Review #2

→ We appreciate your valuable remarks and suggestions. Below, we will reflect on them and we are convinced that the modified version of manuscript is improved thanks to your comments. We replied to your comments (black) in blue, and modifications in the manuscript are expressed as underlined.

We changed words ‘g_{CA}’ to ‘g_i’ and ‘BB model’ to ‘BWB model’ in the manuscript.

General Comments

This paper addresses some important factors that influence the modeling of carbonyl sulfide, with the goal of improving our ability to use it to estimate GPP. Plant-specific optimization of conductance parameters is a really useful way to approach model improvement. The authors used some nice measurement datasets at a couple of different forested sites and were able to demonstrate reduced model-obs mismatch with their optimized setup. Globally this also addresses some of the gaps pointed out by previous studies (e.g. missing sink at high latitudes). While there is still room for improvement, this is a good first step in improving our ability to model OCS. My comments are mostly minor or eliciting clarification. As with RC1, I was also hoping the authors would circle back to the impact on Vcmax and ways to potentially optimize that independently, but there is always the next manuscript!

Specific Comments

- Para from line 41: Soil emissions may also play a role in some specific regions (e.g. hot areas or agricultural fields). See refs cited in Ogee et al, Biogeosci 2016.

  → Thank you for your suggestion. We added the ‘soil emission’ in the sentence.

- In the manuscript:
  (Original) *(lines 41-42 in the old version):*
  Atmospheric COS mole fractions vary around 500 parts per trillion (ppt) and are primarily influenced by biosphere uptake, ocean emissions, and anthropogenic emissions (Kettle et al., 2002).

  (Modified) *(lines 42-44 in the new version):*
  Atmospheric COS mole fractions vary around 500 parts per trillion (ppt) and are primarily influenced by biosphere uptake, ocean emissions, and anthropogenic emissions (Kettle et al., 2002). Depending on the environmental conditions, soils can act as a COS source or sink (Maseyk et al., 2014; Whelan et al., 2016).

- Lines 44-45: More recently, Hu et al (PNAS 2021) also showed existence of this missing sink at higher latitudes.

  → We added the reference in the sentence.

- In the manuscript:
  (Original) *(lines 43-45 in the old version):*
  Moreover, Berry et al. (2013) showed that a sink is missing, or a source is overestimated at higher latitudes. These findings ask for careful evaluation of all sources and sinks, including the biosphere.

  (Modified) *(lines 46-47 in the new version):*
  Moreover, Berry et al. (2013) and Hu et al. (2021) showed that a sink is missing, or a source is overestimated at higher latitudes. These findings ask for careful evaluation of all sources and sinks, including the biosphere.

- Line 105: add the word 'prognostic' to SiB4 description

→ We added the word 'prognostic' in the first sentence of the paragraph.

- In the manuscript:
  (Original) *(The line 102 in the old version):*
  The SiB4 model is a land surface model that calculates the COS flux as described in Berry et al. (2013)

  (Modified) *(The line 109 in the new version):*
  The SiB4 model is a prognostic land surface model that calculates the COS flux as described in Berry et al. (2013).

- Line 202: technically, GPP is not an 'observation' but a 'derived/modeled quantity' so should not be included in this list of obs.

→ Thank you for your remark. We changed the subheading and expressions concerning GPP. We also rearrange the paragraph from site-based to variables-based: COS flux, GPP, and COS soil flux measurements.

- In the manuscript:
  (Original)
  2.2.1 Observed variables, 2.2.2 Observation-based g_s and g_{CA}
  (Modified)
  2.2.1 COS flux, GPP, and mixing ratio, 2.2.2. Conductances g_s and g_i'

