We sincerely thank the reviewers for their thoughtful comments, insightful questions, and constructive suggestions. Their feedback has greatly helped us clarify our ideas, strengthen our arguments, and improve the overall quality of the manuscript. Our responses are in blue text.

Reviewer 2

Major Comments

1. Non-invigoration aerosol-DCC interactions that could affect aerosol-ETH relationships are ignored. Aerosol-DCC interactions include direct effects on microphysics in addition to indirect effects on updraft strength. The paragraph starting on line 43 starts by referencing aerosol-DCC interactions in general but then the discussion that follows in the introduction focuses purely on updraft invigoration. This is problematic because aerosols can also directly affect microphysical properties (e.g., collision-coalescence, riming), which affects radar reflectivity and thus reflectivity echo top height. These direct effects may or may not be further associated with a change in updraft strength. To assume that updraft strength alone is the cause for changes in ETH assumes that changes in aerosols do not alter the reflectivity profile for a given cloud top. Furthermore, there is an assumption that the relationship between ETH and the true cloud top (the vertical gradient of reflectivity between the ETH and cloud top) does not change with changes in aerosols. It is not clear how valid those assumptions are. What evidence is there to suggest that ETH changes are primarily corresponding to changes in updraft strength?

We thank the reviewer for the insightful questions.

We believe that we have thoroughly described the aerosol effects on both DCC microphysics and dynamics in the introduction. This includes highlighting the high level of uncertainties in the extent of aerosol invigoration or enervation reported in the literature and the dependence of (indirect) aerosol effects on convective updrafts on the (direct) aerosol effects on DCC microphysics. The connection between direct and indirect aerosol effects is explicitly detailed through the microphysical pathways described within each invigoration mechanism outlined in the introduction.

It is indeed challenging to directly address aerosol effects on DCC microphysical properties and processes using the ARM TRACER observational datasets or any similar datasets from comparable campaigns. This is primarily because the TRACER field campaign alone did not include an in-situ cloud observational platform capable of observing quantities related to processes such as collision-coalescence and rimming.

We also acknowledge the uncertainty associated with using Echo Top Height (ETH) as a proxy for DCC intensity or maximum vertical velocity. The lack of direct measurements of convective vertical velocity remains a significant limitation, not only for this study but also for many previous observational studies. Due to such limitations, ETH has been routinely used as a proxy for DCC intensity or updraft strength in the literature (e.g., Liu and Zipser, 2013; Guo et al., 2018; Hu et al., 2019; Veals et al., 2022). One of the advantages of using ETH here provides a means to compare our findings with prior studies and builds on the existing body of knowledge.

As the reviewer also pointed out, a detailed investigation of the correlation between cloud microphysical properties/processes and ETH is indeed a limitation of studies that rely primarily on ground-based measurements or radar retrievals in the absence of in-situ observations of DCC microphysical properties when addressing aerosol-DCC interactions. This limitation certainly applies to the manuscript under review. Even in the case of having in-situ observations of DCC microphysical and updraft properties, the causal links/structures between these quantities and ETH would be difficult to establish without the use of more advanced causal inference models or modeling components.

To address reviewer's concern, we added these paragraphs to the manuscript:

1. to Line 108 where the ETH is introduced:

"Note that this assumption neglects the possibility that aerosols may directly influence cloud microphysical processes (e.g., collision-coalescence, riming), which could, in turn, affect radar reflectivity and, consequently, the DCC ETH. Quantifying such influence is challenging in the absence of in-situ observations of the cloud microphysical and dynamical properties (e.g., hydrometeor phase/size distribution, updraft velocity). The reliance on this proxy also stems from the lack of direct measurements of convective vertical velocity for DCCs investigated here, a significant limitation not only for this study but also for many previous observational studies. Nevertheless, using ETH as a proxy allows for comparison of our findings with prior studies, which is valuable for the scientific community and for providing modeling constraints on simulations of the aerosol-DCC interactions."

2. Additionally, we have added the following text to Section 4.6 where the limitations of the study are discussed:

"In the absence of in-situ observations of cloud microphysical properties, the current analysis cannot account for any 'direct' effects of aerosols on ETH or cloud depth through microphysical processes. Neither does the study investigate the microphysical pathways through which aerosols may cause the changes in ETH. Such examinations require in-situ observations and/or high-resolution model simulations, which forms a key limitation of any study aiming to explore aerosol-DCC interactions using remote sensing retrievals alone."

2. The g-computation model does not provide the causal direction, which still needs to be assumed, even if it is called a causal inference model. This assumption is made in the multiple linear regression model where the predicted convective property is assumed to follow from the predictors. The reasoning for this is that the meteorological and aerosol properties are defined prior to the convective cell properties, which makes sense, but this is similar to what has been done in some prior studies. Furthermore, this time offset

still doesn't ensure the assumed causal direction because there is a lot of atmospheric complexity that isn't being quantified that can affect the properties of the cells and atmosphere offset in space and time. Thus, describing this research as the first to show cause-effect is misleading. The methods do have unique aspects relative to past studies that can be highlighted but there is no reason to believe that the causal direction has been more discerned than in past studies.

We acknowledge that using g-computation, like those traditional methods such as linear regression, still requires assuming a causal link between aerosol number concentration and DCC intensity. However, our approach is superior to bivariate correlations (that do not account for basic confounding effects), because it attempts to control for confounding influences of known or potential confounders. If no causal link exists between aerosols and convection, the estimated causal effect using g-computation would approach zero.

