Overall Feedback:

Zhao and co-authors investigated the summertime HCHO interannual variability in northern high latitudes. Using the GEOS-Chem model and satellite data, they highlighted that wildfire is the major driver in regions like Siberia, Alaska, and North Canada. Yet, biogenic emissions and methane oxidation predominantly drive HCHO variability in East Europe. They also introduce SIF as a potential indicator for biogenic emissions in these northern regions. However, there are specific concerns with the current manuscript that need to be addressed before it can be considered for acceptance.

These concerns are detailed below.

Major scientific concerns:

1. **Linear Relationship Assumption:**

Formaldehyde's dynamics are significantly influenced by secondary productions. Given this, is it appropriate to assume a linear relationship across various sources? How have you addressed the discord between the nonlinear chemistry and your linear assumptions? Moreover, how might this nonlinearity affect your conclusions?

2. **Scope of SIF Analysis:**

The SIF analysis is restricted to a range between 0 to 0.25. While the slope derived might be relevant for higher SIF values in lower latitudes (Figure 6), three of four domains in Figure 5 pertains to higher latitudes. This raises questions on the applicability of the derived relationship to VCD and emission data within these domains. How do you reconcile this?

3. **MEGAN Computation in GEOS-Chem:**

While MEGAN is computed online within GEOS-Chem, there is a significant divergence between online and offline MEGAN results, as noted in the discussion below. Since GEOS-Chem typically recommends having both online and offline modes enabled by default, could the exclusive use of online MEGAN have swayed your results?

Link: http://wiki.seas.harvard.edu/geos-chem/index.php/MEGAN_v2.1_plus_Guenther_2012_biogenic_emissions#Note_on_the_use_of_pre-computed_emission_factor_.28EF.29_maps_versus_EF_maps_computed_online

4. **Potential Biogenic Underestimation of GEOS-Chem:**

Selimovic et al. (2022) showed that GEOS-Chem underestimates OVOCs including methanol, formaldehyde, formic acid and acetic acid by a factor of 3 to 12 in arctic area. Among them, the negative model bias for methanol is attributed to outdated MEGAN. In light of this, is it conceivable that your simulated biogenic emissions are similarly underestimated? What implications could this have on your study's conclusions?

Reference: Selimovic, Vanessa, et al. "Atmospheric biogenic volatile organic compounds in the Alaskan Arctic tundra: constraints from measurements at Toolik Field Station." Atmospheric

Chemistry and Physics 22.21 (2022): 14037-14058. https://acp.copernicus.org/articles/22/14037/2022/acp-22-14037-2022.pdf

General writing issues:

1. **Adherence to Writing Guidances:**

Please ensure that your manuscript aligns with the ACP writing guidelines stipulated. For example, when referring to a section in the running text, utilize the abbreviation "Sect." followed by the respective number, unless the reference initiates a sentence. Thoroughly review the ACP writing guidance to ascertain that the manuscript conforms to all provided specifications.

2. **Quantitative Precision:**

There are too many vague words used in the paper. Please aim to bolster the precision and clarity of your assertions by providing quantitative data wherever feasible. Avoid ambiguous or generalized statements; instead, substantiate your claims with numbers. Not only for your abstract but the main texts as well.

3. **Citation Authenticity:**

It's paramount that you intimately familiarize yourself with the papers you cite. This ensures the citations support your arguments. Refrain from randomly citing papers merely to bolster the appearance of your statements. Authentic and relevant citations strengthen the validity of your work and demonstrate thorough research diligence. While I have not looked into every reference in your manuscript, I have identified some citations that appear to be misaligned or inappropriate.

4. **Conciseness in Writing:**

Many sections of your manuscript contain superfluous or verbose language. Aim for succinctness, ensuring that each word adds value to your statements. Make every word count.

Other comments for each section:

Abstract: Please write in a quantitative way.

Introduction:

Line 29-30: Consider providing a foundational and more general reference to support this statement, rather than solely referencing works from your co-authors.

Line 43: Avoid just using words such as “large”. Provide numbers.

Line 49: The word “predominant” is confusing. What do you mean? The most abundant? The most reactive? or both?

Line 53: Same for line 49.

Line 53-59: The connection or relevance of these points is unclear. Could you clarify the relationship?

Line 60: Given the focus on formaldehyde in the preceding and following lines, I guess you want to say the "Wildfire is another important source of formaldehyde"? If so, ensure each word has impact.

Line 63: It seems the biogenic and the background are still the main CH2O sources but fire may be the driven for interannual trend. Please double check your statements here.

