Reponses to reviewer 2

The authors have performed a matrix of simulations to investigate the relative roles of urban heat island and aerosol forcing on the behavior of an observed storm in Houston. They find that both mechanisms lead to increased precipitation for a single cell on their study day. This is a valuable approach and a necessary investigation of multiple forcing mechanisms.

The authors use a single microphysical scheme in a single model, which leads to the ability to directly compare the results in the simulation matrix. The authors show that the simulations with full aerosol and land cover effects are better in some respects than those without. The aerosol effect contributes the most toward bringing the simulations closer to the observations of surface precipitation in these simulations ... and the effect is quite significant as noted on lines 345-6. The aerosol dramatically accelerates the growth of the cloud through the mixed phase zone, consistent with prior studies investigating the aerosol invigoration hypothesis.

We thank the reviewer for your time and constructive comments. We have provided detailed responses point-by-point as below.

More generally, the single model used in this study is one among many state of the art models, and as with any parameterization must build in a set of claims about microphysical processes. The study demonstrates that the aerosol invigoration mechanism *can* be large *if* these claims are correct. I would like to see this latter point emphasized more fully. The authors should review of what may or may not be correct about the microphysical process claims embodied in the model.

The most relevant portions of the study that could use clarification are the additional aerosol process models described on lines 133-134. These are relatively new and probably are driving much of the aerosol effects in this study. What are the likely errors in process rates in these parameterizations? What are the likely errors in thermodynamic characterization? Are they as large as the run-to-run difference the authors observe?

The point is not to relitigate prior work in full, but to the extent the authors can clarify the precision of their forecasts the more value the study has in advancing our shared understanding of how to properly model mixed phase microphysics and how to accurately partition the relative roles of aerosol and thermodynamic influences on convection.

Those are all great questions. We have added a paragraph in Summary and Discussion section to address these questions, that is,

"For simulating aerosol-deep convective cloud interactions, there are a few key modeling requirements as summarized in Fan et al. (2016), such as (1) the prognostic supersaturation is needed for secondary aerosol activation, condensation, and evaporation calculations, (2) hydrometeor size distributions need to be prognostic to physically simulate the responses of microphysical processes to CCN changes, and (3) aerosols need to be prognostic, and fixed aerosol concentrations gave unrealistic cloud properties and qualitatively changed aerosol impacts on convective intensity (Fan et al., 2012). With thee SBM used in this study, all these criteria are satisfied. Furthermore, for (3), we are not only prognosing aerosol numbers but also aerosol composition and size distribution by coupling the SBM with the chemistry and aerosol components. With this coupling, the spatial heterogeneity of aerosols is considered. Also, aerosol regeneration and wet removal processes can be more physically accounted for compared with the WRF-Chem with two-moment bulk schemes (Gao et al., 2016). The spatial heterogeneity of aerosols was shown to play an important role in simulating a torrential rain event observed over Seoul, Korea (Lee et al., 2018). However, bin schemes also have uncertainties in representing ice-related processes mainly due to our poor understanding of convective microphysics such as ice nucleation and riming processes. In particular, the conversions between different ice categories are also determined by threshold sizes or masses. However, those uncertainties are not expected to qualitatively change the warm-phase invigoration mechanism which occurs via enhanced condensation. In the companion paper Zhang et al. (2020), we carried out a small number of ensemble simulations for the anthropogenic aerosol effects for the same case and the results are consistent with this study, indicating this mechanism is robust with the initial thermodynamic and dynamic perturbations. More sophisticated uncertainty qualifications can be done in future with a larger number of ensembles when computer power becomes more advanced."

Below, I also request clarification of the meteorological setting and terminology used by the authors.

Major concerns

The authors claim SBM_anth in Zhang et al. (2020, preprint) is the same as LandAero in this study, but Figs. 7 and 5, respectively, do not match. Likewise, the associated CFADs of reflectivity do not match, though this may be due to a different analysis box size.

