Response to the editor comments

1) The inclusion of Cb-Tram is valuable, but it is important to be clear about the limitations. Specifically, Cb-Tram tends to lose cores when shielded by anvils, and from the description in the manuscript, it is not clear if this could become an issue. It might explain the different spatial distribution as compared to the model simulations, as active cores tend to be detectable at the anvil edge and the surroundings, not in the anvil centre.

Reply: Thank you for making this point. The Cb-TRAM detection is stacked in 3 stages: 1) detection of convective initiation; 2) detection of rapid cooling; 3) detection of mature thunderstorm cells. As the editor points out, stage 1 and stage 2 are usually not detectable when an anvil cloud is present from a previous/contiguous convective cell. Stage 3 instead aims at active cell centres that are characterized by – among others -- a large local inhomogeneity of the HRV channel reflectivity. As soon as the convective cell penetrates the pre-existing anvil of another cell (or any pre-existing high ice cloud) this newer active convective centre can be detected with Cb-TRAM even if it is not at the anvil edge or in the surroundings of the anvil. Thus, we agree that this issue can affect the comparison, but only to some extent. In fact, it might be that in the central part of Switzerland, in the southern Alps where brightness temperatures in the 10.8-micron channel also show very low values (Fig. 3a), thus indicating thick ice clouds, some convective cells cannot be detected. On the other hand, the granularity of the overshooting tops in Northern Italy in Fig. 5b-e is different from the one in the satellite observations, especially for the TKE simulations, with larger black dots in the observations. We have adapted the text accordingly.

2) Due to the stochastic nature of convection, it is not entirely clear how attributable some of the simulated inter-configuration differences are to the difference in the models. For example, are the differences in Fig. 5 or Fig. 6 statistically robust?

Reply: We thank the editor for this important comment. To assess the robustness of the reported inter-configuration differences despite the stochastic nature of convection, we performed paired Student's t-tests and Wilcoxon signed-rank tests on the overshooting convection metrics derived from the simulations discussed in Figs. 5 and 6. The analysis indicates that the reported inter-configuration differences are statistically significant ($p < 0.05$), suggesting that the differences are robust and not solely attributable to random convective variability. This information has been added to the revised manuscript.