We would like to extend our gratitude to both reviewers for taking the time to review our paper. The final author comments in response to both referee comments are listed below. We uploaded both a marked-up version, where all text additions and subtractions appear in blue, and a final version that excludes the mark-up. The lines referenced below reference the marked-up version. Thanks again, and please let us know if there are any outstanding questions or comments.

<u>Author Updates Unrelated to RCs:</u>

1. In regards to a previous editor question about CERES, we investigated an improvement to our methods comparing modeled to measured SWout. This improvement was successfully integrated, and leads to a more precise comparison. The update can be seen in Fig. 5, and in the text starting at Line 337. The result suggests the model is slightly less biased compared to the satellite than we originally calculated. An equivalent update was applied to Fig. S7 in the SI as well.

2. After submitting our draft we noticed a mistake in our extinction/absorption analysis. For the smoke optical property evaluation relevant to Figs. 8-10, we previously used model values at ambient temperature and pressure. However, the measured data is at standard temperature and pressure (STP). We converted all model values to STP for a like to like comparison and made text edits to better communicate this. This change impacts Figs. 8-10, and text throughout Sect. 3.4. The model now agrees better with measurements. For all data, among all domains and flights, the absorption coefficient normalized mean bias was reduced from -20 % to ~0 %, and extinction coefficient normalized mean bias was reduced from - 57 % to -50 %. An equivalent update was applied to Fig. S12-14 in the SI as well.

<u>RC1 and 2 from Referee #1:</u>

1. Abstract, Line 1: “BBA” should be defined. “Smoke aerosols” can just be replaced with “Biomass burning aerosols.”
   Line 1: Smoke aerosols replaced with “Biomass Burning Aerosols”

2. Lines 48-49: Can the authors provide a range of SEA REs from the references cited in this sentence?

We agree this would be a helpful change, though previous studies typically all examine different spatial regions and time spans making a like to like comparison, and therefore our ability to provide a set range, difficult to determine. This is why we discuss a qualitative comparison for the remainder of the paragraph. However, the cited references do make an effort to translate previous studies into a like to like comparison for just the DRE.

Line 50: Added the DRE range from references cited: “with monthly mean DRE warming currently ranging from 0 to +8 W m^{-2}”

3. Line 56: Can the authors please explain why they claim that IRE and SDRE cannot be calculated from observations? Specifically, the indirect radiative effect is the response of cloud changes to aerosol and has been quantified using ground-based (Mccomiskey et al., 2009; Varble et al., 2023; Tang et al., 2023) and satellite remote sensing (Christensen et al., 2020; Christensen et al., 2016; Quaas et al., 2020; Gryspeerdt et al., 2017) observations.
   Our intention was to communicate that calculating IRE and SDRE from observations is difficult for smoke-cloud interactions because it is difficult to disentangle which interactions belong to SDRE and which to IRE. We have not come across an observational based study that estimates the SDRE. Many of the papers cited in this RC question investigate the Twomey effect, then briefly touch on this disentanglement difficulty. For example, Grysperdt et al., 2017 discusses the limitations in estimating meteorological feedbacks (E3) and additional meteorological confounders (E4). Varble et al., 2023 similarly states that unaccounted for covariations between LWP and CCN need to be addressed further.

   Line 58: Clarified the difficulty is in disentangling IRE and SDRE from observations: “While the IRE and SDRE are difficult to disentangle and calculate from observations for BBA-cloud interactions, it is reasonable to assume all aerosol above low clouds is BBA to compute an observed DRE that is comparable to our simulated above-cloud DREs.”

4. Line 58: To orient the reader such that the proceeding studies were done in a similar region, the authors should note that these DRE estimates were conducted over the Southeast Atlantic.
   Line 58: Added several instances of “in the SEA” to this paragraph

5. Can the authors provide a table, either in the main text or supplement, listing their model experiments and set-up conditions that are described in Section 2.1 and 2.3?
   Yes, great suggestion. We added 2 tables to summarize the content of Sects. 2.1-2.3.

   Line 130: Added Table 1, summary of the model setup common among all simulations run in this study. Added Table 2, summary of the experimental setup.