Yes, the analysis area is different between the two studies. Here we are focusing on the convective cell in Houston. Besides, Zhang et al. (2020, preprint) represents the ensemble mean of the three ensemble runs with three different initialization time. However, due to the expensive computational cost, we didn't do ensemble runs for No_Land and No_LandAero. The analysis in this study showed the result of one member for fair comparison between four simulations.

Compared to the observations, the CFAD shows significantly enhanced probabilities of moderate reflectivity at high altitudes (line 241). It also significantly underestimates the prevalence of 40-50 dBZ values at and above the homogeneous freezing level. What is the origin of this overestimate? One interpretation of this is that the SBM is too aggressive in the production of smaller ice crystals that grow into snow at high altitudes, at the expense of graupel. Given the authors' interest in an aerosol invigoration pathway, it seems that matching the ice population in glaciated clouds is more important than matching higher reflectivity values at low altitudes, which can result from many intermediate microphysical steps.

We agree that underestimating reflectivity at high levels is probably due to the model biases in simulating snow (should not have much graupel above 10 km). We have added a sentence discussing it, i.e., "At the upper levels (> 10 km), the model underestimates the large reflectivities (> 35 dBZ), suggesting the model does not get enough snow" (Line 250-252).

As the authors note on line 250, rain rates > 2 mm/hr last about 25% longer in the simulation. This is in common among all simulations, and so would seem to indicate a general over-production of precipitation duration by the SBM scheme overall. This would be consistent with an overactive snow/ice crystal microphysics scheme instead of a shorter-lived cell tied to faster-precipitating graupel. Why is the standard for "better" maximum reflectivity and areal coverage, but not duration of precipitation?

We did bring up the longer precipitation duration bias with the model. Since all simulations get longer duration than the observation, this model bias is not caused by the effects of aerosol and land surface. It could be related to many things such as problems in meteorological conditions, the interactions of microphysics and dynamics, microphysics parameterizations, the coupling of microphysics with other part of physics (PBL, radiative, ..).

Line 88: while Chen et al. (2011) also use the term "prevailing wind," that term is misleading, especially here in a review of how land-atmosphere interactions *generally* work in Houston. The climatological meaning of prevailing wind (c.f. the AMS Glossary) is the expected value of the wind in a given location over a climatological period. For Houston in the summer, the prevailing wind is onshore, from the southeast (https://mesonet.agron.iastate.edu/sites/windrose.phtml?station=IAH&network=TX_ASOS), and that pattern persists for stations further inland. Chen's study is of an anomalous day that may be representative of bad air quality days in Houston, but it is probably not typical of summer wind patterns or sea breeze interactions in Houston. It seems the authors are concerned in this paragraph with describing the general role of Houston's urban environment in influencing the sea breeze, and so this paragraph should be revised to more accurately reflect Houston's environment.

Thank you for the clarification. We have made changes to the sentences of the paragraph that inappropriately used "prevailing" winds. Now it should be accurate to describe the Houston environment. The revised sentences now read as "The strength and inland propagation of sea breeze circulation can be influenced by land/sea surface temperature contrast, land use/land cover, and the synoptic flow (e.g., Angevine et al., 2006; Bao et al., 2005; Chen et al., 2011). Chen et al. (2011) indicated that the existence of the Houston city favored stagnation because the inland penetration of the sea breeze counteracted the synoptic flow in a case study" (Line 86-90).

Line 112: Neither this study nor Zhang et al. (2020) give a synoptic or mesoscale meteorological overview of this case. Where was the trailing front? How did the meteorology, including the sea breeze, develop during the day in comparison to the expected behavior of the land and sea breeze diurnal cycle in Houston? Rosenfeld et al. (2014) is not listed in the references, and the ACPC website is not a suitable reference for long-term archival purposes.

