General comments

The paper is well-structured and concise. The paper presents a useful tool to perform sensitivity analysis, and an essential building block to create a novel inverse modelling framework to be used in the future. The global variable-resolution mesh of the forward model is a promising tool to keep computational costs limited in this future inverse modelling framework, while avoiding issues related to lateral boundary conditions. The validation of the Tangent Linear (TL) and adjoint models is described in detail. The section on sensitivity could use some additional information on employed resolutions and employed initial meteorological conditions. The language is generally fluent and precise, and the authors refer extensively to existing literature. The introduction is also concise and to the point, and equations are well explained.

I have some specific comments, which will hopefully prove useful to further improve the manuscript.

Specific comments

- Line 11: '… satellite-derived column-averaged'. 'satellite-derived column-averaged CO_2 mixing ratio'?
- Line 15: Loss of information is mentioned here due to CT-L's simulation length and spatial domain limitations. In Section 5, I however only see simulations in which the simulation length is chosen identical to the one from the MPAS adjoint. So, aren't it only the spatial domain limitations that are relevant here?
- Line 49: This sentence reads somewhat strange, MPAS-CO2 addresses the limitation by being based on MPAS-A? The sentence coming after that is very clear though.
- Line 70: in equation 1, what is the meaning of the subscript cu in the last term? It is not immediately clear to me from the description. Maybe this can be shortly explained in the text if you think that is relevant?
- Equations 3 and 4: The statements $F_{bl} = \frac{\partial q_{co_2}}{\partial t}$ and $F_{cu} = \frac{\partial q_{co_2}}{\partial t}$ are slightly confusing to me. I assume that e.g. F_bl equals the local co2 tendency when only the PBL contribution is taken into account, not the full CO2 tendency. When applying the equations as they are written now consistently, it would mean that $\frac{\partial q_{co_2}}{\partial t}$ = F_bl = F_cu, which is presumably not what was intended. Maybe this can be made slightly more clear?
- Next to the above, From eq 1 it is clear that F_bl has the same units as $\frac{\partial(\tilde{\rho} q_{co_2})}{\partial t}$. It is not really clear to me how F_bl can have the units of $\frac{\partial q_{co_2}}{\partial t}$ (as follows from eq. 3) and $\frac{\partial(\tilde{\rho} q_{co_2})}{\partial t}$ at the same time, unless maybe $\tilde{\rho}$ is dimensionless? Adding units to the explanation of the variables could make things more clear.
- Equation 3: you are explaining what all the variables are that are used in equations 1, 2 and 3, but you do not explain the flux term $\overline{(w' q'_{co_2})_h}$ in equation 3.
- Line 102: 'e is 2-dimensional'. I assume 2-dimensional in space, but doesn't it have a time dimension as well?
- Line 117: What is meant with 'an emission driver of the forward model'?
- Equation 9: Maybe you can specify the meaning of the '||' symbols? I assume it is the 'length of vector' operator. The use of this operator in the equation however raises a question to me: Imagine a hypothetical case where the tangent linear model would contain errors. In that case, the vector

calculated using the TL model (M(0,αk), as in the denominator of equation 9) will be different from the vector calculated in the numerator of equation 9. However, this does not always imply that also the **length** of the vector M(0,αk) will be different from the length of the vector in the numerator. Thus, I would think that in this specific case with errors in the TL, equation 9 could still be satisfied, and thus does not guarantee correctness of the TL. Shouldn't this test be performed for each element of the TL output vector separately (thus for each mixing ratio in the grid at time t), thereby avoiding the need to work with the 'length of a vector' operator (see also eq. 6 of Errico 1997 and eq. 25 of Bosman and Krol 2023)? What is your opinion on this?