We softened our language and rewrote the last paragraph:

"Nevertheless, this study demonstrates the potential of using a causal model to evaluate the effects of aerosols on DCC properties, providing new insights into aerosol-convection interactions through observations. It also represents a step forward in addressing the challenges of disentangling aerosol-meteorological co-variability in these interactions. Additionally, this causal framework shows promise for broader applications, offering a valuable tool for exploring complex scientific questions across various disciplines."

We removed the last sentence in the abstract.

3. It is not clear what value the g-computation model provides over the multiple linear regression. If the underlying model were a more complex nonlinear model, there would be some justification for it, but multiple linear regression is used. The multiple linear regression coefficients can be used to describe convective sensitivity to aerosols, giving the same results. Even with using the g-computation model, describing an aerosol effect as just the change in ETH without the corresponding change in aerosols, as is done throughout the paper, doesn't make much sense. It is the sensitivity, i.e., the change in ETH per change in aerosol concentration, that is most relevant with the underlying assumption that this is approximately linear, and this is simply the slope for the aerosol concentration predictor from the multiple linear regression model. What does the g-computation model provide that the regression cannot other than calling the model "causal machine learning"?

We thank the reviewer for the question.

We are not denying the fact that simple linear regression can be used for causal inference, but only under ideal circumstances where individual values are *randomly assigned* to groups. This condition, however, is not applicable to our observational study or similar types of studies in atmospheric science. Fundamentally, whether simple linear regression can infer causal relationship depends on how the data was collected. See the first few paragraphs of the introduction in Chatton et al. (2020) for more information.

In the case of aerosol-convection interactions in nature, it is *impossible* in the current world to randomly inject specific amounts of cloud condensation nuclei (CCN) into naturally developed convective clouds. The CCN concentration at a particular location on a given day can be a result of other factors, such as humidity, wind direction, and/or pre-existing convection. These hidden factors (confounders) could themselves be the true causes of changes in convective intensity. As long as CCN number concentrations cannot be randomly assigned, linear correlation coefficients cannot accurately infer causal effects, as they fail to account for basic confounding effects.

Our method extends the capability of an MLR because it offers an alternative by attempting to control for the confounding influence of known or potential confounders such as CAPE. We achieved this through a “controlled” or “forced” experiment, which, though less ideal than a fully randomized experiment, involves manipulating certain variables while others are held constant or randomized to minimize their confounding effects. In our case, we set the aerosol number concentration for every case to 1 for the polluted condition and 0 for the clean condition. Our identified confounders were kept constant in both scenarios.

Within the g-computation framework, technically, any predictive model can be used in the initial step. However, the choice of model often depends on the specific application and the number of available samples. For our study, we chose an MLR model over a different machine learning model due to the limited sample size. Note that this choice of the Q-model is NOT a direct advantage of g-computation itself; rather, the strength of this approach lies in its ability to control for confounding variables in its follow steps, which simple regression cannot achieve without the random assignment of aerosols into the atmosphere.

I am glad that the reviewer asked the question about the possibility to estimate the change in ETH per unit change in aerosol concentration. It actually is achievable using more sophisticated causal models. One such model that we have been experimenting with is called causal-curve, which allows the estimation of the causal effect of aerosols on ETH as a function of aerosol number concentration. However, this analysis is beyond the scope of the current study where we focused on estimating the *average* causal effects of aerosols. Additionally, the current manuscript is already quite lengthy, and including the description and results from the causal-curve model would make it more challenging for readers to follow. To maintain clarity and focus, we have decided to reserve this aspect for future studies.

4. Tests for multiple linear regression model accuracy and robustness are missing. For example, the predictor coefficients should have 95% confidence intervals computed. In addition, how well does the MLR predict the observed ETHs? What is its r2 value? The r2 is important as it shows how much of the ETH variance remains unexplained by the model, which is relevant for missing information that could still confound the relationships of ETH with the current predictors.

We thank the reviewer for the comments and suggestions. To address those, we included the *adjusted* R2 values for the selected exposure variables in the table below (also added in the supplemental material of the manuscript). The adjusted R2 was chosen over the R2 because it penalizes the inclusion of unnecessary independent variables. Specifically, as more predictors are added to the model, the adjusted R2 will increase only if the new variables significantly improve the model performance. In contrast, using R2, the value either remains the same or increases with the addition of new independent variables, regardless of whether the added variables significantly enhance the model performance.

As shown in the table below, the adjusted R2 values are generally below 0.5 and rarely increase even when all the potential confounders discussed in section 2.2 in the manuscript are included (according to a sensitivity test we performed). On one hand, this is, on some level, expected given the fact that these relationships are predominately nonlinear in nature. On the other hand, this result aligns with our previous statement in the manuscript: other confounding variables, beyond those included or discussed in the manuscript (section 4.6), likely exist but are unaccounted for. These variables may not have been measured or discovered to have a relationship with the outcome variables. Additionally, the small sample size may contribute to the low adjusted R2, as high variability in the outcome variable can artificially suppress it.

It is important to note that the purpose of these fitted MLR models is not to predict ETH but rather for exploration and hypothesis testing in this manuscript. Thus, the focus is on the other measures of the model robustness, making a relatively low adjusted R2 less critical. For example, in the original manuscript, we run model diagnostics (in supplemental material) to ensure the validity, reliability, and interpretability of the fitted MLR model which ensures the robustness of coefficients.

