Reply to referee #1

We would like to express our gratitude for the insightful review of our study. The discussion is very helpful and gives insightful remarks and valuable recommendations. We took them into account while preparing the revised version of the manuscript.

Below, the actual comments are given in blue color.
The text added to the revised version of the manuscript is marked in green.

Major Comments:
While the explanation of how the clouds were altered in the model was very clear and innovative, I have several comments about the general setup of the model and how the simulations were performed and analyzed.
First, why was the ICON model used, and more specifically, why was a global model used for this experimental setup? For such short runs, I would think a regional or at least a nested global model would be sufficient. You state that computational cost limited your runs.

The ICON model has now become a standard tool in atmospheric science in Germany following a seamless concept (one model frame for LES, NWP and climate modelling). It is widely used both by the national weather service and by many university researchers. Forecasts from ICON have been evaluated with DACCIWA campaign observations (separate paper in preparation). As a recently developed model, it fulfills highest standards and runs very efficiently. We agree that a regional model would have been sufficient here but the very flexible nesting capability of ICON allowed us to avoid any undesirable effects at the model boundaries. A classical limited area version of ICON that could be driven with existing ICON global data is currently in preparation but did not exist when we started the study. Therefore we tested the global ICON model with a nest and found that it performed well in our test cases. We also like to stress that computational cost was not the main limiting factor for the analysis, but the large amount of data, even though we restricted the output to West Africa.
We included a short explanation on p6l11:

Currently ICON is only configured as a global model but its high flexibility in terms of one- and two-way nesting allows a regional focus without any undesirable boundary effects sometimes observed for traditional limited-area models. It performed well compared to ERA Interim (ERA-I hereafter) in several test cases we ran for DACCIWA. The comaprison can be found in the supplementary material. The simulation period was not so   much limited by computational cost but by the large amount of output, since many different state variables had to be saved for the analysis.

Do you used fixed SSTs or an interactive ocean (or ocean surface)? Since the sim- ulations are so short, I don't expect the treatment of the ocean surface to have much effect, but it would be good to know.

SSTs are contained in the input conditions, but they are not updated during run-time. We also did not expect a strong effect because of the short simulation times. We added to the sentence on p31l12:

Note that in the NWP simulations SSTs stay largely constant during the short run time; they are initialised with ERA-I, but not updated during a 5-day simulation.

> For the EXPL runs, what is the domain of the nested grid? This is probably unimportant, but could have implications as odd things can happen on the boundary between nested grids.

We tested this and asked DWD for advice (to be on the safe side) and chose the grid large enough. The innermost grid was circular (this fitted best to the ICON base grid, which covers the earth in triangles). It had a radius of 30 degree and was centered on 0°E and 13°N. This way, it covered large parts of North-Western Africa. We added on p7l7:

> The nest has a circular domain centered on 0° E and 13° N with a radius of 30° such that it is large enough to avoid undesirable effects near the nest's boundary.

> Finally, what period is the data analysis averaged over? Is it the final four days of the five day simulations? This would be congruent with five day simulations having a single day overlap when the model was started every fourth day, however, it should be clearly stated.

Yes, that's true. We added on p8l4:

> The averages were created from the final four days of the five day simulations.

> Is this period the same for analysis of both the local and regional response?
> The timescale of the local and regional response is an important factor in interpreting your results, so you should clearly state your averaging period for analysis and support your choice with evidence from your simulations (as in the supplementary material) and/or the literature.

The averaging period is the same for the local and the regional response. We did not change any setting in between in order to keep the two responses comparable.
We added on p22l7:

> To answer these questions, we expanded the analysis of the sensitivity experiments and included the Sahel zone up to about 25° N.

> When describing the ICON model base state and comparing it to obs and reanalyses it may be helpful to also know how the control simulations compare with the obs/reanalyses/Hannak et al. In terms of cloud fraction and LWP/IWP. Is it a model that produces a reasonable amount of low level clouds in the base state? The discussion on pages 8 and 9 somewhat address this, but since the paper is on clouds, it might be nice to just state how the control simulations cloud fields compare to obs/reanalyses.

