We thank both of the reviewers for their constructive comments which have led to an improved revision of the manuscript. We address each of the points raised below. Note that the reviewer comments are shown in italic.

An additional modification is that Fig. 2 has been updated with the latest version of R. H. Brown sounding data. This removes some data artefacts that have been identified in previous datasets. The new dataset also contains 7 additional soundings at 75 W that are now also included in the mean atmospheric profiles shown in the submitted version. The results from this new dataset do not affect the discussion in the paper.

Reviewer 1

General comments

This paper very nicely documents how the Met Office UM represents the Southeast Pacific Stratocumulus by comparing model forecasts with VOCALS observations. Overall, the paper is well written and properly structured. It contains very relevant results in the sense that it is important to document in detail what the main modeling issues are, as the evaluation of models deficiencies is a necessary benchmark for future model improvements). This paper should definitely be published after some minor revisions.

My main concern is that while the authors clearly identify some major model strengths and biases in the forecasts (for instance, the remote maritime clouds are well captured with the MetUM- both the diurnal cycle and synoptically induced variability- but the coastal cloud is poorly represented) they don't thoroughly examine the possible causes for these biases (Is high resolution going to solve the coastal cloud problem?). This paper could substantially improve if the authors provided a more detailed interpretation of their results. Another issue is that the authors only show a few diagnostics of the boundary layer and cloud but leave out numerous others that could be relevant (for instance: TKE profile, entrainment velocity, surface fluxes…). Also, I think the paper would benefit from a more detailed description of the relevant parameterizations.

Specific comments

P16801: My feeling is that the model description could be a little more detailed. I would love to have a brief description of the parameterizations of interest (the PBL scheme in particular).

We have added a more detailed description of the boundary layer, cloud and microphysics schemes in section 2.1 and a new appendix.

P16803: Authors claim that the model PBL height is about 200 m too low compared to observations. Is it accurate? For instance, from looking at a few points (at 80W-10S: Satellite ~ 100m and Model ~ 700m. At 88W-30S: Satellite > 1600m and Model ~ 1100m), it might be more than 200 m.

As mentioned in the text this is a cursory comparison with a satellite retrieval shown in Zuidema et al. (2009). It should be noted that the relevant plot from Zuidema et al. (2009) is from a slightly different time period (Oct 2008 mean) to the model forecast (14th Oct – 19th Nov 2008). The reviewer points out that at certain locations the difference between these observations and the model can be more than 200 m. This is certainly true and whilst the different time periods may account for some of this (it is shown in section 3.2 that synoptic variability can have a strong impact for example), the important point here is that the general spatial pattern of boundary layer height is similar e.g. an increase away from the coast, higher boundary layer heights in the Arica Bight compared to other coastal locations etc. The relevant text has therefore been revised with the reference to 200 m removed at this stage in the paper. The model boundary layer height bias is however evaluated in more detail against in-situ observations from VOCALS-REx later in the paper in sections 3.1 and 3.2. Here it is shown that the model typically has a low bias of ~ 200 m (mode of the distribution in Fig 4). However it can be seen from Fig 4 of the paper for example that this bias can be larger, which may also be consistent with the larger differences noted between the satellite and model at certain locations by the reviewer.

P16805: Indeed the year 2008 was atypical in terms of PBL depth. The 2008 observed cloud-top heights were higher than in previous campaigns (except 2006 that also showed deeper PBL than usual). Do you have a sense how the MET Unified Model would perform for more typical years?

We have not analysed model forecast data from other years as this is beyond the scope of this paper. We do however modify the text to emphasise that there is also inter-annual variability in cloud top height.

P16805: The authors suggest that the inability of the MetUM to represent the boundary layer structure may be related to a poor representation of the Andes topography. On this issue, Richter and Mechoso (2006) suggested that employing a more realistic mean orography is not likely to improve the simulation of stratocumulus.

We have rewritten parts of the text in section 3.1. The emphasis should be that in order to better capture the near-coastal circulations and the coastal boundary layer structure, higher resolution simulations may be required (not just changing the representation (smoothing) of the topography on the same horizontal grid as might be inferred from the current text and which is studied by Richter and Mechoso, 2006). This has recently been illustrated by a paper submitted to the same VOCALS special issue by Wang et al. (2010) which we now refer to, who show that increasing the horizontal and vertical resolution can lead to significant improvements in the boundary layer structure close to the coast. They attribute this improvement to a better representation of the mesoscale coastal circulations in their model simulations. We are planning on doing further studies in the near future to look at the impact on the near-coastal boundary layer structure and cloud with MetUM simulations at a higher resolution than the simulations analysed in the current paper.