  (Original) *(lines 202-214 in the old version):*
  In the optimization of g_s and $g_{CA,}$ we used observed values of the variables required in the COS leaf uptake calculation (Eq. (1)), namely: COS ecosystem flux, COS soil flux, GPP, C_{cos}, temperature, and specific humidity. The observations were obtained at Hyytiälä in Finland during 2013-2017 (Kooijmans et al., 2017; Sun et al., 2018; Vesala et al., 2022) and at the Harvard Forest in the United States during 2012 and 2013 (Commane et al., 2015; Commane et al., 2016; Wehr et al., 2017). COS and GPP ecosystem fluxes were measured with the eddy-covariance (EC) technique.
  For Hyytiälä, the EC processing steps were described by Kohonen et al. (2020) and Vesala et al. (2022) and GPP was derived from NEE using multi-

year parameter fits (Kolari et al., 2014, Kohonen et al., 2022). The effect of storage in the canopy airspace was corrected by collocated COS profiles (Kooijmans et al., 2017; Kohonen et al., 2020).
For the Harvard Forest, we used GPP derived from CO2 isotope EC measurements as reported in Wehr et al. (2016), and we used canopy COS uptake derived from COS EC measurements as reported in Wehr et al. (2017).
In addition to the COS ecosystem fluxes, COS soil flux measurements were available for the 2016 growing season at Hyytiälä, and for the 2012 and 2013 growing seasons the Harvard Forest.

(Modified) *(lines 212-228 in the new version):*
In optimizing the parameters g_s and g_i, we used the following variables obtained from observation to calculate COS leaf uptake (Eq. (1)): the COS ecosystem flux, the COS soil flux, C_{COS}, temperature, specific humidity, and GPP partitioned from NEE measurements. These data were collected and derived at Hyytiälä in Finland during 2013-2017 (Kooijmans et al., 2017; Sun et al., 2018; Vesala et al., 2022) and at Harvard Forest in the United States during 2012 and 2013 (Commane et al., 2015; Commane et al., 2016; Wehr et al., 2017). To validate the optimization results, we used the observation-based g_s and g_i (Sect. 2.2.2).

2.2.1 COS flux, GPP, and mixing ratio
We used canopy COS uptake derived from COS EC measurements for Hyytiälä (Kohonen et al., 2020; Vesala et al., 2022) and Harvard Forest (Wehr et al., 2017). The effect of storage in the canopy airspace was included by collocated COS profiles (Kooijmans et al., 2017; Kohonen et al., 2020).
GPP at Hyytiälä has been obtained from NEE using multi-year parameter fits (Kolari et al., 2014, Kohonen et al., 2022). For Harvard Forest, we chose to use the GPP derived from the isotope spectrometer measurements, because it is more accurate and reliable with frequent and rigorous calibrations (Wehr et al., 2016).
COS soil flux measurements were available for the 2016 growing season at Hyytiälä, and for the 2012 and 2013 growing seasons at Harvard Forest. For the soil flux in other years at Hyytiälä, we applied the monthly average diurnal cycle of the soil flux from 2016 to the other years (2013-2015 and 2017). The seasonal and diurnal variation of the soil flux is small compared to the total ecosystem uptake of COS (Sun et al., 2018). Hence, the averaged value of 2016 can be safely used for other years.

- Line 282: change wording to ‘observation and observation-derived quantities’ since ‘GPPobs’ and ‘gs’ are not direct observables but rather derived quantities.

→ As you suggested, we added the word ‘observation-derived quantities’.

- In the manuscript:
  (Original) *(lines 282-284 in the old version):*
  Figure 3 specifies which observations are used in which step (g_s, GPP_{obs}, $F_{COS, obs}$, C_{cos} highlighted as grey) and which variables are simulated by SiB4 (e.g. T_{can}, F_{LH}, GPP_{SiB4}, CO_{2S}, g_b).
  (Modified) *(lines 295-296 in the new version):*
  Figure 3 specifies which observations and observation-based quantities are used in each step (g_s, GPP_{obs}, $F_{COS, obs}$, C_{COS} highlighted as grey) and which variables are simulated by SiB4 (e.g. T_{can}, F_{LH}, GPP_{SiB4}, CO_{2S}, g_b).

- Fig 4 comment: is the reason for missing hours in the HVFM All plot that there is no data for certain phenological stages (apart from growth and maturity for which you have data at all hours)?

  → Yes! The GPP data from 18h to 3h at Harvard Forest are missing for all phenological stages.

- Line 332-333: does this imply you used the new f(Tcan) estimations for forests and applied them to grasslands as well? That seems like it could cause additional problems.

