal and remote contrasts.

We have added the following justification to Sect. 2.2: "The C-130 and RB were selected as the main measurement platforms for the model evaluation in this study due to two main reasons: 1) both the C-130 and RB provided measurements farther into the remote ocean region (~88°W) of the model domain; this extended longitudinal data coverage is necessary for the purpose of contrasting different aerosol and cloud characteristics over polluted and clean environments; and 2) the RB provided energy fluxes, DMS ocean-to-air transfer velocity, and other near-surface measurements that are important to the model assessment, and those measurements are complementary to measurements obtained on the flight platforms. The selection of the RB as one of the two main platforms leads to the division of the model domain into two regions (remote and costal regions) since the RB sampled more intensively around ~85°W and ~75°W. Within the domain, there are about twice as many samples over the coastal region compared to the remote region from both the C-130 and RB platforms during the VOCALS-REx."

Using weighted averages of data from different measuring instruments may provide 'a more complete statistical picture of observations', but could introduce uncertainties associated with differences in retrieval algorithms, instrument detection limits, measurement size ranges, bin size resolutions, etc. The focus of this paper is the evaluation of model simulations. We prefer to maintain our focus by using the C-130 data consistently for both coastal and remote regions in Fig. 3 to show the variability of aerosol number and droplet number (due to synoptic variability, etc) for both measurements and model predictions.

Temporal and spatial variations of aerosol and cloud properties are affected by synoptic variations and diurnal cycles. In order to account for these variations, model outputs were interpolated to measurement times and locations, and only coincident model outputs and measurement data were used for the comparisons. In addition, we chose to segregate the model results into those below clouds and those within clouds, while Fig. 11 of Allen et al. (2011) are divided into the boundary layer and free troposphere. Therefore, we cannot directly refer to Fig. 11 of Allen et al. (2011). In the manuscript, we did compare the domain-average N_d with those from Bretherton et al. (2010). However, we have added the following in the manuscript: "The observed longitudinal variation in N_d is in general agreement with the variation shown in Fig. 11 of Allen et al. (2011) in which measurements from CDP measurements on both the BAe-146 and the C-130 were included.".

Furthermore, Figure 4 of this work quotes AMS data divided into coastal and remote zones. I realise that the distinction between coastal and remote are chosen differently for this study so
this figure is fully justified but Fig. 8 of Allen et al., also includes AMS data from the BAe-146 - data which are noticeable by their absence here (see below).

Response 1.2.2: Response 1.2.1 provided details regarding why we chose measurements on the C-130 and RB as main data sources. Since the RB had sample intensively near 75W and 85W, the division of the remote and coastal regions took into account of the ship's sampling strategy. There was no intention of using a different zoning strategy from Allen et al. (2011). The inclusion of G-1 data was particularly important because the PILS on the G-1 provided sea-salt measurements that were useful for further cross-checking of the large overprediction in sea salt mass concentrations compared to the measurements from the CIs on the RB. Again, we focused on the evaluation of model simulations, and used coincident measurements and model outputs, which considered sampling time and location differences in the model-measurement comparisons. As suggested, we have added BAe-146 data into the Fig. 4. The following discussion has also been added to Sect. 3.2.2:

"Over the coastal region, the high mean sulfate concentration (1.81 $\mu g\ m^{-3}$, Table 4) measured by the AMS on the BAe-146 is dominated by the high values (mean of 2.82 $\mu g\ m^{-3}$) in a pollution plume study flight on 10 Nov 2008. Excluding this flight, the mean sulfate concentration (1.13 $\mu g\ m^{-3}$) over the coastal region observed on the BAe-146 is in close agreement with those measured on the RB and C-130 (Table 4), which is underpredicted by 30% in AERO. Observed sulfate concentrations from different platforms are 0.30–0.40 $\mu g\ m^{-3}$ (Table 4) over the remote region. These observed values over both coastal and remote regions are in a general agreement with those in Fig. 8 of Allen et al. (2011). Over the coastal region, the higher mean sulfate concentration from the RB in Table 4 compared to Fig. 8 of Allen et al. (2011) is mostly because we only used the RB observations within the VOCALS-REx period (15 Oct - 15 Nov), which is a subset of the RB observations."

In addition, Fig. 6 of Allen et al., could have been quoted to demonstrate the measured PDFs of ozone and SO2 (reported there for all aircraft) that are highly relevant to this study, especially when discussing SO2, which appears to be one of the biggest and most important sources of uncertainty in the WRF-CHEM model and this analysis.

Response 1.2.3: Gas-phase chemistry was not a focus of this study, and the discussion of

SO_2 in the paper was to check whether the oxidation of DMS is underestimated, which could lead to underestimation of sulfate. Since the discussion involves multi-day average values, the SO_2 and O_2 PDFs in Fig. 6 of Allen et al. (2011) would not strengthen that discussion.

3/ Use of BAe-146 data: VOCALS-Rex was an international field campaign. BAe-146 aircraft data have not been used in this study, leading to an incomplete use of the VOCALS dataset. Given the authors' extensive work in this study already, why was this platform not considered? I would suggest that it should be used in a new version of this paper to complete the excellent statistical picture that this work has built. I can offer assistance (without co-authorship) in providing that data. There are important sampling differences between the aircraft noted in Allen et al., 2010 that can only be filled in (as much as possible) by including as much data as possible. Those sampling differences (and the filling in offered by the BAe-146) are absolutely crucial in providing the best distillation of the measured data for comparison to WRF-CHEM output.

Response 1.3.0: As we pointed out earlier, this work focuses on assessing model simulations using VOCALS-REx data with emphasis on comparing results between remote and coastal marines. Please refer to Response 1.2.1 for our justification of choosing the C-130 and RB as main research platforms for this study. The frequent comparisons of our results based on the C-130 and RB to those in the literature, such as Bretherton et al. (2010) and Hawkins et al. (2010), provides confidence in the conclusions we drew from comparisons to the C-130 and RB observations. As suggested by the reviewer we have provided additional comparisons of the model results with the BAe-146 data throughout the manuscript where appropriate.

4/ Section 3.1: Boundary layer structure: In terms of the quality of the existing conclusions and analysis (issues above aside), this section is where I think there may have been some misinterpretation that leads to one or two later potentially incorrect conclusions. This section is an interpretation of radiosonde measurements from the Ron Brown (RB). A selection of sonde data recorded separately in the remote zone and the coastal zone are presented in Figure 1. The figure is used to show a cold, low humidity bias in the coastal zone relative to the remote zone. Herein lies a potential problem – the sonde data recorded by the RB in the remote zone were heavily weighted to its time on station at 85 W and likewise at 75 W in the coastal case. The variability sampled in the sonde data for each location represents only the variability across the times on station at each location, which were separated by several weeks, both of which are subsets of the period-averaged model data. As noted in Toniazzo et al., 2011, the synoptic conditions between these two time periods when the RB was on station were quite different, with weaker and less variable subsidence after November 1st. This alone could be argued to result in the biases seen in sonde data between the two locations. Furthermore, the explanation that the coast is colder than the remote area due to SST differences is not backed up – were measured SSTs at 75 W colder than those at 85 W I say this as 75 W could be too far west for the spatial extent of the Humboldt current. Perhaps a quick plot of SSTs from the RB when on station at these two locations could answer this problem? Furthermore, this argument is confused by a statement later in the paper that SST-air temperature differences make little difference to MBL depth. The issue of MBL depth is about much more than SST differences, with the subsidence rate and coastal dynamics all playing an important role which varies in magnitude with distance offshore. As a solution to solve this question of cold, humidity biases, I recommend using dropsonde data from the BAe-146 aircraft, which should be added to Figure 1 (or perhaps used instead). These were dropped along 20 S on many flights (and measured at all locations on the same day), which will show if the MBL was systematically colder near the coast than in the remote zone, avoiding the disconnect in temporal sampling in the RB dataset. I raise this important issue because these sonde data are quoted repeatedly in the paper to be the cause of systematic differences in boundary layer structure, as well as potential causes of other chemistry and model biases. If you look closely at Figure 1, the model-model bias in MBL temperature and humidity between the two regions is much smaller (< 1K) than that seen in the sonde data. Also, the authors have rightly noted that Bretherton et al., 2010 did not observe a similar meridional gradient in MBL humidity – together these facts (the sampling disconnect in sonde data used, the small bias in model thermodynamics and the previous findings of Bretherton) make cause enough to question the sonde data used and the current conclusions. I recommend using the BAe-146 sonde dataset and the findings of Bretherton et al., 2010, and the authors' existing analysis of radar intercomparisons (Fig. 8, see item 5 below) to reconsider the analysis of MBL thermodynamics versus the model and the conclusions drawn therefrom. I would risk the hypothesis that the model does an even better job thermodynamically than is currently presented. If left unchecked, I could quite easily see an error value of 150 m in MBL depth error for WRF-CHEM being assumed elsewhere without due justification.

