Response to Reviewers

February 15, 2026

We thank the editor and referees for the comments. Reviewer comments are in **bold**, while our responses are in normal text.

Response to Reviewer 1

Major comments
1. Please describe your CAT-mixing submodel in more details. This manuscript is about evaluating this specific parameterization, so do elaborate on the details, and not just place it in the supplementary material. For example, does the CAT mixing submodel act on all vertical elevations of the atmosphere, does it include the boundary layer? What values are set for the parameters (for example h1 and h2 in Eq. 1 and t_{norm} in Eq. 4), and based on what principles are these values chosen?

We have now included more details in the manuscript explaining the altitude region (given as pressure) that CAT act on and the choice of the selected values and summarized it in a table. We added in the manuscript At pressure levels below 70 hPa, vertical tracer mixing by turbulence is strongly suppressed due to the strong static stability. At pressures higher than 500 hPa, deep tropopause fold intrusions might occur and lead to tracer mixing; however, such structures are usually not represented well in a typical EMAC setup with a coarse horizontal resolution (e.g. T42 and T63). Therefore, the CAT parametrization is switched off for pressure lower than 70 hPa and higher than 500 hPa, considering the limited computational resources and to avoid overlap with the boundary layer parametrization scheme. and An example namelist can be found in the electronic supplement. The values used in this study are summarized in Table 2. The constant C in the DVT of the Ellrod-Knox index is selected according to our sensitivity analysis on the frequency distribution (Figure S3). We found that the suggested values of 0.1 from Ellrod and Knox (2010) and 0.01 from Lee et al. (2020) lead to either too strong or too weak DVT compared to the VWS and DEF. The time interval of DVT (nhours) is selected based on Ellrod and Knox (2010) and the limitation threshold (N^2_{lim}) is selected based on Chau et al. (2025). The t_{norm} which moderates the strength of tracer mixing, is an empirical value that shown to improve the ozone representation in EMAC (Figure 2b). We also added a limitation subsection of the CAT approach.

2. Eq. 3, the vertical flux $F \equiv \overline{w'x'}$ should be a vertical velocity scale times the unit of the tracer, which means that the CAT Index should have units of [m/s], right? But based on Eq. 4, the CAT Index as the same unit as CAT defined in Eqs. 1 and 2, which is [1/s]. Correct me if I am wrong, but what is with the dimensional inconsistency?

The flux should be with the unit of $mol mol^{-1} s^{-1}$. The Ellrod-Knox index has the unit of s^{-2}

since the vertical wind shear, deformation and divergence are with the unit of s^{-1}. The vertical wind shear times deformation resulting in the unit of s^{-2}, the divergent trend between two times is also resulting in the units of s^{-2}. The MoCATI also has a unit of s^{-2} since the stability term is dimensionless. The $\frac{\Delta t}{t_{norm}}$ is with the unit of s since Δt is with the unit of s and t_{norm} is dimensionless as well (calculated by the seconds of a day and time scaling factor with s as unit as well). So therefore CAT Index$_{boundary}$ will have a unit of s^{-1}, and the flux of the tracer will therefore with the unit of $molmol^{-1}s^{-1}$.

3. The MoCATI adopt a correction based on the Vaisala frequency alone (i.e., $1 - N^2/N^2_{lim}$). First of all, please define the Brunt-Vaisala frequency, did you use the dry Brunt-Vaisala frequency or the moist one? Secondly, I am curious why the stratification correction, wouldn't a Richardson number-based correction be better?

We used here the moist Brunt-Vaisala frequency with the equivalent potential temperature. We have now changed in the manuscript The N^2 is the moist Brunt-Väisälä frequency calculated by the equivalent potential temperature. However, considering the low concentration of water vapour in the UTLS mostly below 100 ppmv, the difference between using moist or dry Brunt-Vaisala frequency should be small. For the stratification correction, we try to avoid using the Richardson number because the shear is already taken into account by the Ellrod-Knox index, therefore only the Brunt-Vaisala frequency is used to avoid considering the shear twice.

4. Section 3.2, before you present the results for the chemistry species, I think it would help to present a latitude-height distribution of CAT mixing, i.e., the CAT Index$_{boundary}$ in your Eq. 3, so that the readers have a sense of where CAT-induced mixing occur. This would also assist the interpretation of the following figures.

We have added a figure of the CAT Index$_{boundary}$ in the appendix of the manuscript since the original section 3.1 is moved to the supplement. We have changed in the manuscript from The changes in both gases are stronger in the Southern Hemisphere, potentially due to the stronger jet streams and fewer mountains, leading to a stronger vertical wind gradient. to The changes in both gases are stronger in the Southern Hemisphere, potentially due to the stronger jet streams and fewer mountains, leading to a stronger vertical wind gradient and resulting in a higher CAT Index$_{boundary}$ as in Figure A1.