  → Yes, grasslands might have a different temperature of the enzyme CA's activity. However, without observations we cannot say much. The necessary measurement for other vegetation types are mentioned in section 3.5.2. "With more measurements over different vegetation types, these parameters could also be optimized for a wider range of ecosystems."

- Line 344 comment: did you investigate whether 100 was sufficiently large?

  → To show the convergence of the optimized values, we now performed 100 simulations, and plotted the median posterior values against the number of simulations (Figure 1). Overall, we observe that the optimized median converged to their final values somewhere after ~100 optimizations.

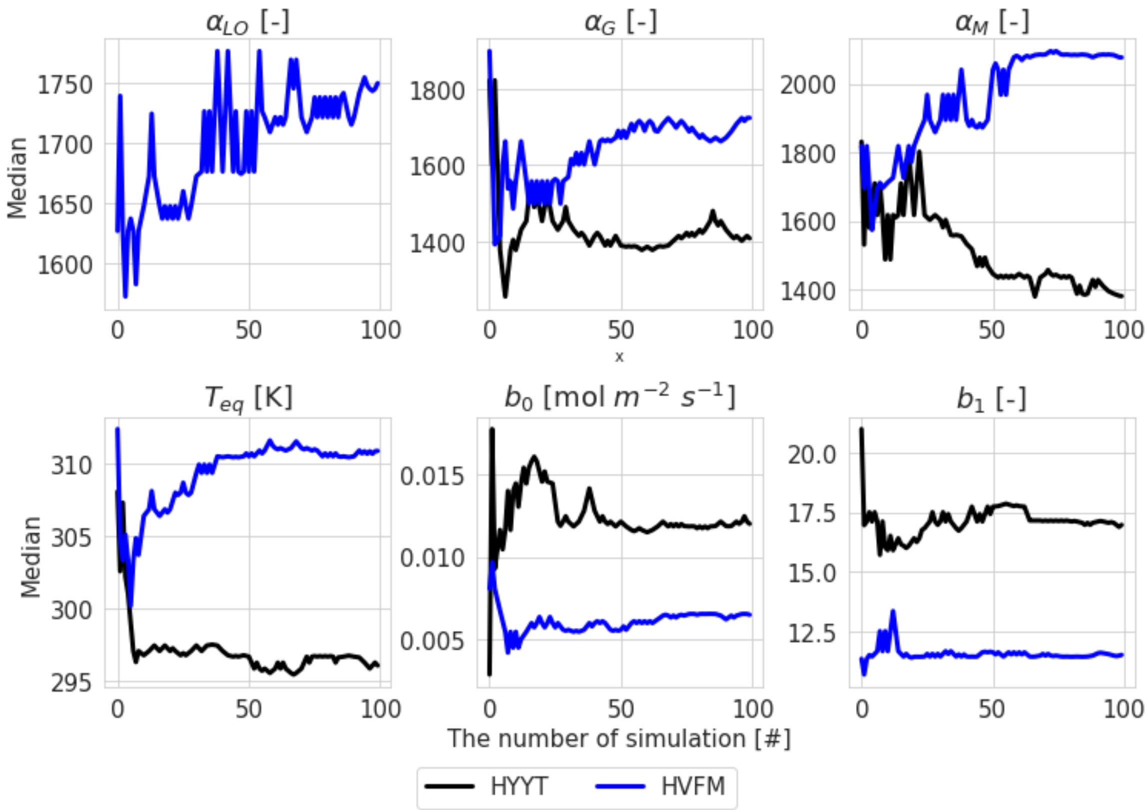

Figure 1. Median values of state variables at the two stations as a function of the number of randomly perturbed simulations (HYYT: Hyytiälä; HVFM: Harvard Forest)

- Table 3: where does the prior error range come from? Perhaps a reminder is in order referencing Appendix A where the prior error is estimated (as mentioned in sec 2.3.2)

→ We agree with your suggestion. We added the related expression to the caption in Table 3.