Response 1.4.0: We thank the referee for bringing up a very good point regarding the sampling time difference in the RB measurements. The conclusion about the drier coastal MBL during the VOCALS-REx is consistent with a recently published study by Jones et al. (2011). They used C-130 measurements to study decoupling over the remote ocean, and found that 'the boundary layer is typically shallower, drier, and well mixed near the shore'. Using several years of cruise measurement along the 20°S in October, De Szeoke et al. (2010) also stated that over the SEP, 'SST, air temperature, sea surface saturation humidity, air specific humidity, and wind speed all increase to the west along 20°S'. According to our WRF-Chem simulation (AERO), MBL is warmer and more humid during November 1-16 compared to that during October 15-31 over both the remote and coastal regions (see Fig. R1). The RB was sampling over the coastal region during November 1-16. Therefore, without the sampling time differences in the RB observations, the observed coastal and remote contrast in MBL temperature and humidity would have been even stronger.

However, the questions regarding the humidity and temperature contrast, and the associated variability (questioned by the 3rd referee) have led to further investigations. We found that there is a newer version of radiosonde data available (an old QC version were used). In addition, it was found that the larger variability seen in the remote radiosonde profiles was due to the inclusion of a few profiles measured in lower latitudes. We have reprocessed the radiosonde data using the most recent version (v4.2) data and applied a latitude constraint of 20 ± 2.5°S for this comparison. As suggested by the reviewer, the BAe-146 dropsonde data have also been included in Fig. 1 with the addition of two panels for the coastal regions based on the dropsonde measurements. The BAe-146 dropsonde data are only used for comparisons over the coastal regions in Fig. 1, since the remote region is 78-90°W in longitude, and nearly all of the BAe-146 dropsonde were located at the coastal region (east of 78°W) with the exceptions of two dropsondes on November 13 (one dropsonde at ~79.83°W and one at ~78.98°W), and they are unlikely to be representative of the defined remote region (78-90°W) in this study. The BAe-146 dropsonde data have also been included for studying longitudinal variations of the MBL heights (top panel of Fig. 2). Since BAe-146 dropsonde data were released within a narrow time window during the day (14-16 UTC hour), they have not been included in the diurnal variability plot (bottom panel of Fig. 2). The mean MBL temperature and humidity differences over the remote and coastal regions are not statistically significant anymore after the new processing (although we tend to believe the contrast might be disguised due to the sampling time issue). Note that the temperature and humidity contrast in the near surface layer are still statistically significant (with 98% confidence level) in both the RB observations and model simulations, which are mentioned briefly in the discussion section of the revised manuscript in the context of surface fluxes. The related discussion in the main text has been modified according to the newly obtained results and the addition of the BAe-146 data. Some main revisions are as follows:

"…The coastal region has a stronger temperature inversion with a 10–12 K increase in θ_v within inversion layers (also see Table 1). The mean observed humidity and temperature in the MBL over the coastal and the remote regions, however, are not statistically significant (at 98% confidence level) in the difference between the two regions.

Differences between AERO and MET in mean profiles of θ_v and q_v are small, in general, except within the simulated inversion layer. In the coastal MBL, mean temperatures in the MBL from both simulations have biases of -0.6 and 0.6 K compared to radiosondes and dropsondes, respectively. The positive bias in mean temperature compared to dropsondes originates from biases in the upper MBL where the mean simulated temperature deviated from observed well-mixed values (Fig. 1). Over the coastal region, the simulated q_v from both simulations, on average, are biased high within MBL (biases of 0.5-1.1 g kg^{-1}) and in lower free troposphere (above MBL and < 2 km, biases of 1.6-2.2 g kg^{-1}) with more significant biases seen in the comparisons with the dropsondes. The larger biases in AERO and MET compared to the dropsondes are associated with the larger biases in model predictions during Nov. 2-10, when the observed temperature inversion is weaker and the vertical variability in humidity is large (with possible presence of multi-cloud layers). Toniazzo et al. (2011) also noted early November to be a period with reduced synoptic-scale variability, lower inversion heights and increased cloud cover. The RB had measurements on Nov 2-3 before a 6-day break in sampling and the mean profiles were less affected by profiles measured during this synoptic episode. Over the remote region, the simulated mean temperature and humidity are in excellent agreement with AERO simulations with the simulated mean values being not statistically different (at 98% confidence level) from observations at most vertical layers from the surface to 3 km. Over this region, the biases of the simulated mean profile in MET are mostly within the inversion layer."

The conclusion about the coast being colder than the remote area due to SST differences is actually backed up by the statistics in Table 3. However, this sentence has been deleted in the revised manuscript due to the new results.

We said in the manuscript that 'temperature and humidity contrasts between coastal and remote regions might be related to cooler SSTs near the coast in addition to differences in cloud characteristics between the two regions'; however, we did not try to establish a direct relationship between the SST and MBL height here. The referee might be confused by our statements regarding the impact on the MBL humidity and temperature in the two regions versus the impact on the MBL height. The changes in thermodynamics might be related to the changes in MBL height, but we don't see a clear and straightforward relationship; for example, Bretheton et al. (2011, J. Adv. Model. Earth Syst.), pointed out that 'In marine stratocumulus-capped boundary layers under strong inversions, the timescale for the thermodynamic adjustment is roughly a day, much shorter than the multiday timescale for inversion height adjustment'. As also pointed out by the referee, the MBL height is influenced by different processes at different scales. That is why MBL heights do not appear to be tightly related to variations in air-sea temperature differences. This sentence, however, has been deleted in the revised manuscript due to the aforementioned new analysis results.

The AERO predicted mean MBL height is about 130 m (a more exact number is used in the revised version rather than round it up to multiple of 50s such as 100, 150, etc.) lower than the MET predicted values, and is biased low by about 110 m compared to radiosonde observations. To better address the reviewer, we conducted additional investigations regarding the potential reasons for this reduction. It was found that the mean subsidence rate at the MBL top predicted by AERO is ~0.15 cm s^{-1} higher than that of the MET simulation. The domain-average entrainment rate in the AERO simulation is about 24%

lower compared to that of the MET simulation (0.0088 cm s^{-1} vs. 0.0067 cm s^{-1}). We have added these new diagnostic results to our explanation of the reduced MBL depth in AERO in Sect. 4 as follows:

"The horizontally and temporally averaged subsidence rate in the layers above the MBL is consistently stronger in AERO than in MET; at 850 hPa, the average subsidence rate is 0.24 cm s^{-1} in AERO, which is about 0.015 cm s^{-1} higher than in MET. The AERO-predicted MBL top entrainment rate (0.67 cm s^{-1}) is about 24% smaller compared to that of the MET simulation (0.88 cm s^{-1}). The entrainment rate was estimated using the Eq. (4) of Yang et al. (2009) with 5-min model outputs of tracer concentrations. It is worth noting that the estimated entrainment rate is larger than the nighttime entrainment rate (0.4 cm s^{-1}) estimated by Yang et al. (2009) using budget analysis of measured DMS during VOCALS-REx. This is very likely due to our vertical resolution near cloud top and in the inversion layer. LES model simulations show entrainment is very sensitive to the vertical resolution and generally use finer resolution (e.g., 5 m in Bretherton et al. 2011) than in our simulations. However, there is no reason to expect that the bias (relative to the Yang et al. estimate) would differ significantly between the AERO and MET simulations, so the 24% difference in entrainment rates between the two simulations should be meaningful. The 130-m lower MBL depth in AERO, compared to MET, could thus be explained by the stronger entrainment and somewhat stronger mean subsidence in AERO. "

In Sect. 5 and the abstract, "Particularly, inclusion of interactive aerosols in AERO strengthens the temperature and humidity gradients within the capping inversion layer and lowers the marine boundary layer depth by 150 m from that of the MET simulation. " has been revised to "Particularly, inclusion of interactive aerosols in AERO strengthens the temperature and humidity gradients within the capping inversion layer and lowers the marine boundary layer depth by 130 m from that of the MET simulation, which are associated with the weaker entrainment and stronger mean subsidence. "

5/ Figure 4: Linked to point 2 above, I recommend adding 146 AMS data to this figure. Also, how are the data weighted in this figure? Have the data been properly weighted for sampling frequency/total sampling time? Also, how did the authors ensure that all the measurements used were in the MBL (i.e. was an altitude constraint or some other proxy used)?

Response 1.5.0: As suggested by the referee, we have added BAe-146 AMS data to this figure. A simple arithmetic average of all the valid AMS/CI/PILS data points for a particular platform and region were used in Figure 4. Model results were compared against corresponding observational data, so the sampling frequency and total sampling time were taken into consideration. In Fig. 4, measurements from different instruments/platforms were compared against model predictions individually. We used a combination of the flight elevation, relative humidity (or difference in temperature and dew point), and liquid water content measurements to ensure that the measurements are within the MBL. We have added the clarification into Sect. 2.2.2.

