We thank the reviewers for their helpful comments and suggestions. Individual questions and issues are addressed in-line below.

Anonymous Referee #1

General Comments:

1) It is stated that the initial conditions are chosen such that the BL characteristics are in adequate agreement with the observation available 8h into the simulation. Here I would like to be given more detail. In particular it would be interesting to know how strongly the BL characteristics and the roll structures are influenced regarding slight changes in the initial conditions, and how strongly such changes impact your ship track. For instance, as discussed in the paper, the roll structure strongly confines the track if the emissions are set along the roll axes. In particular could the low bias in track width and the smaller simulated roll width be related, and perhaps contribute to the high bias of Nd within the track? Considering the length of the paper, one could consider to present a discussion of the sensitivity to the initial conditions in an appendix.

We agree that the boundary layer is sensitive to the details of the initial conditions, and indeed a series of pilot runs were needed to adjust the geostrophic wind profile and subsidence to achieve a circulation with well-developed roll structure and reasonable agreement with the observed profiles. While the nuanced response of the roll structure to the forcing is an interesting boundary layer dynamics question, the thrust of the paper is to address the aerosol-cloud-precipitation interactions in a case study anchored with aircraft observations, and an extended analysis of the dynamical sensitivities of the initialization and forcing seems a bit outside the scope. The work of Chlond (1992) and Etling and Brown (1993) and references therein provide much more depth on the sensitivities of the roll structure. The LES study of Glendening (1996) provides some insight into the dynamical sensitivities of roll form boundary layers, and has been added as a reference for the interested reader.

2) The aerosol background concentrations and hence background Nd are continuously decreasing with time, as the only source of aerosol, the surface flux, decreases with time. Can you make any statements about the realism of this feature and what causes it? I presume it is linked to the roll circulation, which also reduces in strength as the simulation continues. Could there be issues here, as you consider only one mode of rather large particles? Thereby ignoring any source of particles in the activation range by condensational growth from the Aitken mode as well as new-particle nucleation?

While the Nd value decreases continuously with time after the track is injected at hour 8, it at first increases during the spin up phase when the wind driven surface flux is strong relative to the active removal/dilution processes, which scale with the liquid water path, cloud and rain droplet size spectra, and entrainment rate. At hour 8, when the liquid water profiles and turbulent circulations are most comparable to the aircraft sounding of the background state, aerosol number sink processes most closely balance our parameterized source strength. As our relatively simple model omits nucleation and condensational growth from the Aitken mode, it is possible that this parameterized source strength is unrealistically weak.

Since the aircraft sampled the track for a relatively short period of time, observations do not reasonably constrain the evolution of aerosol number, so it is difficult to say how representative the model behavior is. The results of Mechem et al. (2006) and Berner et al. (2013) suggest that either a large entrainment source (i.e. polluted free troposphere with strong entrainment), which helps to buffer the boundary layer aerosol concentration, or an exceptionally strong surface or nucleation source of aerosol would be needed to balance losses due to collision-coalescence after precipitation strengthens beyond a few mm/day.

3) The discussion of the SensHiAer simulation, although stepping away from the observation, adds scientifically to the paper, as a different ship track behavior is seen in terms of secondary aerosol-cloud interactions. However, the description of the involved processes is not clear yet. Differences in the background state, which behaves differently in terms of LWP and precipitation evolution (Fig. 15), remain largely unexplained. Furthermore, the hypothesized mechanism of the ship track LWP increases one hour after emission is not clear.

Some reorganization of section 5.5 has been done in an effort to clarify the mechanism responsible for the differences in boundary layer organization (suppression of precipitation allowing for a stronger buoyancy flux, more robust rolls, and stronger entrainment), as well as the differing track response (difference in entrainment response dependent on cloud water path and droplet size distribution).

Technical Comments:

1) The symbolism of Na, Nad, Nd is not clearly enough stated at beginning of paper and sometimes mixed up:
- P24424: Nd is used in caption, but Na is used in figure.
- P24395L13: Na used without definition, which follows later in P24399L25.
- In the text Na=Nad+Nd+Nr should always be referred to as total aerosol number concentration and Nad as dry (or maybe unactivated) aerosol number concentration for clarity.

Fixed and clarified as requested.