We have added Fig. 1 and the following discussion to give an overview of the synoptic meteorological conditions of the case, i.e., "As shown in Fig. 1a and Fig. 1c, along a trailing front extended zonally across the southeastern United States, the isolated weak convective clouds formed in the late morning. Deep convective cells over Houston and Galveston bay areas developed in the afternoon with the increased solar heating and strengthened sea breeze circulation (Fig. 1b, d). The sea breeze circulation will be shown in a detail in the result section and it was among the typical summer day sea-breeze conditions (Kocen et al., 2013)." (Line 113-118)

Rosenfeld et al. (2014) has been added now. The ACPC website should be permanent. Also, there is no published paper to cite yet for this intercomparison.

Line 227: What is the "sea breeze wind intensity"? How was the sea breeze perturbation wind identified? This is especially important as the authors discuss Fig. 12 (lines 358 and following), where the sea breeze intensity is said to cause the transition to mixed phase cloud.

As we stated in Line 232-235: "the sea breeze wind intensity at a specific time is calculated by averaging the horizontal wind speeds below 1-km altitude along the black line UO in Fig. 4a."

Minor concerns
Line 64: CG lightning enhancements can also be caused by the presence of tall towers (e.g., Kingfield et al., 10.1002/2017gl073449).

We add this to Line65-67 as: "Kingfield et al. (2017) also found that cloud-to-ground lightning enhancements can also be caused by the presence of tall towers".

Line 157: not using met from domain 1? Not clear to me. Revisit after reading rest of paper. Maybe condense into a chart or truth table. I see the point that all domain 2 sims use same meteorology to control for the fact that different aerosols will affect the meteorology on domain two? But if the meteorology is processing the aerosols, how much is the spatiotemporal distribution of aerosols conditioned by the meteorology?

Since there are two Domain 1 simulations with different aerosol scenarios (anthropogenic emissions on and off), and the simulations accounted for the small-scale urban land and aerosol effects on meteorology already. If Domain 2 uses the meteorological fields from Domain 1, the initial/boundary meteorological conditions for the cases with high aerosol loadings (LandAero and No_Land) would be different from those with the low aerosol loadings (No_Aero and No_LandAero). To use the same meteorological fields to drive all simulations carried out over Domain 2 (including those with and without anthropogenic emissions), also to avoid using the forcing that already accounted for small-scale urban land and aerosol effects, we choose MERRA-2 for the initial and lateral boundary conditions for meteorological fields. We have revised the text in that paragraph to be clearer (Line 163-167). Either way (using Domain 1 or MERRA-2) has pros and cons and we chose the one fitting our purpose the best here.

Line 258: The reduction is not 45% from the larger LandAero value. Perhaps the authors used the wrong denominator? Reading approximate values from the graph to confirm the visual impression,
(6.75-4.8)/6.75 = 0.29
(6.75-4.8)/4.8 = 0.41
The claim is repeated on line 449.

Note we are talking about the enhancement of precipitation by urbanization, which should refer to the case without urbanization, i.e., the denominator should be No_LandAero. The calculation is (6.8-4.7)/4.7=0.45.

Line 327 and Fig. 10: Does this total water content include the vapor phase?
Water vapor is not included. This has been noted in the Figure caption now.

Line 331 and Fig. 10: The line marking the onset of the mixed phase cloud (first appearance of blue above the 0◦C level) is marked too late in panel b, and perhaps a bit in panel d, making the duration of mixed phase growth artificially short.

The 0◦C level is just an approximate mark there. If there is no continuous growing to deeper clouds after the water is a little above the 0◦C level, it would still be shallow clouds.

Line 393-4: the phrasing here makes for a confusing physical explanation; "the condensate loading effect is small as a result of enhanced condensational heating" but the condensational heating is directly related to the mass condensed, which should result in more condensate loading.

Sorry the order of the phrases in the original sentence was wrong. It has been revised as "…which is mainly because of the increased thermal buoyancy as a result of enhanced condensational heating since the offset effect of condensate loading is small (Fig. 16a) (Fig. 16c, blue lines)".