- Line 149 : I understand the reasoning that there would be no need to test sensitivity to x0, since the calculation of the sensitivity of x_t with respect to the flux scaling factor k involves the TL code of the CO2 emission driver in addition to the TL code of all the CO2 transport processes. But still, aren't the derivatives specifically to x0 untested in this way?
- Equation 10: About the use of $\Delta\hat{x}_i$ in the right-hand-side of the equation: Does the adjoint always 'ingest' the mixing ratios at forecast time, or can it also 'ingest' a time-array of observations between t0 and the forecast time (as is done in 4d-VAR)? In that case equation 10 is perhaps not a complete representation of the adjoint model
- Line 187-188: Reading these lines, it sounds to me as if the TL model does not require meteorological fields to run. Is that indeed true, and if so, why is this? If it does require meteorological fields (as I expect from lines 236-237), perhaps the sentence can be written slightly more clear.
- Line 190: '…during the adjoint model's forward sweep…'. Isn't it rather a forward model run during which the meteorological fields can be saved, instead of a 'forward sweep' of the adjoint model? Perhaps I am just used to a slightly different system of using an adjoint.
- Line 206: 'adjoint variable of flux scaling factor output from the adjoint model integration' Do you mean adjoint variable of flux scaling factor, which is output from the adjoint model integration from forecast time backward to the initial time? Perhaps this can be written slightly differently to make it more clear?
- Line 217: 'the agreement between the LHS and RHS of Eq. (11) is about −5.15×10−15'. Perhaps you can mention somewhere in this paragraph that it is not exactly zero due to (machine) rounding errors.
- Line 217: −5.16×10−15 instead of −5.15×10−15?
- Line 222: As mentioned on line 203, equation 11 should be satisfied for any combination of Δx and Δk. Therefore, I do not really understand why you refer to equation 9 here to justify your choice.
- Line 230: The forward model is not validated in this paper, I assume this is done in the Zheng et al. (2021) reference. So perhaps the sentence could be slightly reformulated.
- Line 236: 'the TL and adjoint model simulations using the same configuration take approximately 10% longer, indicating that the majority of the computation time is used for integrating the meteorological fields'. This confuses me somewhat: Aren't the meteorological fields calculated during a forward model run as well?
- Line 262-263 'When a uniform time-invariant surface flux is used, the sensitivity to the surface flux scaling factor calculated by the MPAS-CO2 adjoint model is the observation footprint.' Isn't this only true when the flux is set to 1, otherwise it should still be multiplied with a fixed factor? And shouldn't S(k) be taken into account as well?
- Line 265: What function is chosen for S(k) in equation 6?
- Line 267: Where do the initial meteorological conditions (for generating the meteorological fields for the adjoint) come from?
- Line 270: I read here an intake height of 475 m for the WKT tower, but the captions of Figure 2 and Figure 3 mention 457 m…
- Caption of figure 2: The units of the sensitivity of CO2 at the tower to the surface CO2 flux scaling factors are given here as ppm/µmol m−2 s−1. However, given equation 6, I assume the scaling factors to be dimensionless. Because of that, one expects the units of the sensitivity to be ppm instead.

However, because we use a fixed flux of 1.0 μmol m−2 s−1, the sensitivity to the flux scaling factor is equal to the sensitivity to the scaled CO2 flux (e in Eq. 6). Perhaps it might be good to clarify the origin of these somewhat counterintuitive units for the sensitivity to the surface CO2 flux scaling factor.

- Figure 2: In all of the panels, there is a peculiar 'red blob' present in Antarctica. As it doesn't look like this has a physical origin, it might need some explanation/correction. Also in e.g. Figure 8 I notice this.
- Caption Figure 3: the height of only one of the towers is given here. Perhaps better to give either none or both.
- Line 277: '…indicates that the variation in the sensitivity magnitude decreases with time' Do you mean spatial variation or temporal variation? Maybe good to state this explicitly.
- Line 279: 'The triangles in Fig. 3 show that the magnitude of the standard deviation of sensitivity to the CO2 mixing ratio decreases rapidly with time for both towers'. I find this statement somewhat confusing, given that time runs backward in the figure, and thus σ actually increases over (forward) time. Perhaps you could find an elegant way to make this clear. The same holds for line 285.
- Line 286: Interesting finding, but how can we know for sure that a small change rate of the σ of the footprint also means a small change of the footprint itself?
- Line 294: Oregon and Nevada are not mentioned?
- Line 306: How well do the levels of the two models correspond? Is it exact?
- Line 318: The comparison would be easier when the left and right panels in figure 5 would use the same color scale, now there is a notable difference between the scales. This is also the case in Figure 6 and 7.
- Line 330: 'Both the MPAS-CO2 adjoint and CT-L footprints for 500m are generally confined near to the sounding location'. Looking at the figure I would say there is still quite some sensitivity to further away regions as well.
- Line 338: Is it a single adjoint model run, or a single run for each height level? This is not immediately clear to me the way the sentence is written.
- Line 339: Am I interpreting it correctly, when I say that the sensitivity at a single height level in Fig. 8 is the sensitivity of the average mixing ratio along the OCO-2 track at the specified model height, whereby the average is taken horizontally over all cells that are part of the OCO-2 track? Perhaps it might be good to add a little more explanation at this point in the paper.
- At which resolutions are the CT-L and adjoint models run in Section 5? I suggest to include this information.
- Line 352: 'at higher heights from 4,500m to 14,000m, the footprints calculated by CT-L tend to be of much higher magnitude compared to the MPAS-CO2 adjoint model'. Looking at the figures, this only holds for a fraction of the footprints, definitely not all of them. Perhaps good to make this clear.
- Figure 9: Just a suggestion, but wouldn't it be interesting to include one or two simple metrics in the subfigures to numerically quantify the difference between the adjoint and CT-L? Now it is purely visual. Perhaps just the difference in the means and a measure for correlation, or maybe a (scaled) root mean squared difference or something… The discussion on page 12 could then also be based on these numbers.
- Caption Figure 9: '…extracted as the average value of 2°x3° boxes within the range of the CT-L spatial domain (10-80°N, 180-10°W)'. However, at line 257 the longitude of the CT-L domain is given as '0°-180° W', which is different.
- Line 362: Is the adjoint forcing added to all the cells where the OCO-2 track passes? It is not mentioned where the forcing is added.
- Figure 10: To what extent do these 20 pressure levels coincide with MPAS-CO2 pressure levels?
- Line 375: 'Profile 2, …, appears to be more sensitive to the stronger convective transport of surface CO2 flux in the tropics than in the extratropics'. Do you mean 'appears to be more sensitive than profile 1 to the stronger convective transport…'? Perhaps good to make this clear. Next to that, I don't understand where this statement comes from. In the figure I see that the footprints of Profile 2 are generally smaller than those of footprint 1 in the tropics, so shouldn't the statement be the opposite