2) I follow the reasoning for using mg-1 as units of number concentration, but this should be done consistently throughout the paper. If not possible with the observations, I suggest to keep everything in cm-3. Otherwise statements like P24401L8-10 become cumbersome for the reader.

We agree that P24401L8-10 is cumbersome by virtue of the direct comparison between dissimilar units. Throughout the paper, we have made sure like units are always compared. However, we prefer not to convert all the observations to mg^{-1}, as it confuses the issue for anyone referring back to Taylor and Ackerman (1999). In the observational community, it is standard to report Nd and Na in cm^{-3}, while the aerosol modeling community has adopted mg^{-1} for aerosol number; this makes conversions somewhat inevitable when referencing previous work and attempting to maintain readability.

3) P24388L10: change “of cloud” into “of the cloud”.

Changed as requested.

4) P24389L10: Include Christensen and Stephens (2012: Microphysical and macrophysical responses of marine stratocumulus polluted by underlying ships: 2. Impacts of haze on precipitating clouds) in references.

Added as requested.

5) P24393L17: Change “liquid ice static energy” to “liquid static energy”.

Changed as requested.

6) P24397L25 and in equation below: Two different values of S are given. Why?

The first value of S is what we use in our study, and we compare this to the S used by Taylor and Ackerman (1999) in the modeling portion of their paper. We are simply noting that we have chosen a weaker source strength than the previous 1-D version of the case study and believe it to be observationally justified by Hobbs et al. (2000) and Ferek et al. (1998). We have reorganized slightly for clarity.

7) P24398L6-7: Statement unclear.

Rewritten for clarity; we mean that we initially tried the sensitivity test with the same S used in BaseTrack, but the microphysical effects of the perturbation were negligible due to higher background aerosol in the SensHiAer configuration. In order to generate a sufficient microphysical perturbation, we elected to increase the source strength by an order of magnitude.

8) P24398L23: Specify what statistics? Cloud, thermodynamical, dynamical... And change “BaseCtrl, BaseTrack, and SensPerp run suggest” to “BaseCtrl run, and background of BaseTrack and SensPerp”

Altered to explicitly assert the similarity of the thermodynamic, dynamical, and microphysical statistics; changed as requested.

9) P24399L23: Mark cross section in Fig.4. If there is a particular reason for choosing 6.4km as cross section location, it should be stated.

Cross section mark added; 6.4km is just the center of the domain, and any other choice would produce a similar result.

10) P24402L17 & P24403L5ff: Fig. 7 suggests that precipitation increases in the track as soon as LWP is increasing. Only when looking at Fig. 15, which is discussed much later, it becomes obvious that drizzle really remains suppressed with respect to the background as LWP increases.

The color scale perhaps makes it difficult to gauge; while light (<1mm day) surface precipitation does redevelop as LWP increases, the pulses of >4mm day rain present in surrounding rolls are essentially absent until after hour 14. Some clarification has been made in the text to address this.

11) P24401L27: I see precipitation change only after 09:30 in Fig.7, which does not add up with 20min given in text.

As above, the first indication of an effect on rain rates is in the loss of the peak surface rain rates in excess of >4mm day. The maximum in the central roll disappears just before hour 9. The description has been amplified to more specifically note this effect, followed by the cessation of virtually all surface precipitation in the central roll after hour 9.5.

12) P24404L25: Tendency of autoconversion is very small. Can that really be taken as evidence?

Sentence reorganized for clarity as requested.

13) P24405L10: Specify that entrainment in track is 3rd largest term. In background its autoconversion.

Changed as requested.

14) P24407L18: “scaleupdrafts” to “scale updrafts”

Fixed.

15) P24411L15-16: Reformulate. Precipitation is continuously increasing from 9am onwards. It does not start 5h after emission.

Changed to “..reverses five hours after the aerosol injection as increasing precipitation intensity alters the microphysics and entrainment dynamics.”

16) P24416L0-1: Although I agree with your message, I would refrain from speaking about convergence, as the simulations are too short to see real convergence of the curves in Fig.15. For instance panels a1,a3 and b3 of Fig. 15 are clearly not converging in this time period.

“Converging” replaced with “tending” and “convergence” replaced with “approach”.

17) P24417L13-16: Consider breaking up sentence into 2 for readability.

Sentence reorganized for clarity as requested.