- In the manuscript:
  (Original) *(lines 424-425 in the old version):*
  Table 3. Original (Org) and optimized (Post) state vectors for Hyytiälä and Harvard Forest in different phenological stages as defined by SiB4. Values of Posterior in parenthesis indicates posteriori errors. Detailed error reduction is described in Appendix B.

  (Modified) *(lines 450-452 in the new version):*
  Table 3. Original (Org) and optimized (Post) state vectors for Hyytiälä and Harvard Forest in different phenological stages as defined by SiB4. Values of Posterior in parenthesis indicate posterior errors. The definition of the prior values is outlined in Appendix A, and the error reduction is described in Appendix B.

- Fig 7a comment: the red and orange lines don't seem that different here, perhaps cite some calculated statistical significance to emphasize that they are different?

→ Indeed, the red and the orange lines at Hyytiälä are almost the same because optimized alphas (1316 for Growth, 1331 for Maturity phase) are similar to the original value (1400). Thus, we included g_i statistics only for Harvard Forest in the text.

- In the manuscript:
  (Modified) *(lines 465-467 in the new version):*
  In contrast, when the optimized value of α is included (red line), the amplitude of g_i is improved. Compared to the optimization that excluded α, the MBE is reduced from 0.006 to 0.003 $mol\ m^{-2}\ s^{-1}$.

- Fig 7b: why not also show the equivalent to the orange lines for Harvard Forest? (i.e. with optimized f(Tcan) but original alpha.

→ We drew the orange line in Fig. 7(a) to check the improvement of g_i. However, F_{COS} is influenced by many more factors (e.g. g_s and g_i), and the effect of α cannot be isolated. For this reason, the orange line is missing in Fig. 7(b).

- Line 475: your result seems to imply that above-canopy RH is a better observational quantity to use to derive gs, but this is counterintuitive in that the 'gs' specifically involves resistance (or conductance) at the leaf surface, and so theoretically we should use RH at the leaf surface. One alternate explanation here is that it could be incorrect leaf temperature which can lead to a bias in leaf surface RH which propagates to gs.

→ Thank you for your suggestion. There are three reasons that could lead to F_{LH} being overestimated: (1) water vapor flux in the boundary layer to the leaf surface is too large, (2) boundary conductance is too small, and (3) leaf surface temperature is too small. Although we cannot compare the observed leaf surface temperature with the estimated one due to the absence of observation, the temperature in the canopy air space fits well with the observed air temperature (See Figure 1 below). We speculate that the overestimated water vapor flux is the main reason because the observed RH above the canopy explains the diurnal fluctuation of g_s, as shown in Figure 10. As you mentioned, we theoretically should use the leaf surface RH. Since we could not obtain the observed leaf surface RH, we used the RH above the canopy instead. Furthermore, factor (2) about boundary conductance should be evaluated with observations.

- In the manuscript:
  (Original) *(lines 475-478 in the old version):*

However, SiB4 still tends to overestimate gs in the morning and late afternoon. In contrast, when we base the gs calculation on the observed RH above the canopy, the diurnal cycle is better simulated (orange dashed line). This implies that SiB4 has the tendency to underestimate the humidity stress in the late afternoon when converting observed specific humidity above the canopy to humidity at leaf surface level.

(Modified) *(lines 503-511 in the new version):*
However, SiB4 still tends to overestimate g_s in the morning and late afternoon. The overestimated F_{LH} in SiB4 can result from three factors: (1) an overestimated water vapor flux in the boundary layer to the leaf surface, (2) an underestimated boundary conductance or, (3) an underestimated leaf surface temperature. Since we do not have observations of the leaf surface temperature, we confirmed that the estimated canopy temperature has a tight 1 to 1 relation with the observed air temperature. We speculate that the main reason for the overestimated F_{LH} is the uncertain water vapor flux. When we base the g_s calculation on the observed RH above the canopy, the diurnal cycle is better simulated (orange dashed line in Fig.10). The overestimated water vapor pressure implies that SiB4 tends to underestimate the humidity stress in the late afternoon when converting observed specific humidity above the canopy to humidity at leaf surface level. We suggest evaluating the boundary conductance (point (2) above) with observations.

- Line 484: what are the alternatives to 'stomatal transpiration'?

