

This is an interesting study that represents important research that should be of interest to readers of this journal. The results of this research will add to a growing list of studies dealing with the important topic of convective transport of reactive radical precursors to the UT and LS. This paper is within the scope of ACP and meets the scientific quality of this journal. However, after having said this, this reviewer has some major concerns regarding the results as presented and would like to see more supporting evidence in the areas discussed below. Accordingly, this reviewer believes this paper should be accepted for publication after some **major revisions** are made.

First, CO and CH₄ may not be the best species to account for entrainment/dilution during both vertical transport as well as horizontal transport out of the anvil. As shown, the contrast in mixing ratios between the convective outflow (OF), the free troposphere, and the boundary layer inflow (IF) are small and thus information on the entrainment rates may not be reliable. Measurement of other species with much more dynamic vertical profiles like various hydrocarbons would be preferable in determining entrainment rates. Can the authors employ their canister measurements of hydrocarbons like *i/n* butane and *i/n* pentane and their ratios to address this as well as to further verify that the outflow is coherently related to the inflow?

Nevertheless, given the near unity CO OF/IF ratio of 0.93, one cannot assume that entrainment does not exist. It's hard to imagine there is no entrainment dilution during convective transport from the BL to ~ 10 km, followed by no entrainment dilution of UT background air during the OF. Can these assumptions be wrong? Very similar CO OF/IF ratios were measured during DC3, and yet entrainment was still found to be important. In fact, using your CO IF and OF values in the altitude dependent entrainment model of Fried et al. [2016] with estimates of your background CO values in each 1km altitude bin, I get a net entrainment rate of 3.6%/km. Using this entrainment rate, I calculate that the HCHO value at the storm core should be ~ 2.054 ppbv, which should then be used to compare with your 1.45 ppb OF value, which has to be further modified for production and destruction. Even though my calculations are crude (mixing together entrainment from vertical transport and horizontal outflow in the anvil), they serve to illustrate that dilution of background air should not be ignored.

However, my second and biggest concern relates to the appropriate IF values to use for this analysis. The authors attempt to address this in their discussion section on page 11, by stating that although it is not possible to unambiguously identify the inflow area, their HCHO and H₂O₂ boundary layer mixing ratios near Dresden are similar to other boundary layer observations during HOOVER II and should thus be representative of the convective IF values. However, as shown by Fried et al. [2016] if this assumption does not hold, then one can obtain both higher (~80%) and lower (~20%) HCHO scavenging efficiencies (SEs), depending upon the circumstances. In particular, large changes in boundary layer isoprene mixing ratios, as one example, can cause erroneous HCHO SE determinations unless one can be certain that the inflow is related to the outflow. The authors need to provide more convincing evidence to this effect in order to reconcile if their much lower SEs with other studies for both HCHO and H₂O₂ are caused by this or by differences in storm dynamics and microphysics. Do the authors have measurements of isoprene and/or any other sources of HCHO in the boundary layer to help this argument? Simply invoking differences in ice retention factors cannot explain the lower HCHO and H₂O₂ SEs in the present

study with the DC3 results. As discussed in the DC3 studies in the case of HCHO, large changes in calculated HCHO ice retention factors from 0.25 (and most recently 0.15) to 1.0 all result in calculated HCHO SEs near 100%. It is only when HCHO is completely degassed from ice (ice retention of 0) can the modeled results reproduce the $\sim 50\%$ SE results deduced from measurements. There is no scenario where changing the ice retention factor produced lower SE results. Likewise, for H₂O₂ Bela et al. [2016] and Barth et al. [2016] in their simulations found that with H₂O₂ ice retentions ≥ 0.25 , the H₂O₂ SE approached 100%, and with ice retentions of 0, one obtains H₂O₂ SEs of $80\% \pm 12\%$. Again, ice retention factors cannot explain the differences.

Therefore, the authors need to seek other explanations for the differences with DC3 results. Can differences in storm dynamics and microphysics be the cause or can differences in IF and OF airmasses be the cause? In the case of the former, the authors should try and contrast differences between the studied storms here and DC3. To eliminate the latter, the authors need to provide more convincing evidence that the IF is related to the OF. In addition, the authors need to raise the possibility that in contrast to most DC3 measurements acquired in the anvil, the measurements here were obtained in clear air and this may allow the hydrometeors a chance to evaporate, thereby degassing the dissolved species resulting in low SEs.