6/ Figure 8 and interpretation on P. 22683: In the final para on this page, the authors used Wyoming Cloud radar data on the C-130 to intercompare cloud thickness and cloud base with model data and conclude that they are in excellent agreement. Certainly, the mean data do compare excellently and as the authors rightly point out, there are important subtleties in the tails and shapes of the three different distributions. However, this conclusion is at odds with the suspect earlier analysis of the RB sonde data – here the means compare very well, whereas the means of MBL depth diagnosed from the sonde data disagree by up to 150 m in the AERO case. If cloud base and cloud thickness agree well when compared to radar data, then by proxy,
MBL depth would also agree well in the radar intercomparison. This casts further doubt on the earlier interpretation of the RB sonde data and may serve to suggest that WRF-CHEM is doing even better in terms of thermodynamics than currently concluded.

Response 1.6.0: The MBL heights calculated from the RB radiosondes agree well with the cloud-top heights from the C-130 cloud radar. As shown in Table1, the mean PBL height is 1360 m based on the RB, and is ~1330 m (991 m base + 341 m thickness as shown in Table 1) based on the C130. The direct comparisons indicate that AERO-predicted cloud tops are about ~160 m lower than the C-130 observations. We thank the referee for the very careful reading. Further investigation uncovered an error in the calculation that the mean values of the cloud base and thickness in AERO were replaced by those from observations. It has been corrected in Table 1 and in the upper right corner of Fig. 8. This mistake did not affect frequency distributions shown in Fig. 8. The mean AERO-predicted cloud thickness now agrees with the C-130 observations within ~25 m. Comparisons with both radiosonde observations and cloud top observations from the C-130 indicate an underprediction of MBL heights in AERO. Related discussion has been revised as follows:

"Both AERO and MET modeled cloud base and cloud thickness are in excellent agreement with the observations (Fig. 8 and Table 5), with better estimates seen in AERO (mean biases of < 7 m for mean cloud base height and < 1 m for mean cloud thickness). " has been changed to "Both observation and simulations show that cloud thickness increases with distance from the coast. Both AERO and MET modeled cloud thickness are in excellent agreement with the observations (~340 m), with ~25 m low and high mean biases (Fig. 8 and Table 5), respectively. While the mean cloud base height from the C-130 over the remote region (~1100 m) is within the 1000 - 1200 m range (median) reported by Bretherton et al. (2010) for their transition and remote regions, over the coastal region our observed mean cloud base height (~900 m) is lower than their cloud base height (median, 1000 m), which is most likely due to their inclusion of only flight data along 20°S. Cloud base heights are better predicted in MET than in AERO. The underprediction (~160 m) of the mean cloud base height in AERO compared to observations on the C-130 is consistent with its low bias (mean bias of ~110 m) in predicted MBL heights compared to radiosonde measurements on the RB. "

7/ Section 3.6: Rain rate. The 2DS instrument data from the BAE-146 may be a useful addition to this analysis.

Response 1.7.0: As suggested, rain rates based on the 2D-C measurement on the BAe-146 have been added to Table 7. The paragraph below has been added to Sect. 3.6.

"The in-cloud and near-surface rain rates from the C-130 and BAe-146 agree within 30% over the coastal region with the differences most likely associated with different particle size detection ranges of the instruments. This good agreement provides additional confidence in the derived rain rates from both 2D-C probes in this study. The rain rates over the remote region from the BAe-146 are much smaller than those measured by the C-130 and are not representative of the entire remote region since the BAe-146 2D-C measurements only covered a small longitudinal range over the east edge of the remote region (78-81°W with mean longitude of 79°W). Therefore, the rain rate discussion that follows is based entirely on C-130 observations."

Below is a question from a colleague working directly with WRF-CHEM:

8/ Cumulus parameterisations (or lack thereof) in the model settings: The model is producing an order of magnitude less rainfall than was observed - yet the authors don't seem to have tried to find out if including one of the sub-grid cumulus parameterisations which are included with WRF-CHEM would bring their predicted rainfall closer to the measurements. If the intention is to find out how the aerosol phase affects the model resolved cloud formation only then they shouldn't only be running the model at 9km (which is too coarse to really get the cloud dynamics right as noted by the authors), but instead should be running at a higher resolution instead (say 3 km, or even smaller). What I think is needed is either a better explanation of why they've not included any sub-grid cumulus parameterisation or/and some sensitivity studies showing what effect including sub-grid cumulus parameterisations would have on their model results. Failure to do so ultimately undermines the ability to make any useful conclusions for the modelling community from the current model output in the context of coupling prognostic aerosol with cloud microphysics.

Response 1.8.0: The decision to omit a sub-grid cumulus parameterization in our simulations was due to the following reasons:

1) The 9 km horizontal grid spacing is at the upper end of the grey area where cumulus clouds and their effects are not entirely subgrid but are not resolved either.

2) In the VOCALS–REx study, the predominant clouds were marine stratocumulus, which are generally not treated well by cumulus parameterizations. Toniazzo et al. (2011) noted that broken cumulus, which the cumulus parameterization could potentially treat within the model, were infrequent.

3) We conducted a 9-day test run (including a 5-day spinup) using the Kain-Fritsch cumulus scheme, and no apparent improvements in cloud properties were found. For example, the cumulus parameterization led to a large reduction in cloud fraction along the southern

inflow boundary region (see Fig. R2).

4) Current cumulus schemes in WRF-Chem have not yet been coupled with aerosols. Such coupling has not been done for most climate models as well.

5) Resolving shallow cumulus would require a grid spacing of 1 km or less, and it is unrealistically expensive to conduct WRF-Chem simulations at the AERO settings at this resolution over the same domain size for the month-long period.

As suggested, we added a brief justification in sect 2.1:

"Sub-grid cumulus parameterizations were turned off for the simulations in this study. The predominant clouds during VOCALS-Rex were marine stratocumulus, which are generally not treated well by cumulus parameterizations. In a short 9-day test run using the Kain-Fritsch cumulus scheme, there was no apparent improvement in simulated cloud features and precipitation. Also, the cumulus schemes currently in WRF-Chem have not been coupled with interactive aerosol."

Specific (minor) Comments:

Abstract: Line 8 – may be worth explicitly stating which satellite measurements, i.e. MODIS, CERES and GOES-10.

Response: It has been changed as suggested.

P. 22670, line 1: Forgive my ignorance but what is “scalar advection”? Is this anoxymoron - isn’t advection a vector quantity by definition? If you mean bulk mass exchange through grid boxes or the like, then please restate or qualify further for non-modellers.

Response: The “scalar” here follows the same definition as in physics and mathematics. Examples of scalars are tracer (aerosol, chemical species, and hydrometeor) number mixing ratios. We have provided examples of scalars to help readers who are not familiar with the terminology, and “scalar advection” has been replaced with “advection of scalar quantities”.

P. 22675, line 29: Can the observed sonde variability really solely be due to open cell dynamics as stated? I would amend to say that ONE source of variability MAY be POCs and quote the other sources of variability (synoptics, SSTs etc).

Response: Thanks for your suggestion, however, the related sentence has been deleted in the revised manuscript corresponding to other revisions.

P. 22676, line 20: Which "large-scale dynamics" are being referred to here and what may be the proposed process link between cloud bulk property changes due to interactive aerosol and changes at the larger scale? I can't see cloud bulk properties influencing mean MBL flow or subsidence rate for example. This idea is used rather loosely and ambiguously as a potential source of reducing MBL depth in the model by a significant 150 m versus the observations. Are there other potential reasons for this reduction in depth that can be diagnosed from WRF-CHEM? This is perhaps one of the most important thermodynamic errors that warrants further discussion.

Response: The "large-scale dynamics" was loosely used here to refer to "the mean thermodynamics and mean flow". We have changed the wording to clarify. Clouds and aerosols are important to radiative fluxes, thus energy budget. In the model, the microphysics scheme is coupled with radiation, and various aerosol effects (e.g., direct effect, semi-direct effect and 1st in-direct effect) are considered, so the changes including interactive aerosols could potentially modify the mean thermodynamics and mean flows within MBL and above MBL through altering the radiative fluxes, and thus energy budgets.

We have conducted further investigations regarding the potential reasons of this 130 m MBL height reduction in AERO, please see part of Response 1.4.0 for more details.

Section 3.2: Here is where I think the authors really need to fully discuss their results alongside those of Allen et al., 2011 and Bretherton et al., 2010.

Response: As shown earlier, we have added additional discussion to compare our results with those in Allen et al. (2011) and Bretherton et al. (2010). Additional discussion has also been added to the main text for comparisons of rain rates from this study and those from Bretherton et al. (2010) as follows:

"Our calculated mean rain rates are higher than those of Bretherton et al. (2010) derived from a different 2D-C probe and the maximum radar reflectivity. However, the median values of our in-cloud and near-surface rain rates agree reasonably well with those of the radar derived rain rates in Bretherton et al. (2010). We used the precipitation sizing data observed using the 25-μm resolution 2D-C probe which is more reliable than the other 10-μm resolution 2D-C probe on the C-130, of which the true resolution was later found unstable (personal communication, Allen Schanot, 2010). "

P. 22684, line 17: Can you quote the time frequency referred to?

Response: The time frequency is every 10 minutes, and it has been added to the text.