18) P24424: Please complete figure caption. ql is not defined and T in caption should be T_abs as in Figure. As mentioned above, there is an inconsistency in notation between Na and Nd.

Caption expanded and details corrected as requested.

19) P24425: In caption write T_abs instead of T.

Changed as requested.

20) P24426: Caption states hour 8, header of Fig. says hour 6. Please change accordingly.

Fixed.

21) P24429: In Fig. Na should be <Na> consistent with caption, where it should be mentioned that <Na> is MBL-depth averaged Na. Also contour spacing of <Na> is not clear.

Caption clarified as requested; <Na> color scale is logarithmic, intermediate values added for clarity.

22) P24430: In particular dashed stream lines are hard to identify in bottom panel. Please make it clearer using either larger panels or thicker lines.

Panels enlarged.

23) P24433: A_TOA and A_Cld are not defined in caption.

Definition added to caption.

24) P24434: State scaling of rain mixing ratio in caption.

Changed as requested.

25) P24436: Why are not the same times shown as for Fig.6? It would allow for direct comparison.

While showing the same times would allow direct comparison, the uneven spacing makes it more difficult to gauge the dispersion rate, which is much more rapid in the cross roll case.

26) P24437: panel rows (a), (b), (c) and (d) are swapped between Fig and caption.

Fixed.

27) P24432 and P24439: Replace cyan with different color, as very close to blue when printed.

Changed as requested.

28) P24439: Consider defining RRTM in caption or putting in a reference to the text, where it is defined.

Acronym definition added to the caption.

Anonymous Referee #2

General comments:
It is very interesting to see the simulated mesoscale organization of cloud rolls, for which the formation mechanism is definitely worth exploring. I am sure that the authors must have looked into this, but not much is discussed in the paper. What are the key environmental conditions (e.g., shallow boundary layer, strong winds, large-scale forcing, etc.) for the LES model to produce such cloud rolls? As also noted in the paper, the presence of such roll structures can have a profound influence on turbulent mixing and cloud microphysical processes in the marine boundary layer, compared to the relatively more common open or closed cell structures. Thus the impact of ship emissions on clouds could be quite different as well. Wang and Feingold (2009b) showed the suppression of cloud formation surrounding the ship track in their drizzling open cell case, which limited the increase of domain average albedo caused by the ship emissions. This suppression effect does not show up in the simulations of this paper. Is this due to some unique mesoscale circulation or cloud dynamics for the cloud rolls? It would be nice to discuss more on this in section 6 in the context of aerosol impact on cloud regime shift.

The focus of our paper is on aerosol-cloud-precipitation interactions as opposed to the nuances of boundary layer dynamics and the observed roll structure. The review of Etling and Brown (1993) extensively covers the details of boundary layer roll dynamics. We have added a reference to the LES study of Glendening (1996), a pioneering example of LES applied to roll form boundary layers. These references together address the environment necessary to generate a roll form boundary layer, as well as the unique dynamical characteristics.

We agree that the effects of ship emissions of clouds in a roll-form boundary layer are indeed likely to be somewhat different than those of other low cloud cellular modes, such as those investigated by Wang and Feingold (2009b), and this is briefly discussed at the end of section 5.6. This is not merely a curiosity, as the study of Durkee et al. (2000a) found that ship tracks are most commonly found in boundary layers with a depth less than 600 m, where roll organization is more common. Our work, in addition to that of Berner et al. (2013), Wang and Feingold (2009b), and others, indicates that the impact of aerosol feedbacks on precipitation and cloud macrophysical properties is strongly dependent on the dominant mode of boundary layer organization, meaning these effects will be difficult to parameterize in larger scale models.

Specific comments and technical edits:

1) P24390, L5-10: it does not seem appropriate to use “positive or negative first/second indirect effect” to describe the increase or decrease of LWP. Reduction of LWP has nothing to do with the conventional aerosol indirect effects. Suggest remove “positive” and change “negative” to something like “opposite to”. Same for a few other places throughout the paper.

We are not the first to include liquid water path changes modulated by aerosol-cloud interactions within the second aerosol indirect effect (e.g. Stevens and Feingold, 2009). LWP reductions resulting from aerosol feedbacks on precipitation and macrophysics will necessarily impact cloud radiative forcing and thus constitute an aerosol indirect effect. Albrecht’s cloud lifetime effect implicitly requires a link between aerosol and LWP, and this is considered an indirect effect, so it seems artificial to consider more general aerosol-cloud macrophysics feedbacks as very different.