We looked at that and compared the profiles of low-level cloud cover, LWP, and wind speed to ERA-I. We found the base state to be satisfactory in our study region and wrote a paragraph about this. However, because the article became too long, we decided to remove it again. Now we have put a figure into the supplementary material and describe it as follows:

In addition to the general characterization of the meteorological conditions in southern West Africa for the variables precipitation and radiation, a brief discussion about the diurnal cycle of the vertical atmospheric structure is given for the wet monsoon season July, August and September 2006.
Comparison of ICON CLIM with ERA-I and observations confirms the applicability of the ICON model for the sensitivity experiments of the main article.

Figure S1 shows average profiles of cloud cover (CLC), relative humidity (RH), horizontal wind speed v_horiz as well as specific cloud water content q_c for 00, 06, 12 and 18 UTC (corresponding to local time in our study region). At 00 UTC the NLLJ is already well established and the low-level cloud deck is beginning to form (Fig. S1a).
ICON shows a considerably stronger jet than ERA-I reaching 7 m s^{-1} at 925-950 hPa and consistently lower values in CLC, RH and q_c. In contrast to ERA-I, ICON tends to concentrate cloud water in the upper parts of the cloud deck around 850 hPa.
The relatively small differences in RH between the two datasets (more than 90\% from 830 hPa downwards) in contrast to differences in CLC and q_c illustrates a substantial sensitivity to the subgrid-scale cloud scheme or possibly differences in spatial variance of RH, as the dependance of CLC on RH is quadratic in ICON. The tendency of stronger NLLJ and less cloud was also found in many climate models (Hannak et al. 2017). At midlevels around 560 hPa ICON shows a secondary peak in v_horiz, CLC, RH and q_c not found in the overall smoother and moister ERA-I profiles.

At 06 UTC the NLLJ is very similar to 00 UTC but the low-level cloud deck increases markedly in cover and q_c accompanied by an increase in RH to values well above 95 % below 900 hPa (Fig. S1b). Maximum CLC occurs at 950 hPa reaching 25% in ICON and about 45% in ERA-I, which is more realistic (cf. van der Linden et al., 2015).
Overall the discrepancies between the two models are qualitatively similar to 00 UTC (Fig. S1a). At midday (Fig. S1c), radiative heating lifts and dissolves the low-level cloud deck shifting the maximum in CLC and RH to 850 hPa, where a pronounced peak in q_c develops. Surface heating and turbulent mixing markedly slows down the low-level jet (e.g. 4.5 m s^{-1} in ICON) and decreases RH to under 90% below 900 hPa with ICON being substantially drier and less cloudy in that layer.
Finally at 18 UTC (Fig. S1d) the low-level jet starts re-accelerating, keeping the generally higher values in ICON found at all times of day. The deep daytime mixing has reduced CLC and q_c and created an almost vertically constant offset between the two modeling systems. RH is already increasing at this time of day, particularly in ICON, where also the sharp gradient in v_horiz suggests a beginning decoupling of the surface. Such an early evening transition is consistent with observations as documented in Fig. 3d in Schuster et al. (2013).
The comparison between the two datasets shows considerable biases at all times of day with generally higher low- and midlevel wind maxima in ICON but moister and more cloudy low levels in ERA-I. Investigating the reasons for these discrepancies is beyond the scope of this paper but the overall agreement in vertical structure and diurnal cycle suggests that sensitivities tested with ICON should be qualitatively meaningful.

The entire paragraph at the end of Page 12 and beginning of Page 13 is a bit confusing and unclear. (Page 12, Line 26 through Page 13, Line 5) Why do you connect the discrepancy in the longwave component with a dynamical response, but not the discrepancy in SSI? Could the change in upper-level clouds account for this difference?
It is important to point out that July 2006 might not have been the most average year, but I'm a bit perplexed by why that can explain SSI and not OLR. Maybe I'm missing

The main point was to find out, how much of the change in cloudiness could be attributed to the dynamical response in the fully non-linear model as compared to the radiative effect of the static cloud cover itself. Therefore we compared with the study by Hill et al. that only takes the latter into account. Unfortunately it is not easy to compare, because the time periods considered are not the same. This is why the characteristics of July 2006 relative to the climatological context matter. In addition, of course, clouds simulated by ICON can deviate from what CERES sees in reality. Therefore we did not make a statement about SSI because it is not possible to disentangle the two problems. For longwave is easier to interpret because there are qualitative changes (the changing sign of the variation in OLR).