The 95% confidence intervals for the independent variables were also calculated and included in the table below. We notice that the values for the exposure variables sometimes cross 0, indicating the difficulty to conclude that the exposures have a clear and meaningful influence on the outcome. In our case, it suggests that aerosol exposure may not have a significant impact on DCC ETH. This finding is consistent with the relatively small or minimal causal effects shown for these scenarios in Figures 8 and 9.

To make this information clear, we made a few changes to the paper:

1. We modified lines 393 to 396: "We run model diagnostics to ensure the validity, reliability, and interpretability of the fitted MLR model as well as ensuring the robustness of coefficients. This is achieved by examining the key assumptions (i.e., linearity, homoscedasticity, normality, independence, and multicollinearity) of the MLR models as described in Text S4 in the supporting information. Overall, all valid scenarios presented in Section 3.2 satisfy these assumptions. In addition, we also calculated the adjusted R2 values, the 95% confidence intervals for each independent variable (Table S4 in the supporting information). The adjusted R2 values are generally

below 0.5 and rarely increase even when all the potential confounders discussed in section 2.2 are included. This result infers those other confounding variables, beyond those included or discussed here, likely exist but are not accounted for. These variables may not have been measured or discovered to have a relationship with the outcome variables which will be discussed in section 4.6. Additionally, the small sample size may contribute to the low adjusted R2, as high variability in the outcome variable can artificially suppress it."

2. We modified lines 470 to 473: "Interestingly, when conducting causal analysis on the "invalid" scenarios, the estimated average aerosol causal effects are mostly negative (Figure 8), highlighting the potential for contradictory results when a different exposure variable is used. Even for the "valid" scenarios, the significance of the estimated causal effects is challenged by the inconsistent 95% confidence intervals for the coefficients of the exposure variables in the fitted MLR models (Table S4 in the supporting information). Specifically, the 95% confidence intervals for the exposure variables sometimes cross 0, making it difficult to conclude that the exposures have a clear and meaningful influence on the outcome. This finding is consistent with the relatively small or minimal causal effects observed for these scenarios in Figures 8 and 9, which are likely to fall into the uncertainty range of the measurements or related to the sampling methods we used."

Table: The Adjusted R^2 values for the fitted MLR models and the 95% confidence intervals for the independent variables. The outcome variable is the 30 dBZ ETH and the CAPE is calculated when assuming the most unstable parcel would rise.

Exposure, Distance to the ARM site	Adjusted R^2	95% confidence intervals for the exposure variables	95% confidence intervals for CAPE	95% confidence intervals for ELR3
All cases				
Ncn, 40 km	0.2	[-0.01 1.53]	[0.03 0.78]	[-0.02 0.76]
Ncn, 50 km	0.2	[-0.07 1.37]	[-0.02 0.69]	[0.35 1.06]
Nufp, 20 km	0.1	**[0.66 3.76]**	[-0.67 0.66]	[-1.16 0.39]
Nufp, 30 km	0.2	**[0.16 2.02]**	[-0.11 0.76]	[-0.12 0.81]
Nufp, 40 km	0.2	**[0.11 1.66]**	[0.01 0.75]	[-0.04 0.73]
Nufp, 50 km	0.3	**[0.32 1.73]**	[-0.06 0.63]	[0.32 1.01]
Sea breeze cases only				
Nccn1, 30 km	0.2	[-0.01 4.06]	[-0.63 1.18]	[-0.65 1.32]

Nccn1, 40 km	0.2	[-0.43 2.73]	[-0.44 1.02]	[-0.32 1.10]
Nccn1, 50 km	0.3	[-0.68 1.61]	[-0.00 1.06]	[0.29 1.42]
Nccn08, 30 km	0.2	[-0.01 4.06]	[-0.63 1.18]	[-0.65 1.32]
Nccn08, 40 km	0.2	[-0.57 3.01]	[-0.47 1.04]	[-0.51 1.08]
Nccn08, 50 km	0.3	[-0.70 2.05]	[-0.14 1.01]	[0.14 1.40]
Nccn06, 30 km	0.2	[-0.01 4.06]	[-0.63 1.18]	[-0.65 1.32]
Nccn06, 40 km	0.2	[-0.29 2.95]	[-0.33 1.04]	[-0.52 1.02]
Nccn06, 50 km	0.4	**[0.29 2.66]**	[-0.15 0.87]	[0.01 1.15]
Nccn04, 30 km	0.3	**[0.74 5.30]**	[-0.63 1.05]	[-1.32 0.90]
Nccn04, 40 km	0.2	[-0.57 3.01]	[-0.47 1.04]	[-0.51 1.08]
Nccn04, 50 km	0.4	**[0.06 2.62]**	[-0.23 0.87]	[0.02 1.20]
Nccn02, 20 km	0.4	**[0.75 7.44]**	[-1.07 1.20]	[-2.76 0.75]
Nccn02, 30 km	0.3	**[0.37 4.27]**	[-0.32 1.36]	[-0.79 1.17]
Nccn02, 50 km	0.3	[-0.84 1.54]	[-0.02 1.06]	[0.29 1.46]
Nccn01, 30 km	0.3	**[0.37 4.27]**	[-0.32 1.36]	[-0.79 1.17]
Ncn, 40 km	0.2	[-0.41 2.41]	[-0.17 1.15]	[-0.30 1.11]
Ncn, 50 km	0.4	**[0.22 2.26]**	[-0.12 0.89]	[0.41 1.38]
Nufp, 50 km	0.4	**[0.05 2.16]**	[-0.10 0.93]	[0.33 1.35]

5. The argument for activation of ultrafine aerosols in updrafts leading to increases in ETHs lacks evidence. Activation of ultrafine particles seems highly unlikely given the high concentrations of larger aerosols for most of the samples assessed (Figure 7). Activation of the ultrafine particles would result in cloud droplet concentrations of a few thousand per cm3. Are there aircraft measurements (e.g., during ESCAPE) to support such high drop concentrations? Assuming a favorable composition for nucleation, what would the supersaturation need to be to activate particles at a certain size (e.g., 10 nm) given observed aerosol size distributions? This could be assessed in a parcel model to show if the argument being made is even physically possible.