2) P24392, L10: measured by which instrument onboard C-130?

The albedo was measured using the multi-channel radiometer on the C-130. Full instrumentation details are included in Taylor and Ackerman (1999).

3) P24392, L14-17: what’s the size range for aerosol particles above cloud? Seems that they are not represented in the bulk aerosol scheme. Also, Nd and Nad (Na in the plot) should be clearly described in the figure caption. Any measurements of interstitial aerosols in the MBL? What is the reason to set initial MBL aerosol concentration substantially smaller than the measured cloud droplet number concentration?

In the case considered, there was negligible aerosol of any size in the region immediately above the cloud. Aerosols entering the marine boundary layer from the free troposphere are generally smaller than the accumulation mode and thus not explicitly represented within our model. In previous work, we simply ‘considered the particles entering the boundary layer via entrainment to instantaneously grow via condensation into viable CCN in order to avoid the complexities of multiple aerosol modes. While this ignores potentially important aerosol mechanisms and the time lags associated with their action, it still appears to be a useful approximation.

Nd and Nad are now defined in the figure caption. Interstitial aerosol measurements in the boundary layer were not available, as the C130 instrumentation included PCASP, FSSP, and 2-DC probes. The PCASP, which sizes smaller aerosols, suffers from droplet shattering artifacts within cloud, so in-cloud interstitial retrievals are not possible. Furthermore, the C-130 was unable to fly below cloud during the observations due to the extreme low cloud base, so it was not possible to reliably examine interstitial aerosol in the boundary layer in the case studied here.

The initialization of the boundary layer aerosol to a lower value was done as part of the tuning process. Because it takes the model some time to develop the boundary layer cloud and roll structure, the parameterized surface aerosol flux increases the boundary layer aerosol number concentration over the first few hours, and starting at this lower initial value then yields the approximately observed conditions once the cloud rolls and turbulence are fully developed.

4) P24394, L1: change "precipitation" to "raindrop". Having Lookup table for cloud droplet sedimentation as well?

Changed as requested; cloud droplets fall velocities are also obtained from a lookup table.

5) P24394, L11 and L15: no need to use "dry" to describe aerosols, which usually take up water to various extents.

Changed as requested.

6) P24394, L20: not quite clear why such a high model top (29km) is needed for the radiation schemes when using diurnally averaged radiative forcing in the simulations. I assume that the sounding does not cover to that height. What profiles are used then?

The sounding above the C-130 profile was inferred from ERA-Interim reanalysis. While the radiation is diurnally averaged, it still includes a full shortwave and longwave radiation calculation from RRTM. The version of the radiation code used with the build of SAM used in our study did not easily handle a truncated vertical domain, so it was simplest to extend the domain rather than reengineer the radiation code.

7) L24396, L4-5: Could the "large scale updraft" be verified from reanalysis? I am wondering how strong a large-scale lifting can really affect droplet nucleation.

We have reworded this sentence for clarity; the intended meaning is that the C-130 profiled through the updraft between boundary layer rolls that forms the cloud bands, not large scale in the sense of anything resolved by reanalysis.

8) L24396, L25: please describe how the divergence was converted to subsidence and used in the model.

Constant divergence from 3000m to the surface implies a linear subsidence profile. This is applied as a large scale nudging tendency on scalar fields

9) P24397, L11: which scheme is used to parameterize surface salt fluxes? Only the accumulation mode is accounted for?

We use a modified form of the Clarke et al. (2006) flux parameterization which accounts for only the accumulation mode contribution that constitute immediately viable CCN; this is briefly mentioned in section 3.

10) P24397, L15: "m" is missing in the units of radius.

Fixed.

11) P24397, L28: please verify if S = 2.3e16 or 1.5e16. The former is inconsistent with the value given in L15, and does not seem to give s=15000 per mg.