of what you state now? If not, Perhaps you can explain your reasoning how you go from the information in the figure to the statement you make (also in the paper)?

- Line 377: 'These convective transport differences…' Convective transport differences between the 2 adjoint simulations? Given that the two adjoint simulations use the same meteorology etc., this doesn't read very clear to me, or maybe I don't understand it properly. Do you mean differences between the 2 footprints, originating from convective transport to higher levels of the atmosphere?
- Comparing line 398 with 354-355, it seems the given list of reasons for the differences has now been replaced by 'likely arise from variations in vertical transport'? Perhaps not fully consistent? Or is there a specific reason why you can narrow it down here? Please check also whether the related part in the abstract it is consistent with lines 354-355.
- A small note on the code and data availability section, for your information: GMD encourages the submission of code to perform calculations described in the text, see https://www.geoscientific-model-development.net/peer_review/review_criteria.html.
- Line 408: Perhaps mention that the forward model is included as well.

Technical comments

- Line 64: 'conclusion**s**'?
- Line 91: small typo, the word 'two' occurs twice
- Line 106: Why is x_t written as X_t here? Or is X_t another variable?
- Line 172: 'Eq. (9)' instead of (9)
- Line 214: I think the word 'ratio' is missing after 'mixing'
- Line 222: 'configuration' instead of 'configurations'?
- Caption Figure 1: Small typo in first sentence, 'A' instead of 'An'
- In the rightmost column of Table 1, it is perhaps better to use the same number of significant digits or decimal places for all lines, now there is a sudden 'jump' from 1.0 to 0.99999999999998, while the precision of both numbers is actually the same or very similar.
- Caption figure 3: just a suggestion: maybe change the rather complex first sentence 'The variation of the standard deviation (σ) of sensitivity to the initial CO2 and the sensitivity to the flux scaling factor (footprint) over time' into something like 'The variation over time of the standard deviation (σ) of two quantities: the sensitivity to the initial CO2 and the sensitivity to the flux scaling factor (footprint)'.
- Line 281: small typo, 'the footprint spread…'
- Caption Figure 4: 'black crosses in the figures on the right panel' instead of 'black crosses in the figure on the right panel'?
- Line 303: Small typo, I assume '2106' should be 2016.
- Line 362 '…and running backward in time…' The grammar seems not fully correct in this sentence.

References

Bosman, P. J. M. and Krol, M. C.: ICLASS 1.1, a variational Inverse modelling framework for the Chemistry Land-surface Atmosphere Soil Slab model: description, validation, and application, Geosci. Model Dev., 16, 47–74, https://doi.org/10.5194/gmd-16-47-2023, 2023.

Errico, R. M.: What is an adjoint model?, Bulletin of the American Meteorological Society, 78(11), 2577‑2591, 1997.