P16807: It is not clear to me whether observations and forecasts are averaged over the same time period (in Figure 3). Do the flight time periods span a complete daily cycle?

In Fig 3 of the paper the model data are daily averaged values whereas the C-130 aircraft data do not correspond to a complete diurnal cycle (a typical flight would last ~ 9 hours). Measurements in the remote maritime region (80-85W) are weighted toward early morning whereas coastal measurements include early morning and mid-afternoon observations. We choose to retain the model daily averaged values for the individual flight days in Fig 3 of the paper though (red and blue lines) as they can be compared to the model statistics over the complete simulation period, which are shown with the grey shading. This helps to illustrate that the observations on the 23rd October for example are atypical for the VOCALS-REx period. However, when constructing the histograms to better quantify the model bias in cloud top height as shown in Fig 4 of the paper, the nearest hour and model grid box to each individual observation is used. This more quantitative comparison therefore addresses any sampling bias that might result from sampling different parts of the diurnal cycle. We clarify in the text that the model data in Fig 3 is daily averaged data and that in Fig 4 it is from the nearest model time-point to each observation.

P16811: The authors clearly describe the problems in the diurnal cycle but don't indicate the causes of the biases. Any clues about the origins of the biases?

The largest model bias in the diurnal cycle of cloud is clearly in the near coastal region. The model is unable to capture the coastal clearing that is often seen in the satellite observations. It is shown in Fig. 2 of the paper that the coastal marine boundary layer is too shallow and moist in the model simulations. The enhanced moisture in the model prohibits the break-up of the stratocumulus during the daytime as diagnosed by the model cloud scheme, which assumes a fixed triangular distribution of moisture around the grid-box mean value (Smith, 1990). When the grid-box mean relative humidity exceeds a critical threshold value, RH_{crit}, then large-scale cloud is diagnosed in the model grid-box. For the MetUM simulations the value of RH_{crit} changes as a function of height and is set to 91% at the surface and decreases to 80% at 1 km altitude and higher. In the model the typical marine boundary layer height at Iquique is ~ 500 m which corresponds to a RH_{crit} value of 86%. In the figure below we plot the maximum boundary layer RH from all of the soundings at the coastal location of Iquique in red and from the corresponding model data in black. The appropriate RH_{crit} value of 86% is marked with a dashed line. It is evident that the model relative humidity is typically higher than in the observations (which can also be inferred from Fig. 2 of the paper). It is also clear that the model peak boundary layer RH is typically above the RH_{crit} value of 86%, and so the cloud scheme will nearly always diagnose some large-scale cloud. This leads to the model being unable to break-up the cloud layer at the coast sufficiently during the day time.

It is also clear from the figure below that there is a marked diurnal cycle in the observed peak boundary layer RH, with a drying out of the boundary layer in the day time which will likely contribute to the break-up of the cloud layer. We note that this diurnal cycle in RH is much weaker in the model simulations. As mentioned above, improving the representation of the mesoscale coastal circulations (which have a strong diurnal component) around the steep topography in the model may be required to improve the forecast of the boundary layer structure and cloud at the coast.

We add a discussion of the above to section 3.3 of the revised version of the paper.