P. 22688, line 4: "horizontal advection may affect tendencies of MBL depth when its gradients exist along wind directions (Rahn and Garreaud, 2010)" – I don't understand this, can you clarify or rephrase?

Response: We have rephrased to clarify: " the local change of MBL depth with time may be affected by the horizontal advection of the temperature and moisture by winds at the top of the MBL (Rahn and Garreauh, 2010);"

Below are some further comments from colleagues more familiar with WRF-CHEM and aerosol production:
Page 22679 lines 21-26: The authors say that their model does not include oceanic emissions of organic compounds and yet they say in clean maritime air masses the contribution of organic compounds to organic matter could be as high as 71

Response: The current version of the WRF-Chem does not include oceanic emissions of primary organic aerosol, which leads to the underestimation of OM over the ocean. Overall, according to Hawkins et al. (2010), "Measurements of submicron organic aerosol functional groups and trace elements show that continental outflow of anthropogenic emissions is the dominant source of organic mass (OM) to the southeast Pacific with an additional, smaller contribution of organic mass from primary marine sources". To avoid confusion, we rephrased the sentence as follows:

"According to Hawkins et al. (2010), OM over the SEP has a dominant contribution from anthropogenic sources, and an additional, smaller contribution from primary marine sources based on measured functional groups and trace elements during VOCALS-REx."

Page 22689: modelled DMS is over-predicted by a factor of 3 and yet MBL SO2 is underestimated by a factor of 3. Speeding up DMS oxidation shows little sensitivity to modelled sulfate. There seems to be a huge disconnect here. Perhaps, as they say wet removal of sulfate is too high but I agree with the authors that this needs further investigation. – Perhaps this issue should be restated in the conclusions, given its importance to VOCALS and beyond?

Response: The point is that in the model, the predominant DMS loss pathway in the MBL is oxidization to SO_2, and the predominant SO_2 loss pathway is oxidation to sulfate in clouds. For a multi-day period that exceeds the DMS and SO_2 lifetimes, the DMS emissions balance the SO_2 production. Speeding up the DMS oxidation lowers the DMS gas concentration, but has little impact on either the SO_2 production or the subsequent sulfate production. As suggested, the following sentence has been added to the last paragraph of the Conclusions: "Future work related to model development includes improvement of the DMS emission scheme, further investigations of the underprediction of MBL SO_2 and the overestimation of supermicron sea salt, etc. "

P. 22690, line 10: How does drizzle inhibition lead to high stability? I would have thought that drizzle acts to stabilize the MBL through cooling. Perhaps there's a chicken or the egg issue to disentangle here in how this para is phrased – do you mean that a background stable MBL leads to less drizzle formation and hence lower CWP? In which case, it is the stable atmosphere that is the controlling factor and not the moisture supply due to drizzle inhibition as suggested.

Response: According to an LES modeling study by Jiang et al. (2002), "Drizzle evaporation below cloud base is the main process that influences the stability of the air layer below cloud base. In cleaner air, cooling from evaporating drizzle destabilizes the layer just below cloud base with respect to the surface and promotes the formation of thicker clouds, thus higher CWP. In polluted air, the high CCN concentration leads to smaller droplet sizes, thus formation of drizzle is reduced, and the reduced drizzle below cloud base results in a relatively stable air and a less effective supply of surface moisture into cloud ". We have rephrased the related text to clarify: “Despite the highly complex relationship between drizzle and stability, the LES modeling study by Jiang et al. (2002) showed that, relative to a clean environment, where cooling from drizzle evaporation below cloud base destabilizes the slice of layer below cloud base (relative to the surface), drizzle inhibition in polluted air leads to a relatively stable atmosphere and a less effective supply of moisture from the surface to cloud layer, resulting in low CWP. ”

P. 22690 – this discussion needs to be re-examined in light of the new sonde/MBL structure interpretation.

Response: This has been addressed in the Response 1.4.0.

Technical Corrections:
P. 22668, line 23: “cloud-borne”? You mention interstitial aerosol earlier – does this word relate to activated aerosol? If so, I would just say “activated” as cloud-borne would cover all cloud aerosol.

Response: The suggestion is taken.

P. 22670, line 8: Change “. . .is used that. . .” to “. . .is used, which. . .”

Response: It has been used as suggested.

P. 22675, line 12: You define free troposphere as < 3km – does this include the MBL and the inversion as well?

Response: We meant the layer above MBL below 3 km. We have clarified this in the revised manuscript.

P. 22677, line 9, Change “. . .the variation...” to “. . .the true known variation. . .”

Response: I think you meant line 4 rather than line 9 since line 9 is a subtitle line. The related sentence has been deleted in response to other revisions.

P. 22678, line 15: change ". . .number, they. . ." to ". . .number. They. . ."

Response: Rephrased.

P. 22680, line 21: Change "detect limits" to "detection limits."

Response: Corrected.

P. 22681, line 4: Change "Huneess et al. . ." to "Huneeus et al. . ."

Response: Thanks for pointing this out. Corrected.

P. 22683, line 5: Change ". . .associate. . ." to ". . .associated. . ."

Response: Corrected.

P. 22681, line 5: It is not clear whether the enhancements referred to are in the model or in the satellite data without looking at the plot. Perhaps amend to "broader band of enhancements near the Peruvian coast in AERO".

Response: Changed as suggested.

P. 22686, line 6: Change to "near-surface".

Response: It has been changed as suggested.

P. 22688, line 5, Change ". . .turbulence ..." to ". . .turbulent. . ."

Response: Corrected.

P, 22688, line 15, change "limits" to "supresses".

Response: Changed as suggested.

P. 22690, change “numbers at a polluted. . .” to “numbers in a polluted. . .”

Response: Changed as suggested.

P. 22690, change to “. . .found that a simulated. . .”

Response: Changed as suggested.

P. 22690, line 11: change to “leads to. . .”

Response: Corrected.

P.22691, line 4: Change “clearness..” to “clearance. . .”

Response: Changed ‘clearness’ to ‘cloud clearance’.

Acknowledgements: The following names and datasets are spelled wrong: “A Grant” should be “G Allen”. “GEOS-10” should be “GOES-10”. “B Alan” should be “A Bandy”. Please check that this list is complete as per item 1 in my general comments.

Response: Corrected.

Response to Anonymous Referee #2

General Response: We thank the anonymous referee for the constructive comments and suggestions that are helpful for improving this manuscript. Detailed responses to the comments and revisions to the manuscript are provided below.

This manuscript presents regional scale modeling results from WRF-CHEM simulations that include prognostic aerosols and a representation of aerosol-cloud interactions in the South-East Pacific region. The region contains one of the Earths largest and therefore climatically important marine stratocumulus decks. The role that continental and marine aerosols play in modulating the stratocumulus cloud in this region are however poorly understood. The authors utilize in-situ and remote sensing observations from the vast data-set collected during VOCALS-REx to evaluate the model performance. The paper is interesting and certainly addresses scientific questions that are relevant to the journal. I would therefore recommend publication in ACP once the authors have addressed the following comments.

1. The main strength and novel scientific work described in this paper is that it is the first regional/global modeling study from VOCALS-REx that I am aware of in which the model includes a detailed representation of aerosols and their interactions with marine stratocumulus. It is therefore ideally placed to try and address some of the key VOCALS scientific hypotheses that relate to aerosol-cloud-drizzle interactions (Wood et al., 2011). The authors go some way in achieving this by showing that including aerosols leads to significant changes in the modeled marine stratocumulus e.g. cloud top effective radius, LWP, albedo, drizzle etc. However all of the results presented are averages over the VOCALS-REx time period and I feel that the authors could significantly improve the manuscript with a description of daily/synoptic scale variability. For example while the simulations capture a monthly mean gradient in aerosol, it strikes me that you could just use a climatology of aerosol or cloud droplet number concentration in a model and get the same result. In my mind WRF-CHEM is a model that should be looking to capture some of the interesting events seen in VOCALS-REx such as discrete aerosol plumes being advected away from the coast as was observed in satellite imagery and in-situ observations. After all, including prognostic aerosols in a regional/global model is computationally expensive and the benefit of doing so (instead of using a climatology) needs to be highlighted.