6. Line 360-374: Did the authors adjust either of the scattering or absorption coefficients to match wavelengths before summing? This is likely not to be significantly different from just summing the two, but I am interested if this changes the result at all. In Lines 366-367, it appears the authors took the average of the extinction coefficients to approximate at 530 nm. Is there support for why this approach was done rather than calculating an Angstrom exponent (e.g. (De Faria et al., 2021) EQ. 8-9). The similar question is posed for the LASIC extinction (Lines 369-373) as in Dedrick et al. (2022). It is also important to note that the LASIC nephelometer measurements were not conducted in “dry” (<40%) conditions (Dedrick et al., 2022; Zuidema et al., 2018), which may bias results for LASIC comparisons.
   We did not use the Angstrom Exponent in the submitted version, and we agree that it would be straightforward to use it to make this comparison more precise. We converted all the optical variables pertaining to Figs. 8-10 to 550 nm (at STP, see our earlier “Author update”) using standard equations for the Angstrom exponent (e.g. Eqs. 8 and 9 from De Faria et al., 2021). The wavelength correction introduced a negligible difference to summing absorption at 530 nm and scattering at 550 nm to estimate extinction at 550 nm with respect to ORACLES. The wavelength correction introduced a 5% difference on average relative to our old method of averaging extinction at 405 nm and 658 nm to approximate extinction at 530 nm with respect to CLARIFY. By far the biggest impact to the change in results seen was converting model values from ambient to standard temperature and pressure (author updates #2). An equivalent update was applied to Fig. S12-14 in the SI as well.

   We appreciate the reviewer’s comment about LASIC measurements not being in fully dry conditions. However, in Dedrick et al., 2022, it is mentioned that some drying does take place, and given the non-linearity of aerosol water uptake as a function of relative humidity, they conclude that dry model diagnostics are a better comparison to LASIC measurements than ambient RH model diagnostics. We added a reference to this discussion to the text.

Line 394: Start of text modifications for comparing all properties at 550 nm at STP. These changes impact Figs. 8-10 (and S12-S14 in the SI).

Line 413: Added reference to Dedrick et al., 2022 to note our choice in using dry model diagnostics: "These LASIC measurements were performed with some drying; RH was estimated to be below 25 % for absorption and at approximately 45-60 % for scattering (Barrett et al.,2022). Although not completely dry (RH < 40 %), the results of Dedrick et al. (2022) indicate that dry model diagnostics are a closer comparison to the LASIC measurements than ambient model diagnostics"

7. Line 376-378: please specify that the AMS OA mass is non-refractory.
   Line 420: Specified OA as non-refractory

8. Line 382: Non-refractory OA mass concentrations during LASIC were measured by an ACSM, not an AMS.
   Line 426: ASM corrected to ACSM

<u>RC3 from Dr. Zapata:</u>

L1 BBA abbreviation seems confusing there
Line 1: Smoke aerosols replaced with "Biomass Burning Aerosols"

L49 Would it be helpful to include the ranges of values for each RE?
We agree this would be a helpful change, though previous studies typically all examine different spatial regions and time spans making a like to like comparison, and therefore our ability to provide a set range, difficult to determine. This is why we discuss a qualitative comparison for the remainder of the paragraph. However, the cited references do make an effort to translate previous studies into a like to like comparison for just the DRE.

Line 50: Added the DRE range from references cited: "with monthly mean DRE warming currently ranging from 0 to +8 W m^{-2}"

L128 About nudging the wind only above the BL: could it enhance wind shear at that level and modify the BL dynamics?

Nudging does impact wind shear at that level, however, there is a mechanism built in to accommodate the BL. Nudging is applied gradually, ramping up to full strength over 4 model levels spanning ~1 km. This nudging set up minimizes BL disruption while gradually moving conditions toward measurement. Following Zhang et al., 2014 (reference listed at Line 1041), we believe our approach to be reasonable for the application.

Line 136: Added clarification on the nudging altitude and ramping: “We follow the recommendations of Zhang et al. (2014) to nudge only horizontal (u,v) wind fields above 1920 m altitude, approximately corresponding to the height of the boundary layer (BL), with nudging strength ramping up to full over approximately 1 km (at around 3000 m altitude).”

L129 Nudged simulations: There is no moisture nudging either, correct? Have any previous studies analyzed the effect of nudging the temperature and moisture fields? What is the characteristic nudging time?
Correct, nudging applies only to the horizontal (u,v) wind field, not moisture fields. We have not come across a study that nudges moisture fields, but this would be interesting. Previous studies have examined nudging temperature, however this is not recommended for studies of aerosol-cloud interactions per Zhang et al., 2014 (Line 1041). The characteristic nudging time follows the original recommendation of Telford et al, 2008, who first developed the nudging scheme for this model. This reference is now cited in the text.

Line 133: Added characteristic nudging time and source data time resolution of 6 hours: “The nudged simulations are initialized to UM global operational meteorology on 01 Aug 2017, and thereafter nudged to European Centre for Medium-Range Weather Forecast (ECMWF) version 5 reanalysis, or ERA5 files, which provide fresh data every 6 hours. Nudging is applied at each timestep with a characteristic relaxation timescale (Telford et al., 2008) also set to 6 hours.”

Line 136: Added a clarification on nudging pertaining to (u,v) winds: “We follow the recommendations of Zhang et al. (2014) to nudge only horizontal (u,v) wind fields…”

L138 Reinitialized simulations: I don’t completely follow the procedure for the whole studied period. Does ”1 day reinitialization with 1 day forecasts (1d)” mean that the simulation is 2 days long and only the second day is used?