We reworded the section on p13l25 to read:

The most plausible explanation is that the relatively dry July 2006 had overall less mid- and high-level clouds than the June–September 2006–2010 average used in Hill et al., leading to a largely consistent but relatively larger effect of modifying low-level cloudiness (consistent with Fig. 9 in Hill et al., 2018). This makes it hard to distinguish the purely radiative signal from the fully nonlinear dynamical response of the atmosphere. The latter is more distinguishable in the longwave component. The increase in deep convection with optically thinner low clouds in ICON PARAM leads to a decrease in OLR in the model on the order of 10 W m−2, while the radiative transfer calculations by Hill et al. show even a small increase. In contrast, for SLI the purely radiative
effect is a marked decrease, but ICON-PARAM shows almost constant SLI, likely due to combined dynamical effects of the increase in low-level temperature, deep convective clouds and column moisture (see Fig. 12).

Page 19, Line 15: "most striking" is a bit subjective, don't you think? I agree that it is very striking, but I was immediately more intrigued by the low level qv and qc, whose signals can be interpreted as due to changes in vertical mixing, some of which can be explained by TKE. I think describing and explaining the signals in this manner might be more causal, but this, I suppose, is more personal opinion. This is also a massive paragraph and could probably be broken up between these two features.

Reviewer 2 said exactly the same, so we rephrased it and tried to restructure the paragraph to be more readable.

Figure 8 shows the corresponding profiles for PARAM. Despite the overall consistent signal in rainfall and radiation as documented in Fig. 4, there are many substantial differences between the two sets of experiments.

Despite a larger SSI (see Fig. 4a), PARAM has a lower daytime increase in near-surface temperature, particularly at 15 and 18 UTC, suggesting a possible impact of the earlier triggering of convection in PARAM (see Fig. 5). Near surface q_v (Fig. 8b) is strongly decreased at 09 UTC, probably due to the earlier onset of PBL mixing with transparent clouds, and then strongly increased at 12 and 15 UTC, possibly due to the lack of deep mixing as in EXPL, leading to very large differences between the two sets of experiments. Combined, the changes in temperature and moisture lead to overall less pronounced changes in RH at low levels (both negative near the

surface and positive above; Fig. 8c), associated with mostly negative changes in q_c (Fig. 8e) except for 09 UTC. These explain the somewhat unexpected results for q_c discussed in the context of Figs. 7 and 6. In contrast to EXPL, PARAM operates a positive feedback mechanism, where a reduction in low cloud leads to a further reduction. This may clarify, why so many climate models show very large negative biases in cloud cover (Hannak et al., 2017).

Increased vertical mixing can be observed via TKE (Fig. 8d). Positive signals are restricted to the low levels during the day (09, 12 and 15 UTC), with the latter time showing indications for increased mixing reaching midlevels. All hours from 18 UTC to 09 UTC show decreased TKE below 600 hPa and hardly any change at all above that. One needs to bear in mind, however, that the mixing through convection is not reflected in TKE fields in PARAM. Nevertheless, the PARAM signals, at least at low levels, are in clear contrast to EXPL (Fig. 7d) where TKE increases everywhere. These differences are strong indicators that the interplay between PBL turbulence, shallow and deep convection fundamentally differs between the two model configurations. Particularly during nighttime, PARAM shows a slight stabilization in the temperature profile (Fig. 8a) above 925 hPa that appears to suppress turbulence generation in this layer. This cooling may be related to the enhanced NLLJ (Fig. 8f), but it is not clear why this effect does not work in EXPL, where an even more enhanced NLLJ and also a stabilization is observed (Figs. 7a and f). The change in mixing have profound impacts on many low-level fields, whereas more agreement between EXPL and PARAM is found at mid- and upper-levels, except for some changes in the diurnal cycle.