Line 70: Why using SIF? Is that because other methods having limits? Please provide more context here.

Line 80-81: Do readers also need to know pros and cons of OMI and OMPS before understanding your work?

Observations and Model:

Sect. 2.1: Again, why using OMI and OMPS? What are their pros and cons? How does use these two datasets help each other?

Sect. 2.2:

Line 164-173: Have you done the spin up? Also, please use right reference of GEOS-Chem model. For example, in line 176, I believe van der Werf et al. (2017) is the one for GFED4s. Also you need to cite the GEOS-Chem doi with corresponding version. Please read through the web below in GC website.

https://geos-chem.seas.harvard.edu/geos-chem-narrative

http://wiki.seas.harvard.edu/geos-chem/index.php/GEOS-Chem_versions

Line 178-179: Please cite the original reference of 3-hourly scale factor. You need to read the paper and give credits to people who contribute to the work.

Line 181-182: I believe either Liu or Permar was working on the western US, which is mostly temperate forest. I do not think they are appropriate to compare to boreal forest. If you want to compare boreal forest, you need to cite paper for EF reports from boreal forest or at least mention the EF reference for GFED4. Again, you need to give credits to people who contribute to the work.

Line 185: Results from online and offline MEGAN computation is significantly different. See the web below. I believe offline and online are both turned on by default, which is recommended by GEOS-Chem team. Would the solely online MEGAN influence your results? How do you resolve it?
http://wiki.seas.harvard.edu/geos-chem/index.php/MEGAN_v2.1_plus_Guenther_2012_biogenic_emissions#Note_on_the_use_of_pre-computed_emission_factor_.28EF.29_maps_versus_EF_maps_computed_online

Line 193: Did you explicitly represent the chemistry for isomers or just use the default lumping chemistry for monoterpenes? If I understand it well, the monoterpenes you used is lumped monoterpenes (no?).

Line 195-205: Read through narrative description for more information (web provided above). Using MEGAN as an example, you need to cite Hu et al. (2015) as well. Example of how to cite MEGAN in GEOS-Chem narrative description: Biogenic VOC emissions in GEOS-Chem are from the MEGAN v2.1 inventory of Guenther et al. [2012] as implemented by Hu et al. [2015b].

Line 197: I believe the isoprene chemistry is updated by Kelvin Bate. Again, please check the narrative description!

Line 210: Say three simulations a series of GEOS-Chem simulations. Be specific of what you are trying to write.

Line 211: Change ":" into "including" might be better?

Line 215: Is it fair to assume the equation is linear for VOCs can be highly impacted by secondary production? How do you resolve the nonlinearity here and how it would influence your conclusion?

Line 239-240: Again, is the assumption fair? See above comments for line 215 instead.

OMI/OMPS Evaluation with GEOS-Chem HCHO VCD

Line 250-253: I did not fully read Nowlan et al. (2023) but are you suggesting the correction in that paper does not fix the "high-latitudes bias"?

Line 255-258: What's the satellite retrievals uncertainty here? Try to be specific here.

Line 258-259: What is the "biases in both model and satellites"? Try to be specific rather than using vague words.

Figure 2: Is the carbon emissions from emission inventories?

Line 312: What is Mann-Kendall test?

Line 315-320: What cause the inconsistency here? Any implications we can gain here?

Main drivers of HCHO VCD interannual variabilities

Line 330: Again, does the vegetation production matter? See my comments for MEGAN.

Line 334-335: There are other potential reasons. NOx level is one for sure. It could also relate to the difference of temperature, photolysis, and oxidants level.

Line 340-345: Might need to weaken the words? You are comparing the campaign-average with your custom model results rather than directly compare to the observation at the exactly time and location if I understand it right.

Line 347: It should be easy to use the recommended veg map in the GEOS-Chem and then figure out if it is the reason (no?).

Figure 5: Is that seasonally or monthly HCHO? It seems that you only have four data points for each year.

Line 395: See comments for Line 334-335

Line 397: I might not be familiar with such work but how are BVOC and methane oxidation sensitive to varying temperature?

SIF evaluation on dVCDBio,GC interannual variability

Line 412-413: What is the specific reason for choosing the Standardized Major Axis (SMA) regression?

Line 415-430: As both SIF and dVCDBio,GC are largely driven by surface temperature, why not just using surface temperature instead? What's the implication of SIF ranges (0-0.25)? What's the benefit of using SIF rather than other proxy?

Conclusions and discussions

Line 462: Jin2021 highlighted the importance of OH in HCHO production. Liu and Permar reported the observed fire ER/EF. How do they suggest underestimated HCHO in wildfire emission?