We thank the reviewer for their comment. It was NOT our intent to argue that ultrafine aerosols will necessarily be activated in the updrafts. As the reviewer mentioned, there are no direct measurements of supersaturation or other indicators to definitively determine aerosol activation within the DCCs investigated during TRACER. Our intent was to hypothesize that, IF ultrafine aerosols are activated in the updrafts, they could influence the ETH of the convection. If this activation is physically impossible under the conditions studied, it would imply that aerosols do NOT significantly affect the ETH in this particular scenario.

We softened our language and added these sentences to Line 483: ”In addition, the high concentrations of larger aerosol particles observed under the assessed conditions (Figure 7) raise doubts about the likelihood of all ultrafine particles being activated. This challenges our hypothesis that aerosols may influence DCC ETH under the assumption that all ultrafine particles are activated.”

6. The diurnal cycle needs to be ruled out as a cause of the CN-ETH and UFP-ETH relationships. Over land, ultrafine aerosols often have a strong diurnal cycle just as deep convection does, which can affect relationships between the two. Accumulation mode aerosols often have a much weaker diurnal cycle, which is potentially a hypothesis for why one wouldn't get robust CCN relationships but robust CN relationships with ETHs. For example, Fast et al. (2024) shows this for the CACTI campaign. This occurs because new particle formation processes over land operate during the daytime. What are the typical changes in ETH and predictor variables including CN and CCN over the diurnal cycle? Do CN and ETH variables both peak in the later afternoon? If the hour of day is controlled for, does that affect the aerosol-ETH relationships?

The reviewer raised an excellent point, and we agree that it is important to investigate whether the diurnal cycle influences the causal relationships between Ncn-ETH and Nufp-ETH. In response, we included the timing of convection initiation as an additional confounding variable in the g-computation model along with the original ones we have selected in the original manuscript.

As an example (shown in the figure below), when using 30-dBZ ETH as the outcome variable and the most unstable parcel for calculating the convective indices, the average causal effects showed very similar results. They only differ by 0.1 km compared to the scenario presented in Figure 8, where the diurnal cycle was not controlled for.

We added these sentences to Line 476 in the manuscript: “We also conducted a sensitivity test to examine whether the diurnal cycle affects the causal relationships between aerosol properties and ETH. The results indicate that the average causal effects are only 0.1 km lower than those presented in Figure 8, where the diurnal cycle was not controlled for. This suggests that the diurnal cycle has a limited influence on the aerosol causal effects on ETH under the specific environmental conditions of this study.”

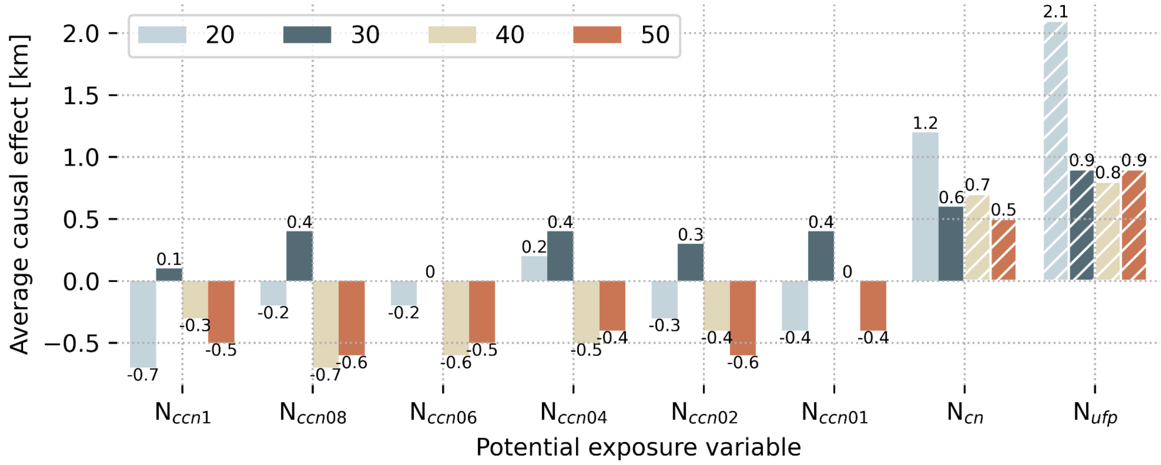

Figure. Average causal effects on 30-dBZ ETH estimated for each potential exposure variable after controlling for confounders which includes CAPE, ELR3, and the timing of the convection initiation. Different colors represent different maximum distances between measurements of environmental variables and DCC properties. The meteorological variables are calculated using ARM soundings (6-hr) when assuming the most-unstable parcel would rise to form a convection. The white hatch lines indicate "valid" results.