→ As mentioned in Sect.2.2.2, observation-based g_s was calculated by estimated transpiration from observed evapotranspiration with fixed fitting equation (FG approach). When evaporation is larger, evapotranspiration can be larger. When we checked the ratio of evaporation to evapotranspiration in SiB4 at Hyytiälä, we found that the ratios are larger in three months: April, September, and October. This might result in overestimated g_s in the observations because we used the same fixed fitting equation for the entire year. Therefore, the alternatives to 'stomatal transpiration' indicates evaporation here.

- In the manuscript:
  (Modified) *(lines 513-517 in the new version):*
  The optimized model still underestimates g_s at Hyytiälä in April, September, and October (Fig. 8b). This might indicate that we did not properly separate stomatal transpiration rates from the observed latent heat flux. The simulated mean ratios of evaporation to evapotranspiration in these three months are 66 %, 60 %, and 95 %, respectively, and these values are higher compared to the other months (43 to 53 %). Thus, we speculate that the observed evapotranspiration does not solely represent stomatal transpiration in these months due to larger evaporation rates, leading to overestimated g_s in the observations.

- Line 493: clarify 'indicating humidity stress only shortly at midday'. Do you mean that the impact of humidity stress is short-lived or only important around midday?

→ It means that the duration of humidity stress is simulated too short by SIB4. We modified the text:

- In the manuscript:
  (Original) *(lines 492-493 in the old version):*
  However, g_s values are generally overestimated and SiB4 simulates two peaks during daytime, indicating humidity stress only shortly at mid-day.

  (Modified) *(lines 526-528 in the new version):*
  However, g_s values are generally overestimated and SiB4 simulates two peaks during daytime. This indicates that humidity stress is only briefly occurring at mid-day in SiB4.

- Line 498: which 'pseudo-observations'? maybe just use 'observationally-derived X' where X is the quantity you're referring to here and mention it explicitly.

→ Both have the same meaning. We changed the word 'pseudo-observations' to 'observation-based'.

- Line 515: and also consistent with Hu et al (2021, PNAS)

→ We added the new reference 'Hu et al., (2021)' in the manuscript.

- In the manuscript:
  (Original) *(lines 514-516 in the old version):*
  The higher uptake at high latitudes and lower uptake at the tropics are nevertheless consistent with inverse modelling results presented in Ma et al. (2021) and would help towards closing the COS budget.
  (Modified) *(lines 548-549 in the new version):*
  The higher uptake at high latitudes and lower uptake at the tropics are nevertheless consistent with inverse modelling results presented in previous studies (Ma et al., 2021; Hu et al., 2021) and would help towards closing the COS budget.

- Line 542: how does the improvement in b0 compare to night-time conductance values calculated for CLM by Lombardozzi et al 2017? (maybe this citation could be discussed earlier where you mention b0 results)

→ Thank you for your idea to compare with the previous study by Lombardozzi et al., 2017. They show a minimum stomatal conductance in a boreal needle-leaf evergreen forest of 0.008 mol m^{-2} s^{-1} with one observation and in a temperate broadleaf deciduous forest as 0.073 (mean) ± 0.084

(standard deviation) mol m^{-2} s^{-1} with 22 observations. Comparing our b_0 with the value from Lombardozzi et al. is problematic due to the limited number of observations (n=1). The results over a temperate broadleaf deciduous forest have large variations, which might vary with season, humidity, etc. Therefore, it is challenging to compare our results to their observations. We need seasonal and diurnal observations to evaluate our optimized b_0 results.

- Appendix A comment: I think your prior errors are based on the 'initial value +/- 1.5 state errors'? So for example prior alpha should then be 1400 +/- 700 (as is shown in Fig B1A for HYYT). But this is inconsistent with Table 3 where you list 1400 +/- 1000. Can you please clarify?

→ In Figure A1(a), the red line indicates the 75 percentile ranges from 400 to 2500. Therefore, we decide to use a prior alpha of 1400 +/- 1000.

Technical Corrections

- Line 109: replace 'heterogenic' with 'heterogeneous'. I think the former is more related to genetic/species aspects.

→ We replaced 'heterogenic' with 'heterogeneous'.