Overall this discussion demonstrates the enormous importance of vertical transport and mixing in a moist tropical environment where the PBL, low clouds and deep convection are closely coupled through radiative effects.

The first panel in Figure 9 is really fantastic! However, I feel that the aspect of it that discusses the differences between ICON and TRMM as well as the description in the text (Page 21, Line 10 through Page 22, Line 1 "As already discussed . . . gap between the rainfall maxima") should be moved to Section 3.1 where you are discussing the differences in the control simulations and the TRMM. When inserted here it subtracts from the main point of the section which is to show how the perturbation effect precipitation into the Sahel.

Thank you. We agree that it distracts from the main topic and inserted the averages into figure 2, where the description was already given. We added the following lines:

The small middle panels in Fig. 2 show the rainfall dependent on latitude but averaged over 8° W-8° E as further analysed in section\ \ref{precip3}. The rainfall maxima over the Niger Delta region and along the coast of Guinea, Sierra Leone and Liberia are not captured in this average, which explains the rather small values between 7-10° N.

We kept the first panel in Figure 9 because it is important for the reader to see the absolute decline in rainfall when traveling northwards.

Page 22, Line 11: "This may suggest that modulations to the WAM allow a slightly deeper penetration of rainfall into the continent but one month is probably too short to make any definite statements on this area." I am a bit confused by this statement. You

are suggesting that only looking at July 2006 might not be enough to make a robust statement?

We agree that this is a bit misleading and reworded the passage to now read:

However, given that in the northern Sahel rainfall is usually caused by few distinct, intense convective systems and that soil moisture perturbations become increasingly important, five-day simulations during one month is probably insufficient to make any definite statements for this area.

Page 25, Lines 32-34: "An interesting implication . . . no significant regional impacts." This statement makes it sound like you are generalizing your result to other regions, how do you support that claim? Or are you focusing on the response to just altering the clouds in the DACCIWA region? Your statement in the summary (beginning Page 31, Line 10 with "Therefore these results . . .") is much more well supported.

Here we refer only to the DACCIWA region, we made this clearer and the lines read now (on p27l1):

An interesting implication of this result is that whatever change in aerosol-radiation or aerosol-cloud interaction is caused through changes in anthropogenic emissions in the DACCIWA region, it will likely have measurable local impacts but probably no significant ramifications elsewhere.

Page 28, Lines 18-22: This paragraph feels out of place. Why are you concerned with how quickly the atmosphere returns to its normal state? The statement "impacts on higher and more remote regions can last days" is misleading as the supplemental material suggests that the residual impacts on remote regions are complicated by the chaotic nature of the atmosphere. I think the plot in the supplemental material is quite interesting and helps explore the timescales of the response, as well as the persistence of changes when the altered clouds are removed. It might be useful to plot the envelop of the anomalies in Figure S1, in order to show the variability between the ensemble members. This would help clarify what is a consistent response to the forcing and what is due to internal variability.

The original idea was to find out, if the experiment that we conducted has a long-lasting effect on the structure of the atmosphere, or if it is a short-lived phenomenon. In the first case, an influence on remote regions would appear more likely. We agree that the chaotic nature of the system makes it difficult to discuss the life-time effect. We plotted the respective figure again and included the envelope in the bias. Since the envelope stays quite constant with time after the switch-off for the boundary layer variables, we believe that our statements are correct. We modified the passage on p29l24 in the main manuscript in the following way:

Impacts on higher levels and more remote regions can last for days but the signals hardly stand out from the high level of background variations indicating the chaotic nature of the atmosphere.

And in the supplementary material in the figure caption of Fig. S1:

> The corresponding differences are shown as green lines in (f)--(j), where the shaded area
> denotes the maximum and minimum values of the time-series.

And added to the corresponding text on p4l13 Where we had already referred to the chaotic nature
of the system:

> The latter is reflected in the growing shaded areas denoting the envelope of all runs in the
> bias in the right-hand side panels of Fig. \ref{longr}.