Response 2.1.0: This is an excellent point. As suggested, we have added Fig. R3 to Sect. 3.2.1 to illustrate the capability of the model in capturing daily/synoptic scale variations such as outflow events. The paragraph below has been added to the main text:

"Aerosol and CCN number concentrations and N_d over the SEP are strongly influenced by pollution outflows from the continent. In Fig. 10 of Bretherton et al. (2010), daily MODIS-derived N_d were compared against aircraft measurements, and it shows the occurrence of a few strong outflow events along 20°S over the SEP. The longitude-time plot of model

(AERO) predicted CCN concentrations (at 0.1% supersaturation) at 975 hPa are shown in Fig. R3 (left panel), and it succeeds in capturing the timing and strength of the observed outflow events shown in Fig. 10 of Bretherton et al. (2010). During the VOCALS-REx, the strongest pollution outflow event along 20°S peaked on October 18, and the cleanest period was around November 8. The four contour plots on the right panel of Fig. R3 illustrate the horizontal distribution of MODIS-derived and model-predicted N_d during the two time periods, respectively. Model simulated N_d compares reasonably well with observations. The model reproduces the outflow pattern from coastline towards the ocean with a band of high N_d several degrees wide in longitude along the coast. Considering the relatively large uncertainties in satellite derived N_d due to averaging over only several instantaneous satellite snapshots and the timing differences between satellite overpasses and model outputs, the agreement in the outflow patterns is remarkably good. The model also captures N_d spatial patterns during the clean event. The accurate predictions of both events demonstrate the model's ability to capture daily/synoptic scale variations of aerosol and clouds, and suggest that the model is suitable for studies at such scales (e.g., pollution outflow study), which is another advantage of using WRF-Chem with prognostic treatment of aerosols and cloud-aerosol interactions."

2. The non-aerosol simulations used a fixed cloud droplet number concentration of 250 cm-3 which doesn't seem to be the most suitable choice for marine aerosol conditions. Could the authors provide some justification for this choice? I imagine that if a more representative value of the off-shore conditions (_ 100 cm-3, fig 3) was used then the non-aerosol simulation would look very different.

Response 2.2.0: We agree that a fixed cloud droplet number of 250 cm^{-3} is more representative for the coastal region affected by pollution outflows but not for the offshore marine aerosol conditions. We did test a lower N_d of 150 cm^{-3} (representing the domain average N_d derived from the C-130 measurements) uniformly applied across the model domain. The lower N_d does not change the main conclusions drawn from the N_d = 250 cm^{-3} case. Also, since N_d = 250 cm^{-3} was the default setting in the Morrison microphysics, keeping this value facilitates comparison with other WRF modeling studies without prognostic aerosols that likely use the same setting.

The simulated cloud fraction is not sensitive to the decrease in N_d, although cloud thickness is reduced by about 20 m over the remote region. As shown in Fig. R4, cloud water path and cloud optical thickness differences between the two cases are very small (less than 3%), which is negligible compared to the about 20% and 60-100% high biases in CWP and COT, respectively, in the MET simulation. There is a 14% increase in r_e which is directly affected by N_d, but it is not large enough to compensate the 37% low bias, and it does not impact radiation in the model since N_d is not coupled to radiation schemes in the MET simulation. As expected, the change of N_d directly impacts precipitation: reducing N_d to 150 cm^{-3} increases rain rate (by ~35%) over the remote region in MET. We have accordingly modified the manuscript to include the effects of changing N_d from 250 cm^{-3} to 150 cm^{-3} in related discussion in Sects. 3.3, 3.4, and 3.6.

We have replaced "The high and uniform droplet number concentration in MET not only causes the underestimation but also limits the variability of r_e." in sect. 3. 3 with "Although reducing the constant N_d of 250 cm^{-3} to a more representative average droplet number concentration of 150 cm^{-3} increases domain-average r_e by 14%, the uniform droplet number concentration in MET still limits the variability of r_e".

Also in Sect. 3.3, "Larger CWP in MET is likely due to the high cloud droplet number (250cm^{-3}) that substantially suppressed autoconversion and drizzle." has been revised to: "In MET, changing the constant N_d from 250 to 150 cm^{-3} only reduces CWP by 3% over the remote region, although the rain rate over this region increases by about 35%; this is due to the very weak drizzle simulated in the MET, which is discussed in more detail in Sect. 3.6".

The following sentence has been added in Sect 3.4, 'In MET, when using a constant N_d of 150 cm^{-3}, cloud thickness is reduced by 20 m over the remote region, agreeing better with observations.'

3. The authors need to justify why no convection scheme was used. The model horizontal resolution (9 km) is certainly not high enough to explicitly resolve convection.

Response 2.3.0: We thank the referee for bringing this up. We have already addressed a similar question from referee 1. Please see Response 1.8.0 for details.

Response to Anonymous Referee #3

Summary

This paper uses the VOCALS-REx field observations and coincident satellite data in the southeast Pacific to demonstrate the utility of the regional WRF-Chem model in representing marine stratocumulus and its interactions with aerosols. In addition to comparing model output to observations, this study also compare runs with and without interactive aerosols to show the improvement made by modeled aerosol-cloud interactions. The authors focus on average thermodynamic, cloud and aerosol variables over the month of observations, as well as quantities related to boundary layer structure and energy balance. This is a good first step in model evaluation and motivates further use and evaluation of WRF-Chem to studying aerosol effects on marine stratocumulus. The authors find that the simulations including aerosol-cloud interactions compare fairly well to observations, and perform better than simulations without aerosols and a fixed droplet concentration. Biases in some meteorological and aerosol variables are noted and potential impacts on other simulation variables are explored. Temperature and humidity biases may influence errors in other variables. Although much more complete and detailed comparisons making use of more VOCALS observational data and more simulations focused on isolating aerosol-cloud interactions should be done in the future, this paper succeeds in demonstrating WRF-Chem is a good tool for reasonably representing the VOCALS-REx mean characteristics. The significant improvements seen by including interactive aerosols seem overstated (discussed in comment section) based on the specifications of the simulations used to make the comparisons. However, the conclusions drawn from the results, such as 'including spatially varying aerosol characteristics is important when simulating marine stratocumulus over the southeastern Pacific", don't overstep the boundaries of what one could determine from the simulations used. It is clear a large amount of work went into producing these results, and these results should be published before more complicated metrics are used to evaluate WRF-Chem. For these reasons I recommend this paper be published following minor revisions.

General response: We thank the referee for the constructive comments and suggestions, which are helpful in improving the manuscript. Below are our detailed responses to the specific comments:

General Comments

1. The authors address the limitations of comparing the specific AERO and MET cases in the summary/conclusion, but restrict the impacts to the variability of microphysics. Some of the improvements in the AERO run could be due to simply representing the spatial gradient in droplet concentration, as implied by the last sentence of the abstract. Reproducing the mean droplet concentration does not require interactive aerosols. Fixing the droplet concentration to a constant in time, but with a gradient representative of mean southeast Pacific conditions would likely also show improvement over the MET simulation, not necessarily just in microphysics. The way this section is written now implies that the improvements seen in MBL structure, energy fluxes at the TOA and surface, and macrophysical cloud properties would be the same if the MET case included a spatial or temporal gradient in the fixed droplet concentration. The statement addressing this issue should be made more inclusive of potential impacts of using a different MET simulation.

Response 3.1.0: We thank the referee for pointing this out. We agree that the importance of including a spatial or temporal gradient in aerosol or cloud droplet concentrations were overstated in this section. Comparisons of MET simulations with different fixed droplet number concentrations (the original 250 cm^{-3} and a new one with 150 cm^{-3}) have been added to the revised manuscript as addressed in Response 2.2.0. It is found that changing the constant droplet number in MET impacts rain rates significantly but leads to relatively small changes of cloud properties in the MET simulation. This is because the droplet number is not coupled with the shortwave radiation scheme in the MET simulation. Thus, the inclusion of full cloud-aerosol couplings and various aerosol effects including the first direct effect and indirect effects are important to the improvements seen in MBL structure, energy fluxes at the TOA and surface, etc. The last sentence of the abstract has been revised to "The overall performance of the regional model in simulating mesoscale clouds and boundary layer properties is encouraging and suggests that treating cloud-aerosol interactions is important when simulating marine stratocumulus over the southeastern Pacific." The related discussion in the summary and conclusion has been revised as follows.

“The well-predicted aerosol quantities such as aerosol number, mass composition and optical properties lead to significant improvements in many features of the predicted stratocumulus clouds…” has been replaced with “The well-predicted aerosol quantities, such as aerosol number, mass composition and optical properties, and the inclusion of full aerosol-cloud couplings lead to significant improvements in many features of the predicted stratocumulus clouds…”

“This study compared two extreme cases: one with prognostic aerosol and the other with a cloud droplet concentration that is fixed in space and time. It might be possible to reasonably capture the spatial and temporal variability of the microphysics using prescribed temporal/spatial gradients of droplet concentrations or background aerosol.” has been revised as the following: “This study compared two extreme cases: one with prognostic aerosol and full aerosol-cloud couplings, and the other with a cloud droplet concentration that is fixed in space and time, and simplified cloud and aerosol treatments in radiation scheme. It might be possible to reasonably capture the spatial and temporal variability of some microphysical variables using prescribed temporal/spatial gradients of droplet concentrations or background aerosol with simplified cloud and aerosol treatments in the radiation scheme.”

