Responses to Referee #1

Wyant et al. 2014, Global and Regional modeling…: the VOCA Intercomparison, ACPD

Thank you for your helpful comments.

This intercomparison contributed to the modeling plan that motivated VOCALS. While the study was professionally done, I was not sure what to take away from it. Comparisons are made and discussed, but little interpretation on what aspects of the individual model behavior limit aerosol-cloud interaction representations is provided, limited perhaps by the diversity of models evaluated.

The conclusion has been expanded to try and clarify the meaning of the study. Also an additional figure comparing mean N_d and CCN concentrations has been added to better illustrate one aspect of the model's capacity to simulate aerosol-cloud interactions. Intercomparison studies are by necessity better for pointing out general problem areas than the reasons for biases in particular models. This still serves the purpose both of informing the scientific community about the current state of modeling as illuminated by the intercomparison, and of inspiring developers to analyze individual models in more depth to provide the basis for model improvements.

If the above goal is too ambitious, could the authors instead provide the community with some guidance on where it should go with such comparisons?

What advise do the authors have for the modeling plans of future campaigns? Would such intercomparison exercises be better served by focusing on models of one type, such a sdone within GCSS, allowing for a further drilling down to specific process representations, e.g., aerosol wet deposition? This would still allow for one study per model type, and provide more of an apples-to-apples comparison.

Some suggestions and advice for a follow-on intercomparison study have been added to the conclusion. For some types of future studies, limiting the types of models participating would indeed allow more specific processes to be studied and problems to be addressed.

Is there a way to rank models, or can the authors suggest useful metrics [vertical resolution?]?

Our main goal for the intercomparison is to survey a broad range of models, see where the models are capable and deficient, have individual modeling groups learn about their model biases, and stimulate model improvement. If the class of models was limited and the input conditions were more stringent, ranking models could be useful. However, a ranked study would probably discourage participation.

On p. 25, we do see a paragraph that compares the performance of Pre-VOCA models to VOCA. Can we conclude from this that incorporating cloud-aerosol interactions does not improve the representation of cloud fraction? Some more in-depth discussion here would be a useful contribution of this manuscript.

This study is meant to complement Pre-VOCA by comparing those models with chemical transport and aerosol prediction capabilities with a comprehensive in-situ dataset from VOCALS-REx that allows far more aerosol, chemistry and cloud comparisons. As significant model and observational details have changed, and the model pool has changed, the results are not meant to be compared with Pre-VOCA and the conclusions to be reached by comparing the two studies are necessarily limited. We have added some text in the conclusions about why we do not think such a comparison would be particularly fruitful: "For many models in VOCA, the representation of aerosol processes is a relatively new feature, and at this stage of model development, we do not expect nor generally observe that their inclusion necessarily improves model performance relative to Pre-VOCA."

I also did not see much discussion of the Ron Brown datasets, despite those comprehensive aerosol and cloud datasets.

The cloud-top height measurements in Figure 5 were from the Ron Brown, and this has now been noted to the caption. We have also added Ron Brown data to Figures 3 (LWP), 4 (cloud-top), and 8 (DMS), and included discussion in the text where appropriate.

Abstract: Many of the statements within the abstract are vague (e.g, "some models simulate the regional low cloud cover well"...."Most models qualitatively simulate...."). Can this [and other statements] be sharpened? Some of the discussion section could probably be summarized in the abstract to add some interest here.

It is difficult to sharpen many of these statements in the abstract without specifying individual models. Even that would be qualified (e.g. no single model has an excellent match for mean cloud fraction). However, a part of the abstract has been enlarged to incorporate a few more of the points raised in the conclusion:

"Most models qualitatively simulate the observed offshore gradients of SO2, sulfate aerosol, CCN concentration in the MBL as well as differences in concentration between the MBL and the free troposphere. Most models also qualitatively capture the decrease in cloud droplet number away from the coast. However, there are large quantitative inter-model differences in both means and gradients of these quantities. Many models are able to represent episodic offshore increases in cloud droplet number and aerosol concentrations associated with periods of offshore flow."

Figures:
fig. 3: worth mentioning AMSR-E is day+night in caption.

Added to caption.

fig. 4: a more general satellite-derived cloud top height would be useful here

As noted above, cloud-top observations from the Ron Brown were added. The uncertainties in satellite cloud-top height are great enough that we do not feel they would substantially improve the figure.

fig. 8, p 17: see also, for discussion of Ron Brown sulfur/DMS results, M. Yang and 20 co-authors, 2011: Atmospheric sulfur cycling in the southeastern Pacific - longitudinal distribution, vertical profile, and diel variability observed during VOCALS-REx. Atmos. Chem. Phys., 11, pp. 5079-5097. doi:10.5194/acp-11/5079-2011 for a discussion of the ship observations.

DMS observations from this study are now also plotted in Figure 8 and this study is cited in the text.

fig. 10: spell out SSA

done

fig. 11, discussed on p. 19: a consistent approach to estimating MODIS Nd from space is not yet in the literature (e.g. doi:10.1029/2011JD016155 and likely others). Would recommend including the equation used.

While there are numerous MODIS retrivals for N_d in the literature, the one we used here (Bennartz (2007) and George and Wood (2010)) agrees well with the aircraft observations in Bretherton et al. (2010). Rather than including the equation used, we have opted to add a reference to George and Wood (2010).

4, 6th line "spring" => "austral spring"

Unsure of the location the comment is referring to. We cannot locate a 'spring' that is not already 'austral spring'

References:

Bennartz, R: Global assessment of marine boundary layer cloud droplet number concentration from satellite.,J.Geophys. Res., 112, D02201, doi:10.1029/2006JD007547, 2007.

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 20º S during VOCALS-Rex, Atmos. Chem. Phys., 10, 10639-10654. doi:10.5194/acp-10-10639-2010, 2010.

George, R. C., and Wood, R.: Subseasonal variability of low cloud radiative properties over the southeast Pacific Ocean, Atmos. Chem. Phys., 10, 4047-4063, doi:10.5194/acp-10-4047-2010, 2010.