Authors greatly appreciate all the valuable comments and suggestions from the reviewers. Line and figure numbers correspond to the ones in the revised manuscript, and texts that are modified are in red colors in the revised manuscript. A comprehensive read-through is done to correct for English/grammar structure.

Main changes in the revised manuscripts are:
-Figure 7 is removed based on a comment from RC2.
-Table 4 is added.
-Figs. 8 and 9 are modified using different color bars that are colorblind-friendly.
-Fig 10 includes additional run with WSM6 microphysical scheme.
-Section 4.2 is modified using three month of data in 2020.
-Appendix A is added.

RC1

Summary: This study presents a new method for estimating latent heating (LH) profiles from geostationary radiances, and compares the result with established methods that use NEXRAD ground-based radar and TRMM and GPM spaceborne radar. The methodology for estimating LH is similar to what is used for TRMM/GPM LH profiles and is based on a database of output from a convection-permitting resolution model. The authors find that the GOES-based LH estimates are similar to those obtained from NEXRAD and GPM, and produce similar (positive) impact on model forecasts when used in model initialization.

General comments: The use of geostationary data to estimate latent heating is interesting, and potentially valuable, as the Geo data provides much more extensive spatial and temporal coverage relative to NEXRAD and GPM. I think this manuscript is publishable, but needs to be supported with quite a bit more explanation of the tools and datasets used, and also should contain additional context and caveats. I would like the authors to consider the following general recommendations.

1. Any LH estimate from remote sensing is by nature indirect - the observation is of the result of a process that involved LH, not of the LH itself. For example, there can only be hydrometeors for the radar to observe after the condensation process has already happened. A change in the cloud top brightness temperature can only happen after the air has arrived at the top of the storm (having already gone through the condensation process). Please comment on this - I think there is a significant unanswered question that relates to the time and space disconnect between an observation of the result of LH and the LH itself.

A paragraph is added in lines 116-123.

2. Convection is identified using time sequences of GOES imagery, yet the LH profiles are binned by the magnitude of the cloud top brightness temperature. It seems to me that an interesting and

more direct comparison could have been made between the simulated LH and the simulated time-difference brightness temperature. Please discuss.

Authors initially considered time-difference of brightness temperature for the same reason, but decided to bin only with the cloud top brightness temperature as the initial step in this paper for several reasons. Since clouds move over time, calculating change in brightness temperature per pixel can include errors due to cloud advection. In such cases, LH profile had to be assigned per cloud, and we thought that assigning the profiles to individual clouds rather than pixels can make profile inconsistent with the cloud top temperature for each pixel. Another concern related to using time-difference of brightness temperature is in case of mature convective clouds. When clouds reach tropopause, the decrease in temperature is rather small or not observed, and thus, the profiles will look similar anyways. Therefore, it remained as future study. This is discussed in lines 353-365.

3. An obvious point of concern in any model-based lookup table is the model construction and configuration. A 3 km horizontal grid spacing is barely convection permitting, and most simulations of deep convection meant for scientific analysis are now conducted at grid spacings of 1 km or less (most often smaller than 250 meters). Studies comparing simulations with sub-1-km grid spacing with those run at ~3 km have consistently shown that updrafts in 3 km grid spacing simulations are too wide and often too strong, and that the latent heating distribution is shifted higher in the coarse resolution runs relative to the fine resolution runs. In addition, studies have also shown that the LH position and magnitude are very sensitive to the details of the cloud microphysical parameterization. I have a number of questions that I would like the authors to address:

- Why did you not run the WRF model at finer grid spacing? Even if this was not computationally feasible, at least one simulation should be run at fine grid spacing and the LH characteristics compared to assess sensitivity.

The reason for using 3km resolution is to match with spatial resolution of HRRR model. As the reviewer pointed out, the magnitude of latent heating can vary depending on the spatial resolution. Thus, while we understand reviewer's concern in using rather coarser resolution, but the purpose of this study is to use retrieved latent heating to initiate convection at 3km resolution model to remain consistent with the operational HRRR model. In order to keep consistency in magnitude of latent heating between retrieved and modeled ones, we used 3km resolution. The discussion is added in lines 317-320.

- What was the sensitivity of the simulated LH to choice of microphysics? I do not expect a detailed study of this, but as with the previous question, one could imagine running companion simulations of the same case, one with Thompson microphysics and another with (for example) Morrison or WSM6. This would at least provide a first order estimate of the sensitivity.

Authors appreciate the reviewer for raising the great point. Comparing results using different microphysical scheme was part of a future study, but we added one simulation result using WSM6 scheme (Figure 10 in the revised manuscript and line 660-662) to address this point.

4. There was not enough detail provided about the simulation database itself. I was missing the following, which the authors should provide:

More detailed information about the simulation is added in lines 313-320, and Table 2 is modified. It addresses the points below.

- What were the lengths of the simulations (in time)?

It was run for several hours when there was convective activity in the scene.

- What were the geographic domains?

It was mentioned that the geographic domain was over CONUS.