> What is the remote (Sahel) response of precipitation for the positive values of f_op = 3
> and 10 in the PARAM experiments? I realize these are extreme values, but it would be
> interesting to see if the same signals appear.

In the Sahel there is not too much difference noticeable due to the already low absolute values of
precipitation. The DACCIWA-box, on the other hand, dries out further as already indicated in Fig.
4c. Below please find a figure (middle panel control run, right panels f_op = 3, left panel
f_op=0.1) that illustrates the regional distribution of this drying.
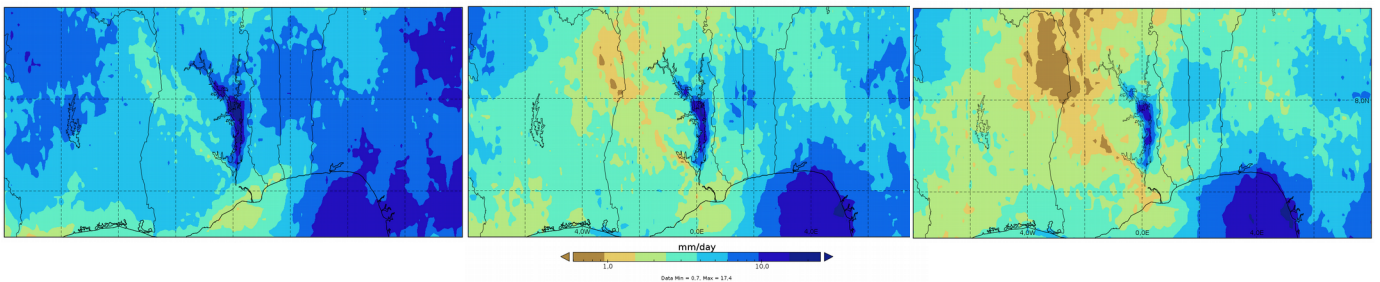
and the same plot as figure 9 in the paper, but with f_op=3 and f_op=10 included:
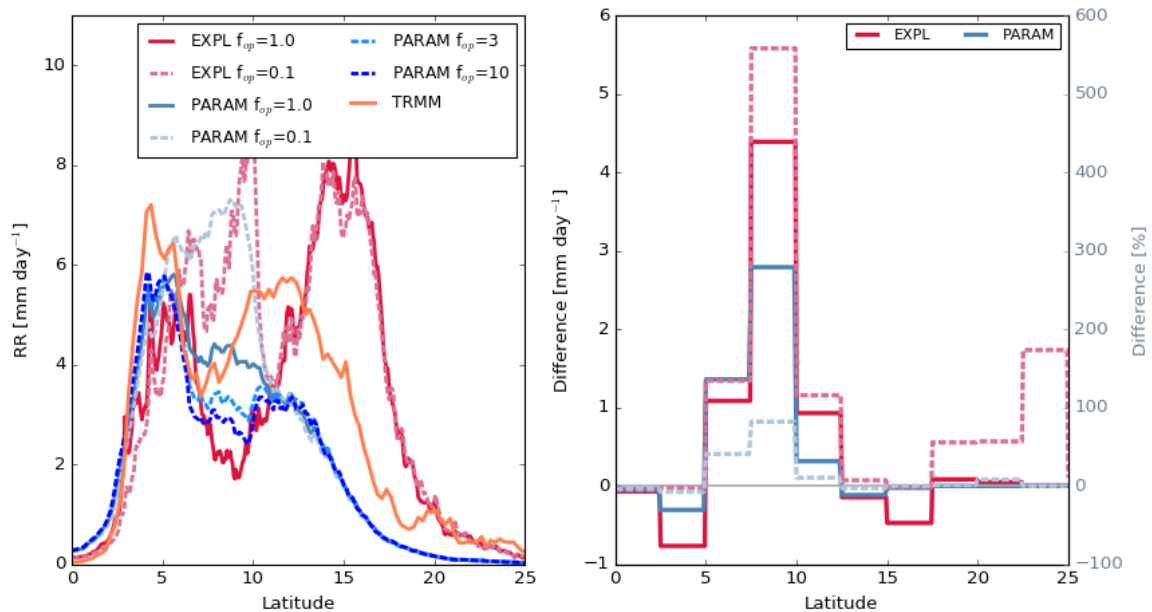
Thank you, we tried our best to improve the punctuation.