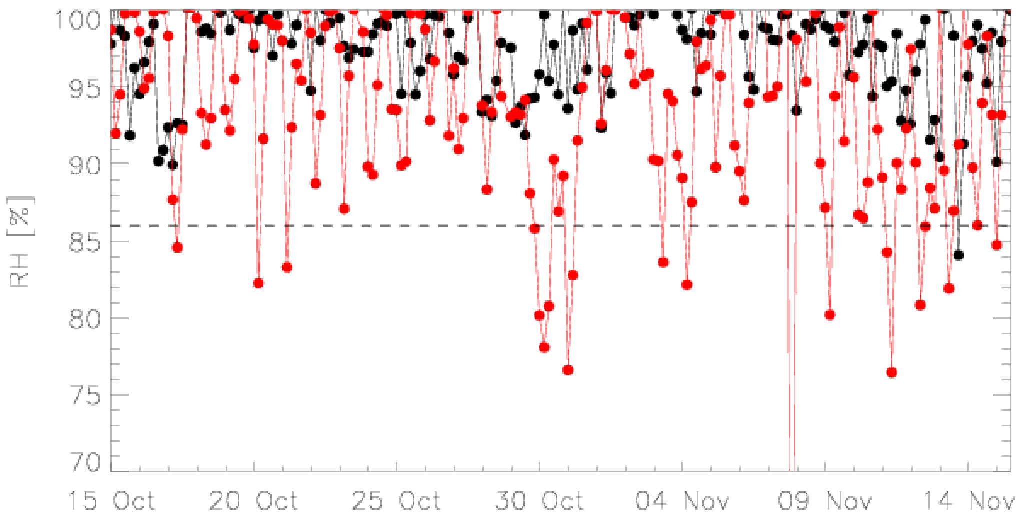

Figure 1: Peak in boundary layer RH at the coastal site of Iquique. Observations are shown in red and the model simulations shown in black. The dashed line is the value of RH_{crit} used in the model cloud scheme to diagnose large-scale cloud.

P16813: Main focus is on diurnal cycle of cloud fraction, LWP, and drizzle. Did the authors look at other variables (latent and sensible heat fluxes, vertical velocities, LW and SW fluxes at the surface etc...)? It would be also interesting to look at vertical distributions of cloud liquid and cloud fraction (maybe comparing with Cloud Sat data).

The main focus of this study was to document the operational NWP models general ability to forecast the synoptic variability and diurnal cycle of stratocumulus cloud in the southeast Pacific, and to highlight areas where significant biases exist. We do plan to look in more detail at the boundary layer and cloud structure and it's evolution in future case studies that will combine additional remote sensing data with multi-platform observations e.g. aircraft operating in conjunction with the Ronald H Brown ship. These case studies will also be utilised for model sensitivity studies to better understand some of the model biases in more detail.

P16816: The model uses a fixed Nd of 100cm-3 over the ocean. While this value is accurate for remote maritime areas, it is too small for polluted coastal areas. What is the impact on the cloud albedo and SW at the surface?

Again we choose to evaluate this in latter studies as part of new model simulations that include the transport of aerosols and a representation of aerosol-cloud interactions. The aerosol-cloud interactions will impact both drizzle production and the cloud albedo in the model and are being performed for the upcoming VOCA modelling experiment (see http://www.atmos.washington.edu/~mwyant/vocals/model/VOCA_Model_Spec.htm for further details).

Technical corrections
- P16800, L21: model's ability (instead of models ability) Done
- P16804, L25: remove one "the" from "with the the inversion" Done
- In the text, should you capitalize South East (instead of south east). The relevant text has been changed to southeast to be consistent throughout the paper.

Reviewer 2

This is a good manuscript describing well an evaluation study of the UK Met Office model performance during the VOCALS experiment regarding marine boundary layer clouds. This paper should be published after some major revisions.

Specific Issues:

a) The explanations of the physics of stratocumulus-topped boundary layers, namely in the introduction, are somewhat confusing and it would be great if the authors could improve the text in this respect.

The relevant text has been updated.

b) The authors need to more clearly describe the parameterizations that are being evaluated. How is the liquid water calculated? What terms are in the equation? How is the cloudy boundary layer scheme coupled to the cloud and microphysics parameterization? Please write the relevant equations.

We have added a more detailed description of the boundary layer and cloud scheme to the model description section. We have however refrained from adding a significant number of equations to the text as these have been documented in previous publications which we reference. However there are some modifications to the version of the microphysics that are used in the model (as compared to previous publications) and these are described in a new appendix – see also the response to point i) below.

c) Please explain in some detail how the model boundary layer height is computed. What is the impact in terms of the analysis of the results (fig.3 and 4) of using different definitions (in models an observations)?