7. Relevance of sounding convective parameters at M1 for some situations needs further inquiry. Convective parameters like CAPE are not stable for 4-6 hours over land, and the study (Prein et al., 2022) used to support this claim on line 209 does not state that so far as I can tell. That study uses a limit of 4 hours difference between observed and simulated MCSs to match them, and MCSs are not the same as isolated convective clouds in atmospheric sensitivities. Other studies such as Nelson et al. (2021) show large changes in low level moisture on distances < 50 km and times of ~1 hour over some land convective regions. The statement after this on lines 209-211 that the M1 site is not heavily affected by maritime conditions is also confusing because the M1 site is close to Galveston Bay, and as noted in the study, a bay breeze often forms. Perhaps the bay air mass is similar to the continental air mass in terms of aerosol and thermodynamic properties, but I'm not sure that can be assumed. It may not be possible to easily assess these caveats, but they should at least be highlighted. Something that could be looked into though is whether the M1 surface measurements are relevant to air feeding cells at nighttime and/or after the bay/sea breezes have passed inland of the M1 site by examining stability at and through the boundary layer up to approximate cloud base to assess the likelihood of coupling to M1 site surface conditions.

We thank the reviewer for pointing this out and agree that the lack of comprehensive spatial coverage and high-temporal-frequency measurements during similar field campaigns is one of the limitations of this type of study. While we applied a few thresholds to select suitable measurements, these are by no means perfect. Instead, they serve as an approach to reduce uncertainty while maintaining a meaningful sample size for drawing robust conclusions.

If we were to restrict the analysis to measurements taken within one hour prior to the initiation of precipitation cores, we would be left with only nine cases (within 20 km of the site), significantly putting the robustness of the results in doubt. This highlights the need for longer-term field campaigns with improved measurement coverage to overcome these challenges, which we recommend as a priority for the broader research community.

Regarding the impact of sea breezes, we investigated their characteristics and effects in a recently published paper (Wang et al., 2024). Based on an analysis of surface variables, we observed that sea breezes have a long-lasting influence on temperature, humidity, and wind, persisting for several hours after the passage of the sea breeze front. An example showing this effect is presented in Figure 4 of Wang et al. (2024).

We rewrote lines 206 to 211: "The choice of a 6-hour time gap and a 50 km distance threshold as the upper limit represents a compromise between capturing representative environmental conditions and maintaining a sufficient sample size. We do want to emphasize the possibility of substantial temporal and spatial variability in the thermodynamic conditions around the M1 site. Local phenomena such as sea breeze, bay breeze, urban effects, and other factors may complicate the extent to which the environmental measurements at the M1 site represent the actual air mass injected into the DCCs (e.g., Rapp et al., 2024; Wang et al., 2024)."

8. More information on the spatiotemporal distribution of cells and cell properties is needed. Because of potentially substantial gradients in aerosol and thermodynamic properties given the coastal and large urban area, it would be ideal to plot the initiation locations and/or locations where the cell ETHs are maximized on maps for different ranges from the M1 site rather than the tracks in Figure 1 that don't provide much information. In addition, it would be helpful to map out cell properties like those in Figure 6 to see if there are spatial gradients in the properties with respect to the M1 site location.

We thank the reviewer for this suggestion. In response, we plotted the locations where the cell ETHs are maximized on maps for cells initiated within 50 km of the M1 site, as shown below (also included in the revised manuscript). The colors in these subplots indicate cell properties previously shown in Figure 6 of the original manuscript. Additionally, we have removed Figure 1 (cell tracks) from the original manuscript.

The spatial distribution of the cell locations reveals clustering along a line that is perpendicular to the coastline and northwest of the M1 site. One possible explanation for this pattern is the interaction between sea breeze, bay breeze, and urban heat island induced circulations, which likely create a favorable zone for cell initiation and/or collisions (Mejia et al., 2024). These cell collisions events may lead to larger cell areas, as observed in Figure c, and slightly longer lifetimes compared to cells located farther from this zone (Figure f). This is consistent with what we found in our recent submitted paper about cell colliding and merging behaviors (Hahn et al. in review). It is not surprising that these cells tend to initiate later in the day (Figure d), coinciding with the propagation of sea and bay breezes and their convergence with urban heat island-induced circulations (Wang et al., 2024). As the reviewer also mentioned, this spatiotemporal heterogeneity introduces complexity into our study, where we rely on point measurements of environmental variables which is a practical compromise in the absence of a more comprehensive measurement network over the region. This again highlights the needs for conducting long-term ground-based field campaigns with additional instruments that cover a larger spatial range.