Not quite, we meant that each forecast simulation is 1 day long and that day is used for analysis. But, for example, in our ‘5d’ simulations, the forecast simulations are 7 days long and only the last 5 days of each simulation is used.

Line 148: We rephrased the description of the reinitialized simulations run in this study: “We tested the following simulation setups, where the interval between reinitializations is always equal to the time period used in the analysis we present: 1 day long forecasts of which 1 day is used in the analysis (1d), 3 day long forecasts of which the last 2 days is used (2d), and two sets of 7 day long forecasts of which the last 5 days are used, one starting on 01 Aug 2017 (5d) and the other on 03 Aug 2017 (5dalt). For example in the 5d setup, shown later in Fig. 4, the seven-day long forecasts start on 1 Aug, 6 Aug and each 5 days until 5 Sept. Simulation results from the whole of the first forecast, the last 5 days of each subsequent forecast, and finally the 7 and 8 Sept only from the last forecast, are evaluated and used in our results. The overlap time, equal to the difference between the forecast length and the time period used for analysis, serves as spin-up time to minimize discontinuity between reinitialized cycles.”

Line 145: We removed several instances of “reinitialization frequency” from the paper starting with this paragraph.

Does it mean that there are N independent 2-day reinitialized runs to cover the N-day period?
Let’s call a reinitialized run a cycle. In the ‘1d’ case there are N independent 1-day reinitialized runs to cover the N-day period. In the ‘5d’ case there are N/5 independent 7-day reinitialized runs to cover the N-day period.

Or does it mean that there is a full N-day simulation, and every 2 days, all fields are reinitialized?
No, see the previous reply.

Same question about the other configurations. Later, you mention overlap time; now I’m interpreting that it means there is no spin-up period in the 1 d simulation.
For each setup,

- Forecast length = total days in each reinitialized run (also called a cycle)
- Spin-up period = forecast length - total days of the cycle used for analysis

Going through each configuration,

- 5d: 7 day cycles, 5 of which are used for analysis and 2 of which serve as a spin-up period
- 2d: 3 day cycles, 2 of which are used for analysis and 1 of which serves as a spin-up period
- 1d: 1 day cycles, the whole 1 day is used for analysis. You are correct, there is no spin-up period associated with this setup. This setup was designed to test the limits of the reinitialization method.

The spin-up period can also be denoted as overlap time because the spin-up period of a given cycle overlaps with the analysis period of the previous cycle.

We hope our rephrasing of the text at Line 148 (new text provided earlier in this response) makes this clearer.

L207 This is an interesting point. Since cloudiness can affect the results, does any existing research, or could future research consider this? Is it possible to use some cloudiness variable and analyze RE as a function of it?
We agree this is very interesting. The change in cloud fraction due to the semi-direct effect, from SI Figure S17, is +5.5% in the nudged simulations. If we assume the IRE is dominated by the Twomey effect (speculatively, Figure S15 suggests this is likely, but we did not do any calculations to back this up) we could therefore infer that the real IRE should be 5.5% higher. This is interesting, and worth further more careful study. However, because 5.5% is significantly smaller than the variability in the IRE between simulations (e.g. 5d and 5dalt IREs differ by 17%), and smaller than the effect of different droplet concentrations between simulations with and without absorption, which we discuss in Section 4.7, it does seem to be a relatively minor effect. We did not follow this up in the revised paper text.

L261 Difference in moisture above the BL: Do you mean that this is explained by differences in wind profiles because of advection?
Yes that is what we mean, differences in smoke advection arising from wind field differences between the 5d and Nudged simulations. Smoke and lower-free-tropospheric moisture are correlated. This point is better made in the paragraph below on evaluating RH, so we moved the sentiment down a paragraph.

Line 279: Removed the sentence about RH from the temperature paragraph

Line 289: Added back in the points deleted above: “ The performance of Nudged$_{bb}$ is again reasonably consistent with that of 5d$_{bb}$ overall, although the nudged simulation tends to have higher free-tropospheric RH compared to observations, while the 5d has lower free-tropospheric RH. Elevated water vapor signals of up to 60 % RH in the free troposphere in the SEA are often associated with smoke plumes (Pistone et al., 2021, 2019). This is therefore a likely signature of different wind fields impacting smoke advection, and will likely affect (ambient) extinction coefficients, aerosol optical depth, and the DRE results between these simulations.”

L263 Regarding the better caption of inversion height, is it 50 m below or above the observations?

Both simulations produce inversions slightly below observations.

Line 278: Added clarification that simulations are lower than observation: “The temperature profiles of both simulations are very similar, with both producing an inversion slightly below the observations on average.”

L274 Does the free tropospheric RH correlate well with the height? Or the upper inversion presence?