The reviewer is correct; on examining our original calculation, we used an f_{CCN} of 0.08, which combined with the air density 1.2 kg m^{-3} from the initial conditions implies S = 2.3e16 s^{-1}. On reexamining Hobbs et al. (2000), f_{CCN} is estimated at 4% to 18% for a range of ships similar to the *Sanko Peace* burning marine fuel oil, with aerosol flux estimates ranging from 4.5e15 s^{-1} to 1.5e16 s^{-1} for the four ships where detailed estimates were made. Ferek et al. (1998) suggested an initial f_{CCN} of 10% that may increase over time with plume dilution. At any rate, the aerosol initialization selected for the plume in our model is compatible with the observational estimates of similar ships, and does not seem unreasonable given the substantial uncertainty.

12) P24398: would be nice to have all simulations summarized in a table.

Since we present only four simulations, which are described at the beginning of Sec. 5, we feel a table of runs is unnecessary.

13) P24399, L11: no need to have "Cloud" here

Changed as requested.

14) P24399, L25: "unactivated, interstitial" is redundant and a little confusing. Suggest remove "unactivated"

Changed as requested.

15) P24400, L11: It seems that the albedo here is not just for the visible wavelengths. How accurate are the equations (2 and 3) and associated parameters for the entire SW spectrum? Does the underlying ocean surface have zero albedo?

The albedo as described on P24400 refers to one deduced from the RRTM SW calculation. The shortwave flux partitioned in the RRTM calculation includes a component outside the visible spectrum, but as the bulk of the flux is in at visible wavelengths, this seems a reasonable approximation.

16) P24400, L25-26: unclear about "the scale" in Figure 6.

Reorganized for clarity; the scale discussed in lines 25-26 is that on the Hovmoeller plots in Fig. 7.

17) P24402, L22-27: Is there a physical explanation why the results are different from those seen in previous studies?

We expect that the lack of suppression hinges on the significant differences in the turbulent organization of the boundary layer. The turbulent circulations in a roll-form collapsed boundary layer are driven largely by wind shear, while the cellular circulations considered by Wang and Feingold (2009b) are driven by buoyancy flux and cold pool dynamics. It may be that the strength of the secondary circulation generated by changes in precipitation and hence cold pooling are much weaker relative to the dominant shear driven flow, whereas in an open cell circulation the secondary flow is comparable in strength, but this is speculation.

18) P24406, L14-15: why lower cloud cover results in a low bias in Nd? Isn't the Nd averaged in clouds?

Changed as requested.

19) P24408, L24: the "360 minutes" after track injection (hour 8) does not match the "hour 15" in the figure.

The figure caption is correct; text amended to 420 minutes.

20) P24415, L7: typo for "occur"

Fixed.

21) P24417, L7: scavenging "of" interstitial aerosol?

Changed as requested.

22) P24425, Figure 3: What does the "large scale" mean here? Why don't use simulated vertical velocities instead of LWP for the conditional sampling?

"Large scale" is meant to refer to the lateral length-scale of the mesoscale rolls. Sampling is done by LWP as it seemed to yield slightly cleaner sampling than the vertical velocity in a y-average.

23) P24426, Figure 4: "hour 6" or "hour 8" in the title of the albedo panel? Strictly speaking, the plots must be from the run BaseSpinup rather than BaseTrack (said in the caption) as ship track does not appear in the albedo plot yet.

The title is incorrect—the frame is from the beginning of hour 8, and the title has been changed accordingly. The frame shown is the first frame of run BaseTrack, as the shiptrack is injected 30s after the run is branched from the end of BaseSpin.

24) P24427, Figure 5: also for run BaseSpinup? Please describe the white (cloud) contours in the figure caption. Why they look so different in the two panels?

The white contour marks the 0.01 g kg-1 cloud water boundary. The difference in appearance is due to the y-averaging of the fields in the upper panel, while the lower panel is a slice across the middle of the domain. The title and caption have been altered for clarity.

25) P24431, Figure 9: why domain-average Nd is shown rather than cloud-average?

Cloud fraction varies strongly with height, and a cloud conditional average would only average over the updraft bases, while higher in the profile would average across cloud in the updraft and updraft. Rather than add multiple layers of conditional sampling, it seemed simplest to average across the domain.

26) P24432, Figure 10: suggest change “sinks” in the titles to “budget” since sources are plotted too. Are these terms averaged within the boundary layer?

Changed to “boundary layer averaged budget”.

27) P24434, Figure 12 caption: change “times” to “hour” and add the experiment name.

Changed as requested.

28) P24437, Figure 15: remove units in the third row, as the quantity is not number concentration anymore.

Fixed.