Minor comments

1. Abstract, please do not abbreviate at first appearance, such as MoCATI, EMAC-QCTM.

Corrected

2. Line 12, what "modified tracers", modified by what?

We rephrase it in the manuscript from This modification is not a pure result of the physical mixing but also the chemical feedback of other modified tracers. to This modification is not a pure result of the physical mixing but also the chemical feedback of other tracer distributions modified by CAT.

3. Line 13, not sure what you mean by "changing the O3 becoming relatively sensitive to NOx".

We mean that the ozone chemical production regime is becoming more sensitive to NOx abundance. As shown in Figure 1 below from Nussbaumer et. al. (2024), a lower alpha represents the ozone production becoming more sensitive to NOx (This does not necessarily mean, that it falls into the NOx sensitive regime). We rephrased it from becoming relatively sensitive to NOx to becoming more sensitive to NOx.

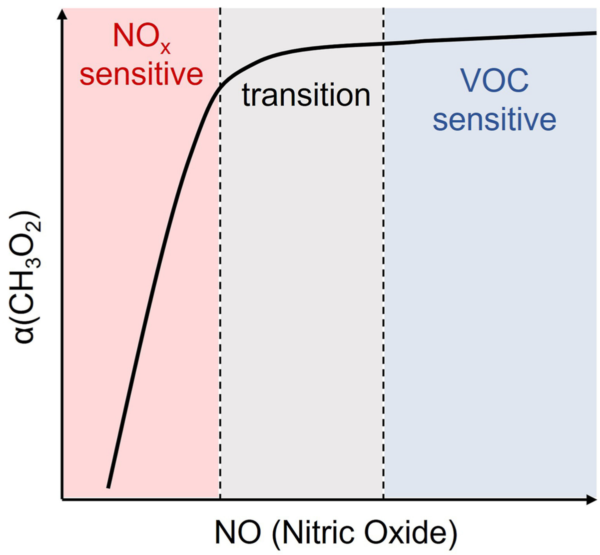

Figure 1: O3 sensitivity using $\alpha(CH3O2)fromNussbaumeret.al.$(2024).

4. Lines 27-28, "vertical shear, which has been recently identified constituting a layer of high probability of shear", please improve your grammar.

We rephrase it in the manuscript from which has been recently identified constituting a layer of high probability of shear occurrence around the local tropopause to which has recently been shown to form a layer with high probability of shear occurrence near the local tropopause.

5. Line 36, remove "for EMAC", you are talking about a physical process, right? Not just for a numerical model.

Corrected

6. Lines 42 and 45, I hope both definitions are given below.

The definition of the Ellrod-Knox Index is defined in Section 2.2. For Ellrod Index, since it is not the focus of the study and is included in the Ellrod Knox Index, we did not specifically define it in the manuscript. The following sentence, "The Ellrod-Knox index is the improved version of the TI by introducing an additional divergence trend term." also indicates the difference between the Ellrod Index and Ellrod-Knox Index.

7. Line 51, at which elevation? Please be more specific.

It is at 250 hPa. Now specified in the manuscript.

8. Line 58, what is "MESSy"?

Corrected to EMAC.

9. Line 60, again, what is "QCTM"?

Corrected to Quasi Chemistry transport Model (QCTM).

10. Line 116, so what time interval was selected for this study, and why?

The time interval of 6 hours is used following the the tested values from Ellrod and Knox (2010). We also conducted sensitivity tests of the namelist parameters and found that the mixing of tracer was insensitive to the time interval. We added in the manuscript The time interval of DVT (nhours) is selected based on Ellrod and Knox (2010).

11. Line 129, How did you set t_{norm}? It is also strange that the turbulent flux (tracer transported per area per time) should depend on the model time step. Some explanation is needed before you proceed with testing this model.

The current used t_{norm} is an empirical value derived from a comparison to observations. We performed simulations with different t_{norm} and selected the t_{norm} that has shown to improve the ozone representation in EMAC compared to SWOOSH. A magnitude larger (weaker mixing) or smaller (stronger mixing) t_{norm} either leads to insignificant improvement or worsens the representation. With a different vertical and horizontal resolution, a different tuning on the empirical t_{norm} might be required.
We added in the manuscript The t_{norm} which moderates the strength of tracer mixing is an empirical value that shown to improve the ozone representation in EMAC (Figure 2b) against observations (SWOOSH). and also in the limitation subsection the namelist parameter C in the DVT is resolution-dependent and t_{norm} is an empirical value as well. Application of the CAT submodel to other model setups of EMAC or other climate models will require retuning.
For the dependency on the model time step. It is because the t_{norm} represents the physical time scale of the overturning by CAT between the two layers. And it needs the time step length to tell the model what fraction of the total mixing occurs during that specific time step.

12. Line 141, please explain "the lapse rate tropopause".

Changed to lapse rate tropopause (temperature lapse rate decreases to 2 K/km or less and stays on average below this value between this altitude level and any level above with the next 2 km). The whole section is now moved to the supplement.

13. Line 142, I assume Fig. 3 is produced from data for all three winters rather than Dec. 01, 2016.

For Fig.3, it is produced from the data of all 3 winters. For Fig.2, it is from Dec 01, 2016. The intention of Fig.2 is to show that both indices perform comparably on a specific time/event. The whole subsection is now moved to the supplement.