- Line 234: delete 'the', or add the word site after 'Harvard Forest'

→ We deleted articles in front of 'Harvard Forest'

- Line 264: delete 'to use'

→ As you suggested, we deleted it.

- Line 266: delete 'of' before 'uncertainties in GPP'

→ As you suggested, we deleted it.

- Line 338: change 'humidity impact' to 'humidity stress impact' GPP'

→ As you suggested, we changed the word.

- Line 340: Capitalize 'we'

→ We added the part of the sentence 'to account for the better humidity impact for the global COS leaf uptake'.

- In the manuscript:
  (Original) *(The line 340 in the old version):*
  we simulated the global COS leaf uptake without the 0.7 threshold of F_{LH} for ENF.

  (Modified) *(lines 359-360 in the new version):*
  <u>To account for the optimized humidity impact on the global COS leaf uptake,</u>

we simulated the global COS leaf uptake without the 0.7 threshold of F_{LH} for ENF.

- Fig 5 caption: replace 'in' with 'at'

→ We replaced every 'in' in front of 'Harvard Forest' with 'at' in the manuscript.

- Line 403: replace 'in' with 'at', and 'In' with 'At' (and in any other instances when mentioning the sites, such as lines 407, 413, 433, Fig 7 caption, 463...)

→ We replaced every 'in' in front of 'Harvard Forest' with 'at' in the manuscript.

- Line 411: 'higher and smaller respectively'

→ We added 'respectively' in the sentence.

- Line 430: replace 'pseudo-observations' with 'derived' or 'observationally-derived'

→ We replaced all 'pseudo-observations' with 'observation-based' in the manuscript.

- Line 491: 'At the Harvard Forest site..'

→ We deleted articles in front of 'Harvard Forest'

- Fig A1 caption: 'where the cost is minimized' or 'where the cost reaches a minimum value'.

→ We modified the part of the caption as 'where the cost is minimized'.

- In the manuscript:
  (Original) *(line 570-574 in the old version):*
  Figure A1: Cost function values plotted against the value of the state vectors elements in Hyytiälä (solid line) and Harvard Forest (dotted line). The red lines indicate a criteria cost calculated by *H(x)* as the 75-percentile value of every three-hourly observation in each month. While the target parameter changes, the other variables are fixed as α = 1400 (Hyytiälä), 2000 (Harvard Forest), ΔH_a = 40 kJ mol^{-1} ΔH_{eq} = 100 kJ mol^{-1}, and T_{eq} = 295 K (Hyytiälä), 310 K (Harvard Forest), b_0 = 0.02 (Hyytiälä), 0.01 (Harvard Forest), and b_1 = 17 (Hyytiälä), 12 (Harvard Forest). These values were decided where the cost has minimum.

  (Modified) *(lines 606-610 in the new version):*
  Figure A1: Cost function values plotted against the value of the state vector elements at Hyytiälä (solid line) and Harvard Forest (dotted line). The red lines

indicate a criteria cost calculated by *H(x)* as the 75-percentile value of every three-hourly observation in each month. While the target parameter changes, the other variables are fixed as α = 1400 (Hyytiälä), 2000 (Harvard Forest), ΔH_a = 40 kJ mol^{-1} ΔH_{eq} = 100 kJ mol^{-1}, and T_{eq} = 295 K (Hyytiälä), 310 K (Harvard Forest), b_0 = 0.02 (Hyytiälä), 0.01 (Harvard Forest), and b_1 = 17 (Hyytiälä), 12 (Harvard Forest). These values were based on the value where the cost reached a minimum.

- Line 585: Reiterating comments from above, GPP is not an observation.

→ Thank you for your remark. We rewrote the GPP part in the sentence.

- In the manuscript
  (Original) *(line 585-586 in the old version):*
  We optimized each ensemble with the same observations (GPP and COS leaf uptake) and state variables but added noise to each ensemble member (Chevallier et al., 2007).

  (Modified) *(lines 621-622 in the new version):*
  We optimized each ensemble with the same y (observationally derived GPP and COS leaf uptake) and x but added noise to each ensemble member (Chevallier et al., 2007).