Review of “Seasonal characteristics of atmospheric formaldehyde (HCHO) in a coastal city of southeast China: Formation mechanism and photochemical effects,” Liu et al., ACP (2022)

Summary

This manuscript describes a set of ground-based observations of atmospheric composition at a coastal urban site in China. The primary analysis focus is formaldehyde (HCHO). Measurements are fed into a PMF model and a photochemical box model to estimate the sources of HCHO and the contributions of HCHO to radical chemistry and ozone production.

The reviewer has substantial concerns regarding the quality of HCHO observations, the interpretation of the PMF and box model results, and the general presentation of data and analysis. Many superfluous details are provided in the text. Text is highly descriptive without drawing out any obvious novel/new conclusions. This is a potentially useful contribution that hopefully will benefit from a hard critique. I recommend rejection with encouragement to resubmit.

General Comments

Regarding the HCHO analyzer described in Sect. 2.1: The reviewer was not able to locate any information about this analyzer on the internet, and there is no citation of literature regarding the design or performance of this instrument. The stated performance is 1 Hz, 50 pptv detection limit, 5% accuracy. This exceeds, by far, similar Hantzch-based instruments. For example, Glowania et al. (2021) report a 90-second time response, 300 pptv detection limit, and 8.6% accuracy (https://doi.org/10.5194/amt-14-4239-2021). Given that HCHO is central to this paper, additional documentation regarding calibration procedures and determination of potential artifacts is warranted.

Regarding interpretation of a highly-constrained model: Throughout the text, attention is given to the difference between HCHO production and loss rates (described as “net production rate” on L181). The model, however, is forced to measured HCHO. How well does the model predict HCHO if this constraint is turned off? If the model performs poorly, this calls into question the utility of the “net production rate” since the HCHO concentration does not match what would be predicted based on the modeled gross production rate.

PMF analysis: Several questions here.

1. Why are other species not included in PMF (CO, NOx, PAN)? In particular, CO should be a clear marker of vehicle exhaust.
2. Are the authors really suggesting that the HCHO associated with isoprene is directly emitted by the ecosystem? Is there any literature evidence of that? It seems more likely that this HCHO was produced by isoprene enroute to the site. Possibly without significant ozone production (e.g. from a nearby forest).
3. What is the real meaning of “secondary formation”? Again, it seems likely that the HCHO from those other sources is a mix of primary and secondary. It seems more accurate to call it “Ozone-associated HCHO.” This is a general shortcoming of using PMF to parse something like HCHO and it should be acknowledged and clarified.

Data and Code Availability: According to FAIR standards, the observations and box model code should be publicly available without having to request them from the author.

Specific Comments

L24: The method for determining HCHO contributions should be stated here.

L59 – 62: suggest deletion of this sentence.

L146: why is the error factor 10% instead of actual measurement accuracy?

L166: How is the boundary layer height determined?

L173: updating constraints at hourly intervals is too coarse and likely leads to model artifacts due to step changes in photolysis and other parameters. 10 – 15 minute time steps are more appropriate for science-grade simulations.

Figure 2: There is little utility in showing atmospheric pressure and all 3 J's. You could remove the bottom panels and replace pressure in the top panels with shaded J(NO2).

L174: JNO2 is strongest in the visible, so applying a scaling factor from this variable alone may not capture variations in the UV (e.g. due to aerosol). How well does this JNO2 parameterization predict other measured J's, like JO1D or JHCHO?

L182: Why 20%? Are RIR values sensitive to this choice? Why not a smaller value (like 1%) so that radical resulting perturbations are locally linear?

L222: What other data supports "replenishment of HCHO primary emissions and accumulation of pollutants"?

L291: "under the intense solar radiation" is vague. Be quantitative. Based on Fig. 3, the J values are higher by 20% in autumn.

L29: The net production rate seems at odds with diurnal cycle of dHCHO/dt. At times, the observations show increasing HCHO when the model predicts loss, and vice versa. This is consistent with my second general comment about the model being over-constrained.

L301: CH3O2 comes from many precursors, so it is not quite fair to distinguish this from other RO2 precursors. This should be somehow stated or made clear, that the "RO + O2" bars in Fig. 5 are lower limits.

L312: "significantly higher" is not quantitative. Also, the loss rate might be higher because HCHO is higher. What is the difference in HCHO lifetimes between the two periods?

Sect. 3.4.1: This section does not describe the impacts of HCHO on atmospheric oxidation.

L421: Autumn OH is higher than any previous observations in China, at least among those cited. It is not fair to say the simulated HOx is "comparable." This is a lot of OH. Can any model comparisons be done to observations to substantiate it?

L430: Here, and elsewhere throughout the paper (L454, L501), are long lists of numbers that don't convey anything meaningful to the reader.

Figure 9: Figure S4 is potentially a more useful figure.

L449: If the model were truly constrained to ozone, there would be no decrease in ozone photolysis between these two runs. As alluded to in the General Comments, this is potentially an artifact of hourly time steps.

Technical Comments

English throughout would benefit from substantial copyediting.

Acronyms should only be defined after their first use.

Throughout the text, numbers are reported with too many significant figures. For example, 2.94 +/- 1.28 ppbv should be 2.9 +/- 1.3 ppbv. Also, it is often unclear what the averages and uncertainties/variabilities refer to (averaged of what period, at what time resolution).