- Which data was used for initial and boundary conditions?

HRRR analysis data are used as initial and boundary conditions

- What was the model output time frequency?

Model is produced every minute, but data every 10 minute are used to create the lookup table.

- How many vertical layers were used? (this can have as large or larger effect on the convection than the horizontal grid spacing)

50 is mentioned in table 2.

- How were the simulations validated? How did the authors ensure they provided a reasonably realistic depiction of storm structure?

It was validated subjectively by comparing simulated brightness temperature and observed brightness temperature.

- Did the simulations span a range of convective types (size, longevity, mode of organization)?

The model was run from the beginning of convective activity to the end over the scene, but the lookup table was not divided into different type of convection as it is hard to distinguish different convective types from observation.

5. The various LH estimates seem to reflect different sources of information on LH. For example, NEXRAD is sensitive to large hydrometeors and primarily obtains information from the lower portions of the troposphere. As such, one would expect the NEXRAD estimates to be biased toward the lower portion of the storm and miss LH in the middle and upper portions. TRMM/GPM radars operate at a shorter wavelength - they will see more of the smaller hydrometeors higher in the storm and may miss some of the heaviest rainfall due to attenuation). One would thusu expect their information to come from the middle portions of the storm but perhaps miss the very lowest and highest layers due to missing detection of heavy rain and small cloud particles. Geostationary data only sees the change in cloud top properties - it's not clear which portion of the storm produces the change at cloud top, but it is likely weighted toward the middle and upper portion of the storm. I would like to see the authors comment on this, and to perhaps discuss how the three sources might be merged in those instances where all three view the same place and time.

Authors agree with the reviewer that there's a potential to merge three products because each observation sees different part of convection. However, the goal of this study is to use LH profiles for short-term forecasts, and DPR product is not suitable for this purpose due to coarse temporal resolution and narrow swath. Yet information from NEXRAD and GOES can be merged through a lookup table in Appendix A which is newly added. Cloud top information from GOES will determine the vertical profile, and the overall intensity can be adjusted using NEXRAD composite reflectivity through the lookup table in Appendix A.

6. There were no caveats listed in the conclusions - one would expect that there are places and times where the GOES data might provide a more reliable estimate of LH and others where these estimates will have larger errors. What are these? Also, there was no mention of future work - what is next? This should also be discussed in the conclusions section.

Lines 690-691 and 700-705 in conclusion are modified to reflect this comment.

Specific comments:

1. June 2017 (the case used to assess impact) is within the time frame used to run the WRF simulations that form the database of profiles. In testing a database-based method, it is common to test on a case that lies outside of the training dataset. I wonder what the results would look like if you compared the estimates for a month from 2019?

We agree that that using June 2017 data is not independent for the testing. We replaced the analysis using summer of 2020. Section 4.2 is modified based on the new analysis.

2. It was clear that there are discrepancies between the NEXRAD and GOES detections of convection. It would be interesting to see statistics on how often these discrepancies occurred.

Statistics of GOES detection accuracy compared to MRMS PrecipFlag product which is different than using 28dBZ but uses NEXRAD radar reflectivity to assign precipitation type is provided in lines 293-294. One third of the three-month data had large discrepancies in detected area (the number of convective grid points from GOES-16 exceeds five times more than the number of convective grid points from NEXRAD and vice versa), and it is added in lines 560-562.

3. The scatter in the plot comparing GOES vs NEXRAD LH in Fig 8 is very large. It is surprising that the correlation was ~0.8. I wonder if the relationship is more obust for smaller LH values than for larger? I suggest using log-log axes for Fig 8 to better be able to examine the smaller LH values.

Thank you for the suggestion. Figure 8 (7 in the revised manuscript) is changed with log-log axes.

4. The phrasing in lines 560-560 on page 16 is confusing - it makes it sound like you are replacing the observed LH with the LH from the model. I think that what you are doing is inserting the observed LH into the model (replacing the modeled LH), right?

Yes, it is rephrased in line 582-583.

5. Follow-up question - are you inserting the observation-estimated LH *profile*? If so, the NEXRAD profile would be bottom heavy while the GOES profile would be top-heavy, right? This would explain the precipitation differences, I would think... NEXRAD LH would produce warming lower in the troposphere, which should result in a much larger effect on buoyancy, relative to GOES.

Yes, the reviewer is correct that we are inserting the vertical profile of LH, and NEXRAD LH would be bottom heavy while GOES would be top heavy. Lines 658-660 are added.

6. While the magnitudes are similar between NEXRAD and GOES estimates of LH, the position of the peak in the vertical matters quite a bit for large scale dynamics. How has this discrepancy been addressed in the literature? Is it assumed that NEXRAD is biased low? Is CSH (and by extension GOES) biased high?

To author's knowledge, there has not been a literature that compares LH from NEXRAD used in HRRR model with LH retrieved from CSH since NEXRAD LH is simply developed to initiate convection in the operational model, and it has not been used to study Impacts of LH in large scale dynamics. Such comparison can be future study.