We have added more information on the calculation of the cloud top height to section 3.2. We choose to use this diagnostic which is calculated by the boundary layer scheme for comparison with the observed cloud top height. The diagnostic is a sub-grid calculation and represents the height at which turbulent mixing from cloud top occurs in the model. The figure below is equivalent to figure 4 in the paper but the model cloud top height is calculated solely from the model profiles of cloud fraction (which are on a discrete vertical grid). Here cloud top is defined as the top of the highest model layer where cloud fraction > 0. We do not include this figure in the revised version of the paper but it illustrates that the typical low bias of ~ 200 m in the model is still apparent when using other definitions of cloud top i.e. this does not significantly impact the discussion in the paper.

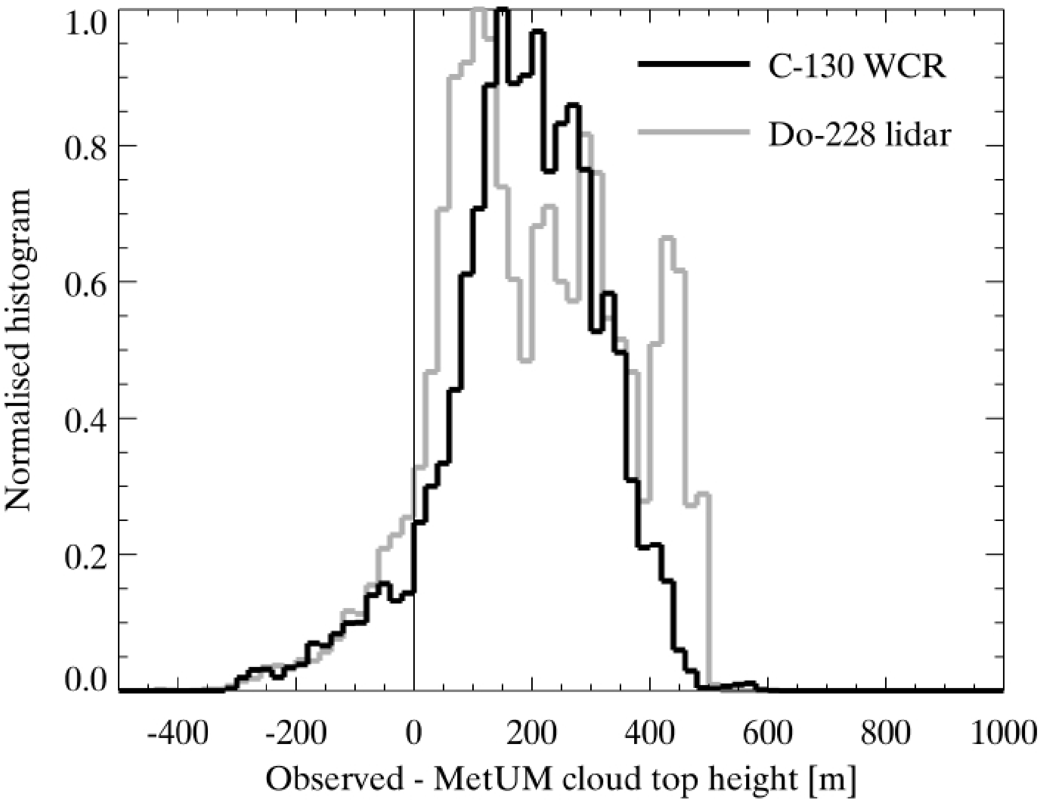

Figure 2: Histogram of the difference between aircraft observations and the model forecast of cloud top height which is defined as the top of highest model layer with cloud fraction > 0.

d) The authors need to provide more physical interpretations of their results. The fact that the authors of this particular manuscript work in the same organization as the researchers that developed these parameterizations is a key advantage that the authors should be able to take advantage of.

We add a more detailed discussion on the model bias in cloud cover at the coast. We believe that this results from the high bias in the boundary layer moisture shown in figure 2 of the paper, which prohibits the break-up of the cloud layer during the day time (see response to reviewer 1 for further details). We also propose further modifications to the microphysics scheme in sections 3.4 and 3.5, that may improve the drizzle efficiency in the model (changing the autoconversion scheme) and to increase the evaporation of drizzle below cloud base (change the raindrop fallspeed relation). Evidence that the model underestimates the evaporation of drizzle has arisen from observational results shown in a recent paper by Bretherton et al. (2010). These observations show that a larger fraction of drizzle evaporates in the sub-cloud boundary layer than in the model simulations. These modifications to the microphysics will be tested in future sensitivity studies.

e) Please state where the boundary layer model levels are in height (since the authors do discuss the role of vertical resolution).