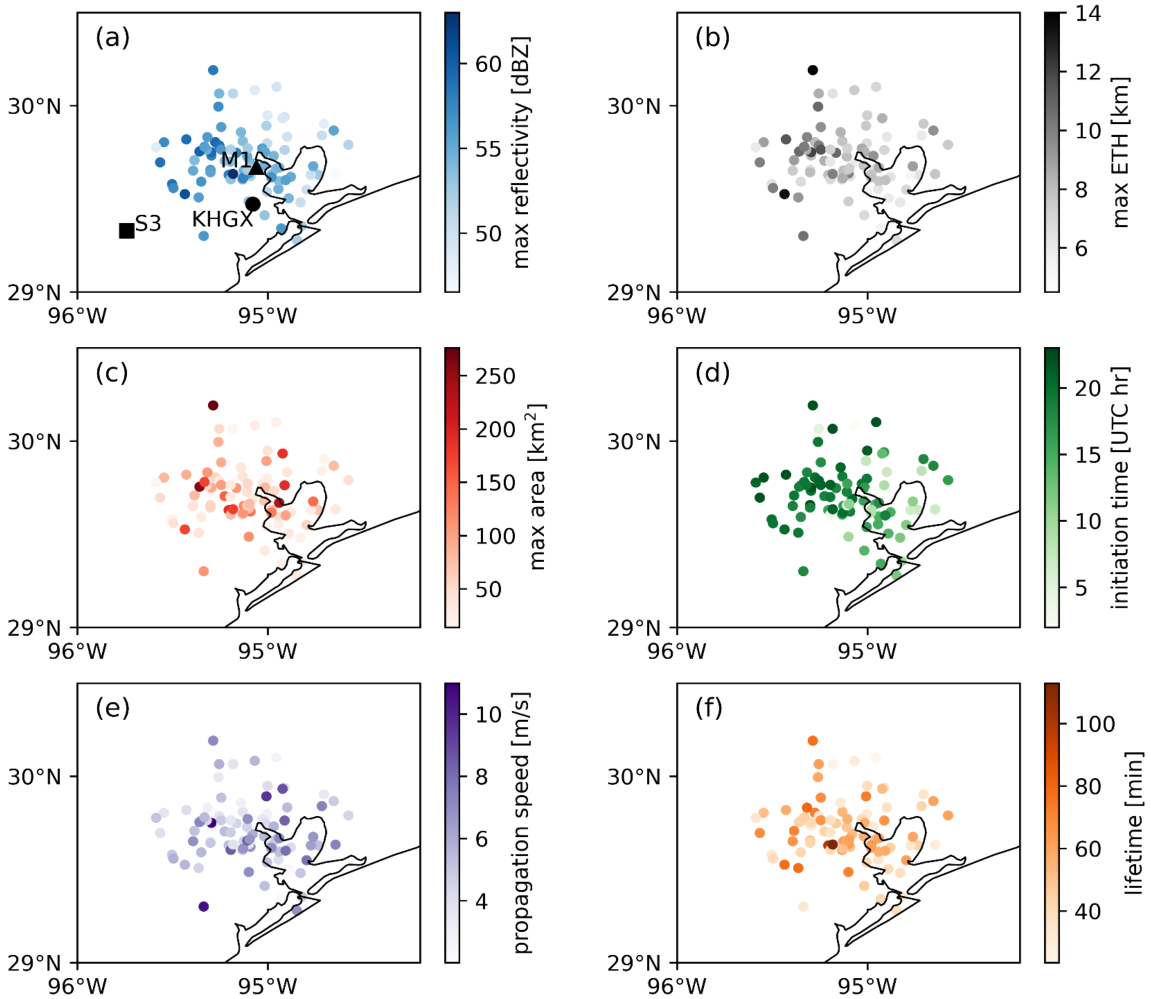

Figure: Dots indicate locations where the cell ETHs are maximized on maps for cells initiated within 50 km of the M1 site. The colors in these subplots indicate cell properties as shown in Figure 5.

We added these paragraphs to the revised manuscript (after Line 442): "Figure 6 illustrates the spatial distribution of DCC properties, showing a notable cluster along a line perpendicular to the coastline and northwest of the M1 site. This pattern can potentially be attributed to the interplay between sea breeze, bay breeze, and urban heat island-induced circulations, which may create a conducive environment for DCC initiation and/or collisions (Mejia et al., 2024). Such events appear to result in larger cell areas, as depicted in Figure 6c, and slightly longer lifetimes compared to cells located outside this zone (Figure 6f), consistent with findings by Hahn et al. (2024). Additionally, it is observed that these cells tend to initiate later in the day (Figure 6d), aligning with the timing of sea and bay breeze propagation and their convergence with urban heat island circulations in this region (Wang et al., 2024). Note that the spatiotemporal heterogeneity of these precipitation cores adds complexity to our study, as it relies on point measurements of environmental variables. While this approach is a practical solution given the absence of a comprehensive measurement network during TRACER, it highlights the need for long-term field campaigns with enhanced instrumentation to achieve better spatial coverage across regions with complex multiscale forcings."

9. Are ETH retrievals from level 2 NEXRAD data unbiased with range from the radar? Related to the previous comment, ETHs should be mapped with range from the radar to see if there are biases related to beam filling and gaps between elevation angles with range.

As shown in subplot (b) of the figure above, there is no clear trend indicating that ETH consistently increases or decreases with distance from the radar. Thus, we do not believe ETH is biased by radar range for cases studied here.

10. ACP recommends making processed data and code openly available in a FAIR-aligned reliable public repository to support study reproducibility. It is likely not possible to reproduce the methodology with only links to TINT and raw datasets given the information provided in the study.

We uploaded the post-processed data and codes for running g-computation to Zenodo and added a link to the revised manuscript.

1. WANG, D. (2024). Post-processed data for "Causal Analysis of Aerosol Impacts on Isolated Deep Convection: Findings from TRACER in Houston-Galveston". Zenodo. https://doi.org/10.5281/zenodo.14298966
2. WANG, D. (2024). Codes for "Causal Analysis of Aerosol Impacts on Isolated Deep Convection: Findings from TRACER in Houston-Galveston". Zenodo. https://doi.org/10.5281/zenodo.14299094

Minor Comments

1. Line 7: Only a single model predicts a significant relationship between an aerosol concentration and convective core area, which is 0.8% CCN within 30 km of the M1 site (Figure 10). The other 31 models are not significant. That seems pretty random, particularly since some models switch sign with changes in range within M1, and not enough to support this statement in the abstract that greater aerosol levels correspond to increased convective core area.

