This paper introduces the storage-release events in the mass-balance air-craft top-down methods. Using Global Environmental Multiscale-Modeling Air-Quality and CHemistry (GEM-MACH) model, the authors present the storage-release events (The author shows the transient storage of the emitted mass within box volume and its later release) can contribute to the overestimation or underestimation of emission rate, up to (-29 – +156%). When certain meteorological conditions are set up (e.g. when temporal and spatial variations in meteorological conditions, so the source emission rate result in mass imbalance through box top and lateral walls and the deposition (storage) to ground.

They also introduced three productive parameters to forecast the storage-release events, such as 1) Richardson number (atmospheric stability/shear), 2) Plume Shift in the transport direction, and 3) Weighted upwind to downwind concentration ratio, emphasizing the potential practical use for both aircraft and model-based forecasts of their parameters.

The paper reads well and designs well based on the thorough consideration of the current issues associated with the mass-balance approach, and potential sources of the uncertainties.

However, there are a few points that would have been considered.

- First, while the authors kept mentioned that the emission estimate can be highly affected by the meteorological conditions, they didn’t show the actual wind conditions for each case. Furthermore, the authors also didn’t consider how the emission estimate can be significantly changed depending on how to handling wind data [Turnbull et al.,
The fluxes can be also varied depending on what planetary boundary layer height (PBLH) is used. Model PBLH is known to be incorrect, so please specify how you handle this PBLH for flux calculation in your GEM-MACH data.

- Second, the storage and release terms seem to be determined by the rate of compound mixing ratio change with time and the temporal variations in air density with time. The assumption underneath here is that the gas concentration change with time during flux measurement (normally 30-2 hours), and the values are significant enough to change the emission rate. But this may be scale-dependent. For example, for the local scale, the variation of the mixing ratio can be small. It is great for authors to suggest “universal” methods to consider all the factors to influence the emission estimate in all possible meteorological conditions. But does this “storage” term really include the surface-emission flux which can be highly varied with extrapolation methods when we use aircraft data in practice [Gordon et al., 2015]?

- Third, the actual data processing is not well described. While mathematical terms are well described, it is questionable whether this can help interpreting field data since all the terms used in this study are obtained directly from the model (e.g. surface deposition rate). Furthermore, the model data may not need to do additional interpolation/extrapolation as air-craft data does, so the uncertainty associated with interpolation can be lower. However, emission estimate methods based on mass balance using aircraft data often adapt the interpolation (e.g. kriging or exponential) and extrapolation, and the values are highly varying depending on what method you use. Since model data is gridded, and somewhat averaged data, data uncertainty is already included in model output. In this way, how can we trust the storage or surface deposition rate from the model is correct? Comparing the model input emission input with the data methods are good enough for trust model-based calculation of storage value? In addition, how do you know the storage and release terms can be obtained “along” the lateral walls of box-shape (i.e. closed shape) flight over measurement time, not throughout the bottom area of the box?

- Finally, the suggestion of multiple flights or several loops with one flight is great and sounds ideal. However, given aircraft constraints, is this realistically feasible for most cases? Some redesign will be necessary to satisfy this.

**Specific comments:**

- Line 84: “Storage-and Release” events sound new. Is this new terminology used in this study?
- Line 187: Good approach. However, can’t we just say that the “storage” term just represents the surface flux and entrainment through the top and bottom flux?
- Figure 7: This is a nice illustration. However, this storage term cannot be measured by aircraft if we only consider “along-track change”. How to validate the model storage term? Comparing MIE with estimates from three different methods are enough? I am not sure how applicable this method will be in a practical way.
- The mathematical formulas in general: The mathematical notations are useful, but ironically these are also very confusing as well. I liked the detailed description of the mathematical formulas, but I often lost and forgot what each abbreviation referred to. Could you please be more specific in the notation? Maybe adding more words in the notation will be helpful for readers to understand what each term refers to, keeping them from looking back at the previous pages.

**References**