2. The ’MET’ simulation fixes the droplet concentrations at a very high value that causes a lot of the biases seen (addressed in the paper) in this simulation. It is not emphasized enough (though is briefly in the summary) that this may cause the improvements noted AERO simulation to appear larger and more significant than they might be if a different constant droplet concentration were used.

Response 3.2.0: We thank the referee for bringing this up. We have already addressed a similar question from referee 2. Please see Response 2.2.0 for details.

3. It would be helpful if the paper would describe the ’MET’ simulation in more detail, specifically what it represents physically. What does ’rain’ mean if droplet concentration doesn’t change? If the aerosol module is turned off, why is it necessary to prescribe droplet concentration to a fixed value? One possible explanation is if droplet concentration can change, droplet loss to rain would represent

a loss of aerosol, even though there is no aerosol module. A cloud-aerosol interaction would therefore still exist. Although this is perhaps intuitive and obvious, adding a description would keep readers from guessing what the MET simulations represent. In the MET simulations is there a fixed aerosol impact on radiation?

Response 3.3.0: In principle, one could allow the droplet concentration to change even if there is no coupling with an aerosol module, for example, if the droplet activation is related to a background aerosol that is constant in time. The use of the constant concentration in MET is because this is how droplet concentration is treated in the default scheme in WRF. In the Morrison double moment microphysics, the cloud droplet number concentration and cloud water mixing ratio are used to determine the cloud droplet size spectrum, which affects various microphysical process rates (e.g., autoconversion). Cloud water is predicted in the microphysics scheme based on sources (primarily condensation of excess water vapor) and sinks (e.g., autoconversion and accretion). When aerosol modules are used, N_d is predicted based on sources (aerosol activation) and sinks (which parallel the cloud water sinks), while in the MET simulation N_d is prescribed with a constant value. The parameterized autoconversion rate depends on both the cloud water and droplet number, whether droplet number is predicted (and varies spatially and temporally) or is prescribed. Thus the rain/drizzle also depends on droplet number. Another difference between AERO and MET is that in the MET simulation, the predicted droplet sizes and number concentrations from the microphysics scheme do not feed into shortwave radiation scheme, where the predicted cloud water and a constant effective radius are used.

As suggested by the reviewer we have added the following discussion to the Sect. 2.1 to describe the MET simulation in more detail.

“In the MET simulation, within the Morrison double-moment microphysics scheme cloud water mixing ratio is predicted but cloud droplet number concentration is prescribed as a constant value, aerosols have no impact on cloud microphysics. However, this constant droplet number affects the autoconversion of cloud water to rain water and thus affects the rain water and raindrop number concentrations. Another difference between AERO and

MET is that in the MET simulation, the predicted droplet sizes and number concentrations from the microphysics scheme do not feed into shortwave radiation scheme, where the predicted cloud water and a constant effective radius are used."

4. It is stated in the abstract that the Morrison microphysics scheme is used because it allows for two-way aerosol cloud interactions, and the paper mentions that it is newly connected to the aerosol code, but specifically how and why this scheme is preferable is not explained. The Morrison scheme is well described, but it is not clear in the paper why it is better than the past used Lin scheme. The Lin scheme also allows for 'two way' aerosol cloud interactions, and although was originally single moment, has been modified to behave as a double moment scheme. What is the benefit to using this new double moment scheme? Have there been runs or studies comparing the Lin scheme to the Morrison scheme in WRF-Chem? Since the reasons for/benefits to using the Morrison scheme aren't described or cited in the paper, either an explanation should be added to section 2.1 or the fact that the Morrison scheme is being used shouldn't be the leading line in the abstract. What is different/better about this microphysical representation, especially with regard to aerosols?

Response 3.4.0: The Lin scheme was the only microphysics scheme coupled with interactive aerosols in WRF-Chem before the release of V3.3 with the Morrison microphysics. The Morrison microphysics has become a popular scheme in the WRF community because of its performance as well as the fact that it was the very first scheme in WRF which is fully two-moment for all species, including rain/drizzle, and the Lin scheme coupled in WRF-Chem is only two moment for cloud water. The selection of the Morrison microphysics as the new microphysics scheme to couple with aerosol modules is also based on the frequent requests from the WRF-Chem users. We are not aware of studies that have compared the Morrison and Lin schemes for stratocumulus, but improvements using the Morrison scheme compared to the Lin scheme have been noted for other cases such as a mesoscale convective system (Luo et al., 2010, JGR). We did perform some sensitivity simulations with Lin and Morrison microphysics schemes (with YSU PBL scheme) and found that the Morrison scheme predicted better domain-average cloudiness in this preliminary testing. The methodology of coupling the aerosol modules and the microphysics is similar for both Lin and Morrison schemes. The use of the Morrison microphysics in this evaluation is also intended to inform the WRF-Chem community the availability of this new functionality.

We have rephrased the first sentence in the abstract.

As suggested by the reviewer we have modified and moved the first sentence of the abstract, and now it is not the leading line of the abstract anymore. "In the recent chemistry version (v3.3) of the Weather Research and Forecasting (WRF-Chem) model, we have coupled the Morrison double-moment microphysics scheme with interactive aerosols so that two-way aerosol-cloud interactions are included in the simulations." has been replaced with "In the recent chemistry version (v3.3) of the Weather Research and Forecasting (WRF-Chem) model, the Morrison double-moment microphysics scheme is newly coupled with interactive aerosols." "

5. In the Lin scheme the autoconversion of rain has a different parameterization if droplet concentration is prescribed than if it is predicted. In the Morrison scheme implementation are there any differences in the cloud physics based on whether Nd is constant or varying? This doesn't need to be explained in the paper as there is likely not. But if there is, it is important when comparing the 'AERO' to the 'MET' simulations to explain all possible differences between the two.

Response 3.5.0: In the Morrison scheme, the autoconversion parameterization is the same for both cases, whether N_d is constant or varying. In the Lin scheme, the original autoconversion parameterization was independent of droplet number, so a second parameterization was introduced for simulations with predicted droplet number. This second autoconversion parameterization in the Lin scheme could in fact be used with a prescribed droplet number.

6. Most of the figures comparing VOCALS data to model output model use one platform, either RB or C-130 without explaining a reason for the choices, or incorporating data from the other platforms that could add a significant amount of observations. RB data is biased by date, and different synoptic conditions could impact the boundary layer structure over the REx period. The first paragraph of section 3.1 describes the coastal MBL as being '_2K colder and _2gkg-1 less humid' than the remote MBL, based on Table 3, which only uses ship data. While the qualitative statement may be correct, using data that is recorded on different dates in different places is not representative of the mean

geographic contrast on average. A multi-platform mean utilizing the flight data to fill in some of the missing data would be a better observation to use to make such a quantitative statement. In addition, there is no mention of number of samples or statistical significance on the results, which would greatly enhance the arguments made. Also, aircraft and ship data are used for comparison of different quantities to model simulations without reasons explaining the choice of comparison platform. With aerosol variables (e.g. Table 4 and Figure 4) comparisons with several platforms are made and it is clear 'observations' refers to any of the platforms, but the discussion of cloud variables needs to be modified to make the reader understand why RB is used in some cases and C-130 in others.

Response 3.6.0: These are very good points. The selection of the platforms has also been questioned by the referee #1 and we have addressed these questions. Please see Response 1.2.0 for details. Regarding the RB sampling time issue and the humidity contrast, they have also been addressed earlier in the response to referee #1. Please see Response 1.4.0 for details. As suggested by the referee, we have included the number of profiles used for the statistics in Figs. 1 and 2 and Table 3. In the discussion related to Table 4 and Fig. 4, we used 'observations' to refer to all the observations from different platforms. We have revised the related discussion to clarify. Regarding the change of cloud variable observations from the C-130 to RB, please see our response to the specific comment c below (Response c).

Specific Comments

a. Table 3: It is unclear why the SST is in units of Celsius, while the temperature is in units of Kelvin. If one wants to compare the two, they need to do arithmetic. I suggest changing the SST values to SI units.

Response a: Changed as suggested.

b. The Table 3 caption should explain at what vertical level dqv/dh and dthetav/dh refer to. Based on the text, I assume this is across the inversion level (section 3.1), but the numbers cited are not the same as the table, as they are not divided by dh. Does it refer instead to the average dqv/dh over the MBL?

Response b: Thanks for pointing this out. The $d\theta_v/dh$ and dq_v/dh in Table 3 were calculated for the inversion layer. We have added the explanation into the footnotes of Table 3 to clarify.

c. Table 3: Why are droplet concentration observations from RB, while aerosol concentrations are from C-130 data? In Figure 3 both aerosol and droplet concentrations are from C-130 observations, but Section 3.2.1 describes Nd from aircraft '(Nd, Fig.3 and Table 3)'. It also describes Nd from RB to compare with Bretherton et al. 2010 values (from MODIS and aircraft). This is confusing and it is not clear why the discussion uses different observation platforms to make different points. First, '(Nd, Fig.3 and Table 3)' should be changed to '(Nd, Fig.3)' because Table 3 doesn't show aircraft Nd. The reason for switching the discussion from aircraft data to 'near surface' Nd from the RB needs to be explained. Perhaps Table 3 could include comparisons of model Nd with both RB and C-130.