We removed the area statement in the abstract.

2. Lines 31-33: This is an odd motivation since ERFaci uncertainty is currently mostly attributed to non-deep convective clouds that are not the focus of this study.

We change the sentences to: "Aerosol-cloud interactions in DCCs are among the most complex and challenging processes to simulate accurately. This difficulty was evidenced in a recent model intercomparison project (MIP) conducted by the Deep Convective Working Groups of the Aerosols, Cloud, Precipitation and Convection (ACPC) initiative (Marinescu et al., 2021)."

3. Discussion of leading invigoration mechanisms in introduction: Semi-direct effects by aerosols that alter atmospheric thermodynamic stability should also be included.

We thank the reviewer for the suggestion and agree that a discussion of the semi-direct effect could provide additional context. However, considering that the introduction and manuscript are already quite long, and our work focuses specifically on the evidence of invigoration and enervation, we prefer to limit the discussion to these effects. Moreover, investigating the semi-direct effect would be equally challenging for us due to the lack of information on the extent of aerosol mixing into clouds and its contribution to this effect, as well as the absence of direct cloud microphysical measurements required to study it comprehensively.

4. Lines 60-63: Some of the studies cited here are not simply questioning the importance of invigoration mechanisms relative to other forcings but showing that there is a spectrum of enervation to invigoration possible, thus suggesting that referring to the mechanisms only in terms of invigoration is misleading.

We rewrite these sentences: "Despite a range of hypothetical mechanisms for aerosol-DCC invigoration, recent studies continue to challenge these theories, revealing a spectrum ranging from enervation to invigoration (e.g., Grabowski and Morrison, 2020; Igel and van den Heever, 2021; Dagan, 2022; Romps et al., 2023; Peters et al., 2023)."

5. Lines 75-76: Though individual modeling studies have quantified aerosol effects, it is important to note that there is still disagreement between these studies, even in the sign of effects, because models and the methods for analyzing them (e.g., discussion in Varble et al., 2023).

We added these sentences to Line 76: "Though individual modeling studies have quantified aerosol effects on DCCs, it is important to note that there remains significant disagreement between these studies, even in the sign of effects, largely due to variations in model configurations and the methods used to analyze them (Varble et al., 2023)."

6. It isn't clear how updraft strength is being defined. Is this referring to updraft mass flux, average vertical wind speed, or maximum vertical wind speed?

In this context, we are referring to the maximum vertical velocity in convective regions. We changed this sentence to "The maximum height of these cores can serve as a proxy for the maximum updraft velocity…"

7. Lines 124-128: Not tracking cells when max 2-km Z < 40 dBZ leaves out more than non-precipitating stages as suggested here. It also leaves out lightly precipitating periods.

We rewrote the sentence: "In other words, the tracked lifetime of the cores excludes the initiation stage of non-precipitating cumulus clouds, the dissipation stage of non-precipitating anvil clouds, and the lightly precipitating periods during either stage."

8. For the meteorological variables, there is almost an unlimited number that could potentially be relevant and tested. Were different shear layers other than 0-5 km tested? Was mid-level RH tested (separate from the boundary layer)?

We did a sensitivity test by adding mid-level RH (3-6 km) and high-level wind shear (5-10 km) as confounding variables alongside CAPE and ELR3 to the g-computation model. We found a decreased aerosol average causal impact on 30 dBZ ETH by less than 0.1 km, which is minimal.

9. What assumptions are made for the lifted parcel calculations (LCL, LNB, CAPE)? Is liquid pseudoadiabatic or reversible ascent assumed?

We added one sentence to line 160: "Note that, in the calculations, we assume that the parcel undergoes undiluted ascent in a pseudo-adiabatic process (neglecting hydrometeor loading)."

10. Line 187: CCN at various supersaturations does not have a temporal resolution of 1 minute or less as stated here. The supersaturation is varied over the course of about an hour usually so there is 1 value at each supersaturation every ~hour or so.

We rewrote this sentence: ”The Ncn and Nufp were measured at a temporal resolution of 1 minute, Nccn at various SSs had two measurements per hour, and radiosondes, used to derive meteorological parameters, were launched four to seven times per day.”

11. Lines 194-195: A t-test may not be valid here if the aerosol distributions are skewed.

A t-test is valid if the sample size (n) is large enough. The general rule of thumb is n>30. According to the Central Limit Theorem, the *mean* of any distribution is approximately normally distributed when the sample size is sufficiently large. In other words, even if the data itself is not normally distributed, the mean of the data *is*. If we were to repeat the experiment (TRACER field campaign) a hundred times and plot the sample *means*, the resulting distribution would be approximately normal.

12. How are DCC tracking results averaged? Does each DCC have a single value for a variable like ETH and then all of the ETHs are averaged together?

Yes, we took the maximum ETH throughout a tracked DCC lifetime (so one ETH for one DCC), then we averaged these ETHs to represent the mean ETH of these qualified DCCs for each corresponding sounding.

We added these sentences to line 204: “More specifically, in terms of ETH, we identify the maximum ETH throughout a tracked DCC lifetime (one ETH for one DCC), then we average these ETHs to represent the mean ETH of these qualified DCCs."

13. Lines 234-235: I don’t follow the argument for why large-scale ascent needs to be avoided, though I can see why MCSs would want to be avoided. Is that the primary reason for avoiding certain large-scale meteorological conditions?