Page 1, Line 7: effect

Done.

Page 1, Line 24: "do not show skillful forecasts of precipitation for the next days", is awkward, perhaps "do not produce skillful short term precipitation forecasts" is clearer.

Agreed.

Page 2, Line 9: "determined by the WAM system", is unclear, what about the WAM system determines these characteristics? "are connected within the WAM system" emphasizes the challenge of the interconnectedness of the WAM.

That's better, thank you.

Page 2, Line 32: I would remove "which is currently gaining increasing attention".

Done.

Page 3, Line 17: misrepresenting

Done.

Page 3, Lines 19-30: I would reorganize this entire paragraph to flow better.
This study is part of the Dynamics-Aerosol-Chemistry-Cloud Interactions in West Africa (DACCIWA) project (Knippertz et al., 2015) that aims to better understand the consequences of the rapid increase of anthropogenic emissions in West Africa on the local air quality, weather and climate. To the best of our knowledge, it is the first to analyze the radiative impact of the low-level cloudiness over southern West Africa on the thermodynamics and dynamics of the regional atmospheric system in a fully non-linear and systematic way. The analysis is based on a number of targeted sensitivity experiments using the numerical weather prediction model ICON (Icosahedral Nonhydrostatic), systematically changing the optical thickness of the model clouds. This allows us to clarify the impact of the inter-model spread in cloudiness found in Hannak et al. (2017) on the overall monsoon development in both parameterized and explicit regimes of convection. Although aerosols are not directly modeled in our experiments, the effects found for imposed changes of cloud optical thickness also help to understand variations of the natural system brought about by aerosol effects on cloud properties and radiation, which in a similar way control the amount of shortwave radiation reaching the surface or interact with clouds through modifications in the diurnal cycle of the PBL (e.g. Deetz et al., 2018a).

Thank you so much, this is really appreciated and we put it into the manuscript.

Page 6, Line 4: Green's

Done.

Page 6, Line 7: (Dee et al., 2011), and do not use data assimilation. (the comma helps separate the idea that the simulations are not using assimilation instead of ERA-I,

Done.

Thank you.

Since reviewer #1 had an issue with this figure, it was modified and now contains a box in each subplot.

Done.

That was actually done on purpose. The range is similar to the other radiative quantities to illustrate that the variability is not so large for SLI. Therefore, we would like to keep it this way.

Done.

Done.

Done.

Corrected.

We totally overlooked that, thank you!

Done.

We added on p27l3:

This quite noticeable impact was found in the simulations with explicitly simulated convection. The influence of parametrization of convection in this experiment will be discussed next.

Page 28, Line 9: remove "just mentioned"

Done.

Section 4, First Paragraph: The second sentence introduces the idea of representation of clouds in models, but then immediately returns to the idea of describing the clouds in reality. The paragraph might flow better this way:
In the present study, we analyzed the role of low-level clouds over southern West Africa on the local meteorology and larger monsoon system. They frequently form during the night close to the surface and often persist long into the following day. At their maximum diurnal extent, they cover a vast area of about 850 000 km2 in southern West Africa (van der Linden et al., 2015). Their formation is linked to cold advection and turbulent mixing associated with the NLLJ and radiative cooling (Schrage and Fink, 2012; Schuster et al., 2013; Kalthoff et al., 2018). These clouds play an important role in the energy budget and diurnal cycle during summertime and tend to be badly represented in many climate models (Hannak et al., 2017). The role of these clouds in the WAM system was assessed here for the first time in a fully nonlinear way via sensitivity experiments using the ICON model from the DWD in NWP mode for July 2006.

We would like thank you very much for this rephrasing, we changed the paragraph accordingly.

Supplemental Material, Page 1, Line 4: Would read better as: "To investigate this we use EXPL experiments, in which f_op = 0.1 is applied for the first 4 days . . ." The phrase "as in EXPL" makes it unclear what type of experiment is being done here.

Thank you, we adapted it.