Response c: The observed droplet concentrations in this study are from the C-130 platform only, being consistent with the observed aerosol number concentrations. The association of 'RB' with N_d observations in Table 3 and one location in the main text was a mistake; they really are N_d from the C-130. They have been corrected.

d. Page 22675 Line 28-29: It seems likely that the larger variability over the remote region is due to open/closed cellular dynamics, but other factors could influence this: Other good possibilities are mentioned, but what about the decreased subsidence away from the Hadley cell (allowing for deepening of the boundary layer), or the distribution/sparseness of available data? The paper should mention the number of observations used for the 'coast' and 'remote' averages.

Response d: We agree that decreased subsidence away from the Hadley circulation allows for the deepening of the boundary layer and is likely to induce larger variability of the MBL temperature and humidity. However, the related sentence regarding variability has been deleted in the revised manuscript due to the change in results after processing a newer version of the RB radiosonde data (see Response e below). The radiosonde observations on the RB have 23 and 54 observations in the remote and coastal regions, respectively. The numbers of observations have been included in Table 3 and Fig 2.

e. Page 22675, lines 1-3. While it is true mixing data from multiple sources can obscure a real signal only observed from one of the sensors, it also possible that the humidity contrast noted in RB data is artificially created by the sampling distribution in time and space. Also, even with systematic differences between flight and ship sensors, both flight and ship should observe some magnitude of zonal humidity contrast if it exists and is detectable. One would need to look at the flight and ship observations separately, and combined to determine if the humidity contrast is indeed obscured in the Bretherton et al. 2010 paper. Is the humidity contrast statistically significant based on the degrees of freedom from RB observations?

Response e: The humidity contrast has also been observed by the C-130 according to Jones et al. (2011). However, after the reprocessing using a newer version of RB radiosonde data (it just came to our attention during the revision process.) and using an additional latitude range constraint, the humidity and temperature contrast is no longer statistically significant. Please see response 1.4.0 on Page 7 for more details.

g. Page 22677, Line 4-5: Is the larger variation in MBL depth during the daytime something noticed in model output, observations, or both?

Response g: Thanks for pointing this out. More detailed statistical analyses suggest the conclusion is not statistically robust. We have deleted the related statement.

h. Page 22680, Line 10: The text says ". . . treating sea salt as NaCl in model implies an overestimation of sodium and chloride emissions by 25% and 10%." What is the source of this information? After accounting for this effect, there is still an overestimation by factor of 1.9. Does this number apply to both sodium and chloride separately by the same amount?

Response h: The composition of freshly-emitted sea-salt particles reflects that of sea water, and the inorganic portion is about 55% chloride, 31% sodium, and 14% other (SO4, Mg, Ca, K, …) by mass (e.g., http://www.marinebio.net/marinescience/02ocean/swcomposition.htm, http://www.soest.hawaii.edu/oceanography/courses/OCN623/Spring2011/salinity.pdf). The model treats sea-salt aerosol as NaCl (61% chloride and 39% sodium by mass), so when 1 kg of sea-salt aerosol is emitted, the sodium and chloride are overestimated by the factors given in the text. The factor of 1.9 applies to both sodium and chloride. The sentence has been rephrased to clarify.

i. Page 22681, Line 18-20: While this may be indicative of the first aerosol indirect effect, strictly speaking the first AIE is described in conditions of constant CWP. While not obvious from the color scale used in Figure 6, there is a longitudinal gradient of CWP in the region that complicates attribution of the first AIE. I suggest removing the mention of the AIE as it is not discussed further here.

Response i: The suggestion has been taken.

Technical Comments

i. Figure 1 caption: include '(red)' after 'RB ship' to be consistent

Response: It has been changed as suggested.

ii. Page 22684, Line 3: The word 'identical' is too strong of a word for this situation.

Response: The word 'identical' has been replaced with 'close'.

iii. Page 22684, Line 9: Refer to Fig. 5 after "AOD"

Response: It has been changed as suggested.

iv. Page 22684, Line 10: Break this into two sentences, e.g ". . .biases in TOA SW. For example, . . ."

Response: It has been changed as suggested.

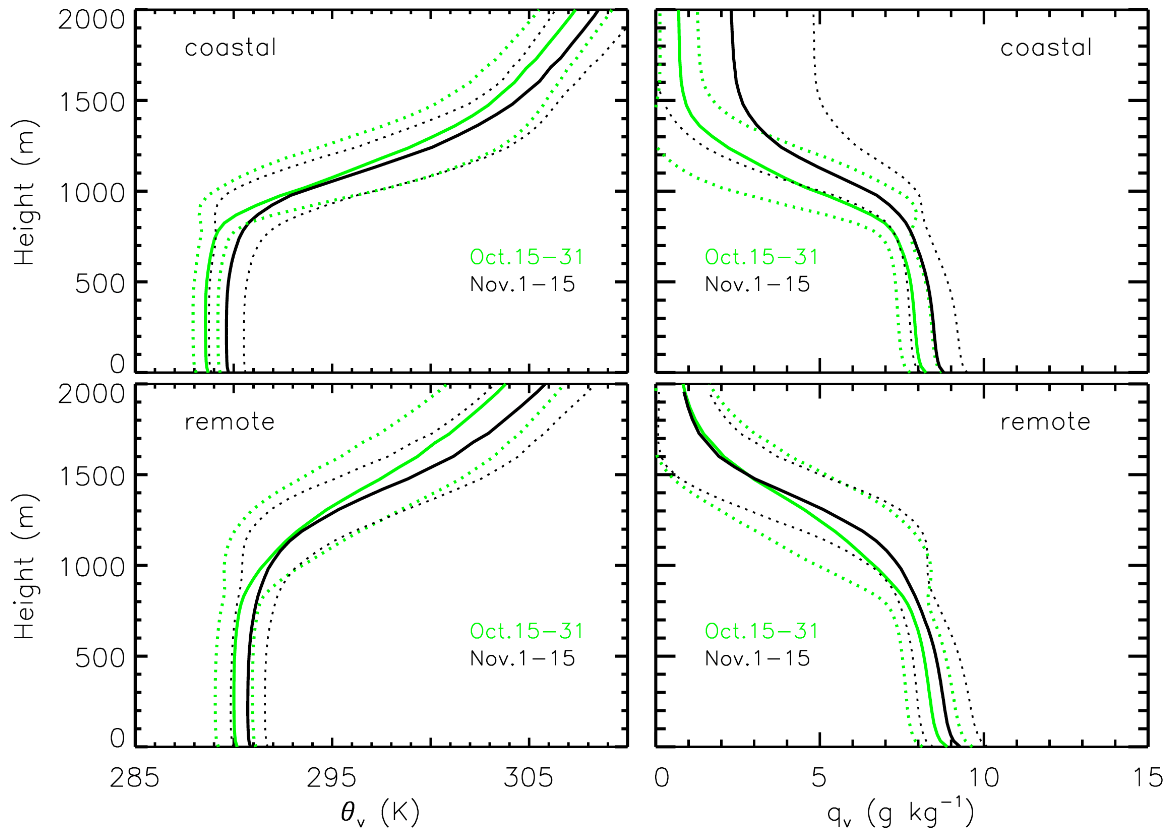

Fig. R1. The average vertical profiles of virtual potential temperature (θ_v) and water vapor mixing ratio (q_v) along 20°S for 15-31 October (green) and 1-15 November (black), 2008, respectively, based on the AERO simulation. The dash blue lines also indicate the ± 1σ.