Our intention was to specify height here. However, smoke absorption should in theory strengthen the presence of an inversion as well.

Line 297: Added clarification this is discussing height

Fig. 4,S5,S6 Why do all cases have an initially similar LWP for the remote domain but not for the coastal domain?

All simulations are initialized with the same file, so the value of all simulated LWP is equal at the starting point on 01 Aug 2017. This is only true for the simulations though. We are unsure why this initial simulation value is closer to the observed value for the remote domain compared to the coastal domain.

It would be nice to know when the 5d bb case is being reinitialized. (Same for all the time series)

For 5d, the simulation runs 7 day forecasts where the first 2 days are spin-up and the last 5 days are used for analysis. The forecasts occur (01 Aug - 07 Aug), (06

Aug – 12 Aug), (11 Aug – 17 Aug), (16 Aug, 22 Aug), (21 Aug – 27 Aug), (26 Aug – 01 Sept), (31 Aug – 06 Sept), and (05 Sept – 08 Sept), with the date of reinitialization being the first date of each forecast. However, the data used to make up the time series consists of the first full forecast followed by the last 5 days of each consecutive forecast. In other words, each forecast has an allotted spin-up time of 2 days not counted toward analysis. We decided the transition between forecasts is the more appropriate addition to this time series, rather than the date of reinitialization which includes the spin-up period not used in our analysis. We therefore added vertical lines to Figure 4 to show the transition between forecasts. Equivalent changes were made to Figure S5, S6, S8 and S9 in the SI. We hope our previous changes to the reinitialized setup description at Line 148, described above, in addition to this plot change allows for an improved understanding.

Figure 4, S5, S6, S8, S9: Added vertical lines to show the transition between forecasts. Figure captions were updated to describe this as well.

L299 Could the difference stem from both decoupling and inversion height mismatch? Could a sensitivity test with finer height resolution help elucidate the difference? (that could be proposed for future research)
Yes, inversion mismatch is another valid potential source. Thank you for the suggestion.

Line 325: Added “Another potential cause of the overestimated LWP is that the simulated inversion heights are slightly lower than observations (Fig. 2)”

L330 It is interesting that the SWout differences are smaller than for LWP. Is there a way to theoretically extrapolate the difference in LWP to an estimated difference in SWout? Maybe comment on other differences like RH above cloud that could contribute to this diminished bias?
As well as LWP, differences in cloud fraction, droplet size, cloud top height, or free tropospheric RH could all affect SWout.. In light of Fig. 2, we don’t think RH is a notable factor for this case. The most likely explanation is differences in subgrid cloud fraction.

Line 348: Added a note discussing the implications of comparing SWout to LWP: “Comparing the small 6 % overestimate in SWout to the more moderate 20 %

overestimate in LWP, it seems likely that the model has a small low bias in low cloud fraction (CF)."

Line 358: Refined our conclusions with respect to cloud fraction. Added: "While there may be compensating errors between CF and LWP, the relatively small biases in LWP together with the good simulation of SWout likely constrains the CF to be simulated reasonably well."

L389 How do these differences relate spatially to the known biases on cloud thickness?

L389 (now L434) pertains to our absorption coefficient results. Through the semi-direct effect, smoke absorption can impact cloud optical thickness. At L311 we note; ''Modeled LWP bias is particularly over-estimated in the northern region of the coastal domain, and southern region of the remote domain." We did not identify a similar latitudinal variation of the smoke absorption that would indicate correlation with the clouds. We therefore do not define a spatial relationship between the smoke absorption and LWP biases.

L603 Any ideas on a way to modify the formulae to overcome this issue?

This is an excellent question, but unfortunately we could not identify any straightforward changes to the formulae that would help here. One possible way to resolve the interference between IRE and SDRE would be to run simulations without absorption in which the cloud droplet concentration is prescribed to equal the concentration in the simulation with absorption. But this capability is not available to us at present: we would need to write a significant amount of new Fortran code if we wanted our no-aerosol-absorption (noaa) simulations to read in an external field of cloud droplet concentrations, and so testing this is outside the scope of this current study.

L685 Could the 5d recommendation depend on the wind magnitude for the analyzed days? Do you think that another period with weaker winds would lead to the same conclusion?

That is a good point. We will emphasize that this result pertains to the SEA and is likely dependent on wind strength. We do expect to see a different result for a different region. However, we would not expect to see a large variation in our results for this same SEA region over a different time period within the African fire season (July-October), because while there are differences in smoke transport within and between fire seasons (e.g. Tatro and Zuidema, 2025, referenced at Line 1015) we think they are not sufficiently large to affect our conclusions. However, these answers are speculative. We still recommend our techniques be explored in future similar research.

Line 734: Refined our conclusion to note results may be dependent on wind strength: “We recommend these techniques for future similar research, as the appropriate choice for forecasting and spin-up length may be dependent on wind strength.”