Yes, mostly. We want to eliminate large-scale, dynamically-driven convective clouds since the aerosol effect may be overwhelmed by meteorological forcing. These situations often also exhibit strong CAPE, in which the aerosol effects are found to be difficult to detect (Storer et al., 2010). Additionally, different types of convection (organized vs. isolated) may respond to aerosol loading differently (Chakraborty et al., 2016). Therefore, we want to focus our study only on isolated convective clouds that are initiated in a similar weak large-scale forcing environment, where the aerosol effect may be more identifiable.

We added this sentence to line 235: “This choice serves to mitigate the potential influence of large-scale ascent on the evolution of DCCs. In other words, we aim to exclude large-scale, dynamically-driven convective clouds, such as mesoscale convective systems, since the aerosol effect may be overwhelmed by meteorological forcing (Chakraborty et al., 2016; Storer et al., 2010)."

14. Lines 273-275: Mesoscale deep convective systems are still buoyancy driven, so I don't understand what this sentence is trying to get across.

We removed this sentence and rewrote the previous one (see answers to question 13) to emphasize that we are only considering isolated convective clouds, which may be more conducive to observing aerosol impacts.

15. Figure 4: Why are values not filled in for the significant correlations less than 0.4? Also, I may have missed it, but are the aerosols in Figure 4 sampled around the same time as the soundings or are they sampled after the soundings?

Given the size of the correlation matrix, including correlation coefficients below 0.3 makes it difficult to visually extract the most important information. Since relationships with smaller correlation coefficients are not the focus of this study, we have only included values greater than 0.4 to improve the readability of the figure.

The post-sounding averaging (a 1-hour period following the radiosonde launch) for aerosols is shown in Figure 4. We added this information to the figure caption. We also plotted the correlations using the prior-rain averaging method (a 1-hour period before the rain). The resulting correlation coefficient matrix is very similar to the one shown in Figure 4, and we have included it in the supplemental materials.

16. In some places, LWS is used and in others, shear is used. It would be best to choose one or the other and be consistent throughout.

We modified the text and used LWS throughout the manuscript.

17. Line 342: Should "accuracy" be "robustness" here?

We changed it to "robustness" in the revised manuscript.

18. Lines 364-365: Including some critical meteorological quantities supports this assumption, but I wouldn't say that it is necessarily sufficient. That is hard to know without an in-depth study of possible confounders.

The reviewer is correct. However, as much as we would like to include all possible confounders, we often need to consider the balance between the number of samples and the number of confounders. An exhaustive list of confounders is ideal, but it may come at the cost of model accuracy and stability when we work with a limited sample size. This is a challenge the community faces today when working with observational data from shorter-duration campaigns, especially when certain data streams are only available at a single location.

We changed the sentence to: "Critical quantities known to influence ETH, such as CAPE and LWS, are explicitly included or discussed, to a large extent, supporting this assumption."

19. Lines 388-389: I don't follow the argument of multi-collinearity supporting standardization. Isn't the reason for standardization stated on lines 390-392?

Yes, it does not directly prevent multicollinearity but can prevent the numerical instability in computations that might arise when multicollinearity exists. We removed that sentence from the manuscript to avoid confusion. We did a test for multicollinearity by assessing the Variance Inflation Factor (VIF), and for all the models, the values are around 1, showing no multicollinearity. This information was included in the original supplemental material.

20. Lines 463-464: There is not enough evidence to make this statement that Ncn and Nupf are causing higher ETH via their activation.

We would like to direct the reviewer to our responses to major comments #5.

21. Line 494: I disagree that a causal link was demonstrated. The only supporting cause is that the aerosols are sampled prior to cells in time, but there is no evidence to show the causal mechanisms, and there are potentially other confounders not accounted for (see major comments).

We would like to direct the reviewer to our responses to major comments #2 and #3.

22. Lines 536-540: It's true that uncertainty renders the max reflectivity results less robust, but the same argument can be made for how well 4-6 hourly soundings and aerosols at a single point represent conditions where cells are growing.

We emphasized this uncertainty source in the second paragraph located in Section 4.2 in the revised version.

23. Lines 603-605: I think this sentence can be clarified. Aerosol is not *robustly* associated with DCC max ETH (not its evolution) given the sampling in this study. That does not mean that it couldn't be if more samples were added.

We rewrote the sentence: "Only a small fraction (16%) of the SLR models are valid, indicating that, in the majority of cases, aerosol loading is not robustly associated with DCC maximum

ETH, suggesting insufficient effects of aerosols on DCC updraft velocity in these situations with the current sample sizes."

References

Fast, J. D., Varble, A. C., Mei, F., Pekour, M., Tomlinson, J., Zelenyuk, A., Sedlacek III, A. J., Zawadowicz, M., and Emmons, L. K.:, 2024 Large Spatiotemporal Variability in Aerosol Properties over Central Argentina during the CACTI Field Campaign, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-1349.

Nelson, T. C., J. Marquis, A. Varble, and K. Friedrich, 2021: Radiosonde Observations of Environments Supporting Deep Moist Convection Initiation during RELAMPAGO-CACTI. *Mon. Wea. Rev.*, 149, 289–309, https://doi.org/10.1175/MWR-D-20-0148.1

Citation: https://doi.org/10.5194/egusphere-2024-2436-RC2

We cited these papers in Section 4.3 in the revised manuscript.