In section 2.1 we have modified the text to "The vertical grid spacing is 20 m at the surface and increases with altitude, such that there are 10 model levels below 2 km (see filled circles in Fig. 2)."

f) The authors use in eq.(10) Zmbl from the model. Could this be an issue? Would it be cleaner to compare brightness temperatures from the model and the observations?

We deliberately chose to derive a cloud cover product from GOES-10 as models are typically evaluated against satellite retrievals of cloud fraction (note that the cloud cover product will be made available for other researchers). The low cloud cover diagnostic that we compare from the model represents the horizontal proportion of the grid-box that is cloudy when looking down from above i.e. as seen from a satellite. The reviewer suggests that comparing brightness temperatures from the model directly would be a cleaner comparison. One could perhaps look to estimate a cloud top temperature from the model profile of cloud fraction for example although difficulties would arise when the grid-box mean cloud fraction is < 1.0 and the SST or lower clouds contribute to the brightness temperature measured by the satellite. To do this properly would involve outputting the brightness temperature (or GOES-10 channel 4 radiances) directly from the models radiation scheme. We do not have these diagnostics available for these simulations.

The reviewer also asks if using the model Zmbl in equation 1 to calculate the temperature threshold function has adverse affects on the derived cloud cover product. We choose to use this as it allows for variation in both SST and boundary layer depth to be utilised in the retrieval. As shown in figure 5 of the paper the derived cloud cover product is comparable to that from MODIS, yet available at much higher temporal resolution. It can also be seen that there is good correspondence between both the cloud cover product (Fig 5b) and GOES-10 imagery (Fig 5a) along the 20S latitude line, and with previous observations at the Stratus buoy (Fig 8a). Further visual examination of the derived cloud cover product with GOES-10 imagery across the wider VOCALS-REx region (not shown) generally supports the choice of the threshold function.

g) First lines of page 16812: is this the main mechanism by which solar radiation modulates the boundary layer structure? Please provide references.

We have removed the sentence from this part of the text and have chosen to revise the discussion on the physics of stratocumulus in the introduction as suggested in point a).

h) Please be precise about the Aqua satellite overpass times.

The Aqua satellite (AMSR-E data) is in a sun-synchronous orbit and the exact overpass times (UTC) differ depending on the location on the Earth's surface. The overpass times in the remote maritime and coastal regions can be seen in Fig 8 c-d of the paper. You can notice a slight difference in the times in the two regions but they are ~ 07:00 and 19:00 UTC i.e. very similar to the time used for the diurnal maps of GOES-10 cloud cover in Fig. 6 of the paper. We have added these Aqua overpass times to paragraph 3 of section 3.4.

i) Section 3.5: Please describe in here, if not before, and in some detail how drizzle is parameterized in the model.

The microphysics scheme used is an updated version of that described by Wilson and Ballard (1999). We have added an appendix to outline the modifications that are relevant to warm clouds such as those studied in this paper. We also discuss potential improvements to the microphysics scheme in Sect. 3.4 and 3.5 that will be tested in future modelling work.

References

Bretherton, C. S., Wood, R., George, R. C., Leon, D., Allen, G., and Zheng, X.: Southeast Pacific stratocumulus clouds, precipitation and boundary layer structure sampled along 20 S during VOCALS-REx, Atmos. Chem. Phys. Discuss, 10, 15921–15962, 2010.

Richter, I., and Mechoso, C. R.: Orographic influences on subtropical stratocumulus, J. Atmos. Sci., 63, 2585-2601, 2006.

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

Wang, S., O'Neill, L. W., Jiang, Q., de Szoeke, S. P., Hong, X., Jin, H., Thompson, W. T., and Zheng, X.: A regional real-time forecast of marine boundary layers during VOCALS-REx, Atmos. Chem. Phys. Discuss., 10, 18419–18466, 2010.

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

Zuidema, P., Painemal, D., de Szoeke, S., and Fairall, C.: Stratocumulus cloud-top height estimates and their climatic implications, J. Climate, 22, 4652–4666, 2009.