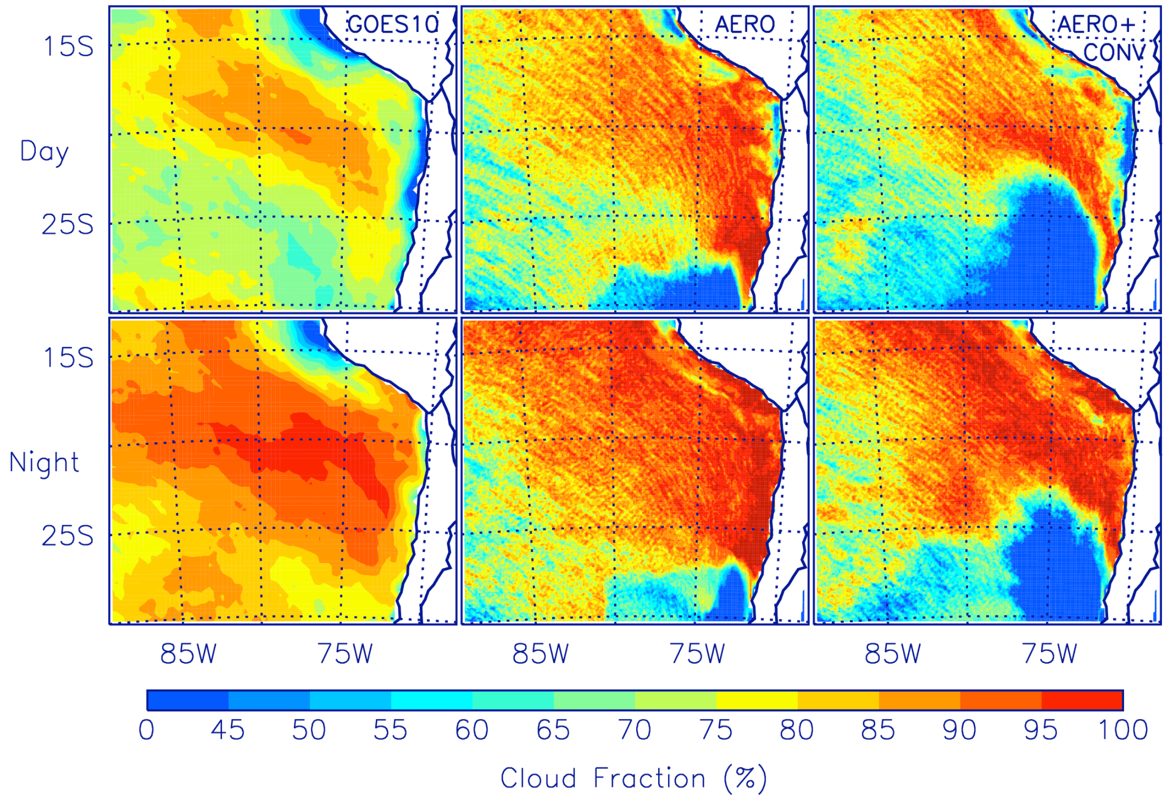

Fig. R2: Mean low cloud fraction during day and night for October 15-18, 2008 retrieved from GOES–10 (left), and those from the AERO simulations with no cumulus convection schemes (middle), and with the Kain-Fritsch scheme (right).

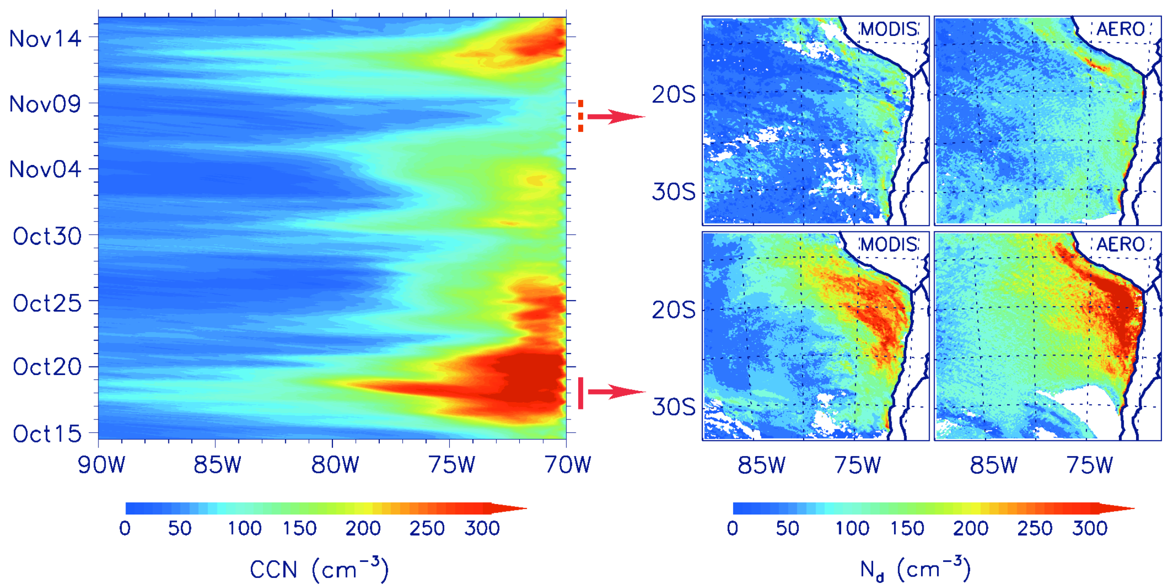

Fig. R3. Longitude-time plot of model (AERO) predicted CCN (at 0.1% supersaturation) concentrations at 975 hPa along 20°S (±2.5° in latitude), and illustration of episodic horizontal distribution of MODIS-derived (Aqua) cloud droplet number concentration (N_d) and AERO-predicted cloud top N_d during a strong outflow event (peaks on Oct 18, 2011, red solid line) and during a clean period (around Nov 8, 2011, red dash line). To obtain a more complete data coverage over the domain, MODIS N_d was composed from available retrievals in 3 days (centered at the peak of the event), and correspondingly only model predictions at around satellite overpass time (18-20 UTC) during the 3-day period are included.

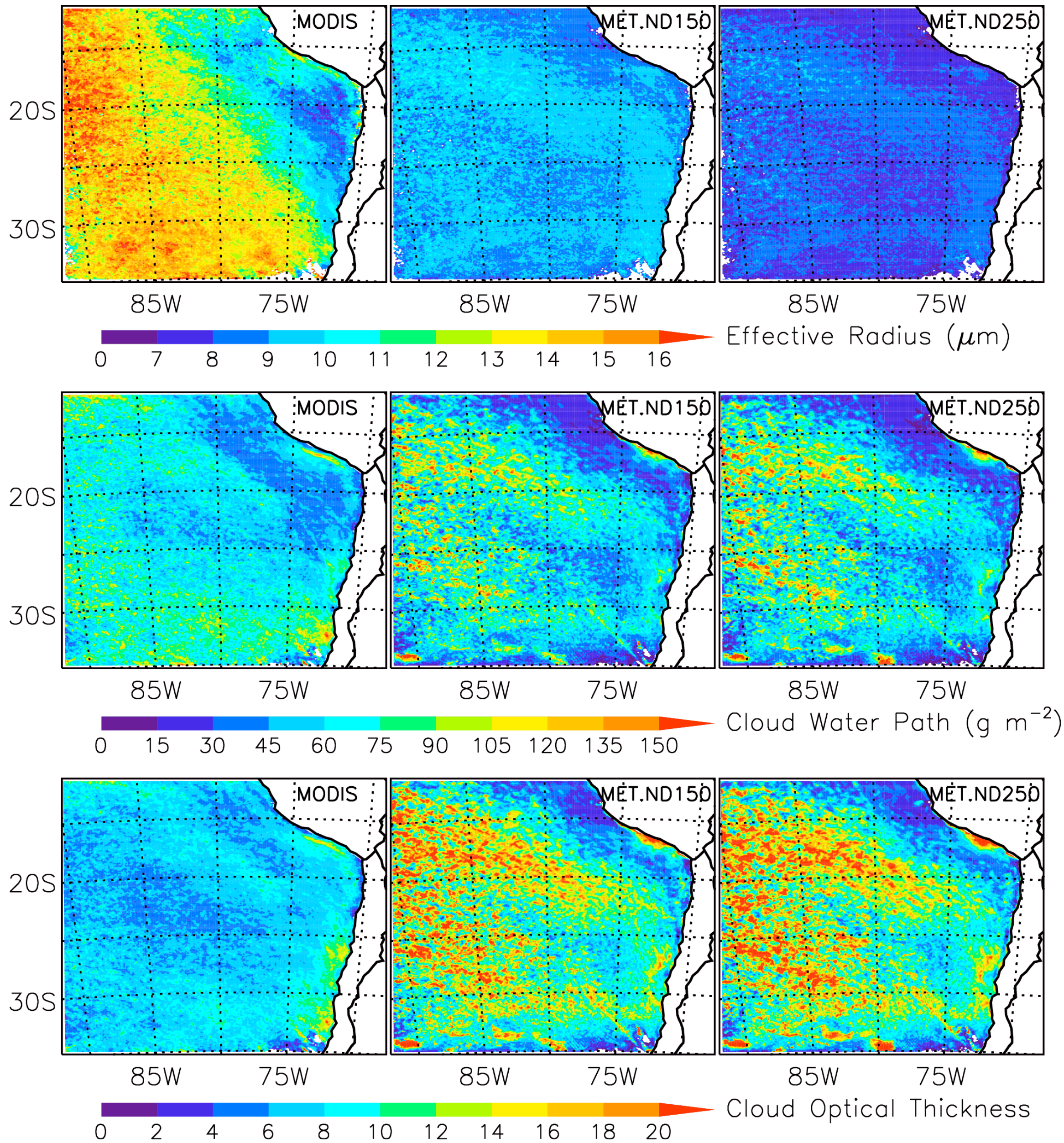

Fig. R4: Effective radius (r_e), cloud water path (CWP) and cloud optical thickness (COT) during the VOCALS-REx period from MODIS (Aqua) retrievals and from the MET simulations with constant droplet number concentrations of 150 cm^{-3} (middle) and 250 cm^{-3} (right), respectively.