14. Fig. 4, I assume both indices are computed over the North Atlantic (i.e., 60◦ W to 0◦ W and 35◦ N to 60◦ N) ?

Yes, in Fig.4 both indices are computed over the North Atlantic. Fig.3 and Fig.5 as well. Now it is moved to the supplement.

15. Fig. 5, If "both indices show a similar performance under different thresholds" is the only conclusion to be drawn from Fig. 5, I would suggest moving this 5-subplot figure into the appendix.

We have now moved the comparison between Ellrod-Knox index and MoCATI into the supplement.

16. Fig. 6, "The contour lines represent the mixing ratio of O3 in ppmv of the respective QCTM simulation.". If so, which simulation did you use to plot the contour lines in subplot 6c?

The NOMIX simulation is used. However, now the original figure 6c removed, replaced with the comparison between SWOOSH and the mixing without the stability modification.

17. Line 162, "significantly reduced" might be an over-statement.

We now added an analysis of RMSE to support the statement.

18. Fig. 7b, is this the same as Fig. 6c?

Yes, but with a different range on the y axis. Swoosh only provides data between 50 to 250 hPa (range of Fig. 6c), Fig. 7b covers from 1 hPa to 600 hPa. Also, the original figure 6c is now being replaced.

19. Line 176, "changes of the gradient", what gradient are you referring to?

It refers to the tracer gradient. Corrected to changes of the tracer gradient.

Response to Reviewer 2

Major comments

1. First, I would like to clarify the major objective or aim of this study. In the abstract, or in lines 56-57, the major objective is deemed to be examining the sensitivity of UTLS chemistry and radiation on vertical turbulence mixing (as a side remark, I think it should be "the sensitivity of chemistry/radiation on CAT", instead of the sentence structure used in the manuscript). However, there are only two model simulations with one having CAT included and the other serving as the control. If sensitivity is the focus of this study, should it be examined by varying the strength of the parametrised mixing by CAT? Please clarify if the term "sensitivity" means different from what is suggested here and explain how this is studied with just the MIX/NOMIX simulation-pair. There is also another statement (lines 58-59), "We demonstrate the possibility of parametrizing CAT mixing based on turbulence indices". However, if this is the major take-away from this manuscript, there should be a rigid evaluation or comparison to observations to validate that the strength of the mixing is reasonable or realistic. Since there is only one MIX experiment with one set of parameters selected, the authors need to elaborate on why this particular setting of the parametrisation scheme is chosen and is meaningful. Otherwise, the scientific relevance of the final conclusions, e.g. the –0.2 W/m2 radiative forcing, is much diminished. Therefore, I urge the authors to clearly define the major goal of this study, or to put the study into a context, which shows that the current attempt is already a breakthrough (e.g. given the computational resources, such a parametrisation being included into a chemistry climate model is very difficult), in order to improve the scientific significance of the manuscript.

We rephrased the wording of this study throughout the whole manuscript. We think that "examining the potential impact of turbulence on UTLS chemistry and radiation " is the proper wording.
We further removed the statement "We demonstrate the possibility of parametrizing CAT mixing based on turbulence indices". We have now explained the selection of the parameters in the manuscript and comments below. We also added more analyses to show the improvement of the ozone representation after applying the CAT submodel and based on the improvement, showing the potential impacts of CAT on both the chemistry and radiation budget (see also the comments below) .

2. Second, I would like the authors to elaborate on their rationale for the CAT parametrisation scheme implemented. There are multiple assumptions behind this parametrisation scheme which require justification.
a) The flux of the tracer is assumed to be linearly proportional to the magnitude of the CAT index (equation 5). However, turbulence diagnostics like the Ellrod-Knox index, while having a certain degree of theoretical basis (e.g. the TI being linked to upper-level frontogenesis (Ellrod and Knapp, 1992), there is some empirical treatment (e.g. the constant C in the Ellrod-Knox index, equation 1 on page 5). Moreover, while stronger turbulence is assumed to be indicated by the higher magnitude of the turbulence diagnostics, the strength of turbulence is not assumed to be linearly proportional to the magnitude of the diagnostic (e.g. in Sharman et al. (2006), the mapping of indice to strength of turbulence is only piecewise linear; Sharman and Pearson (2017) use the log-normal distribution to map the diagnostic to eddy dissipation rate). Since there is no verification or comparison presented

in the manuscript to show how realistic the mixing is in the MIX experiment, this assumption has to be justified.

We agree that the assumption of linearization is the limitation of our studies. We now mentioned this limitation in the new subsection about limitations The current approach of the CAT mixing scheme assumes the tracer mixing flux is linearly proportional to the magnitude of the CAT index, while the strength of turbulence and turbulence diagnostics are only piecewise linear (Sharman et al., 2006).
We also included a more detailed analysis to show the improvement in ozone representation of the MIX experiment in Section 3.2.1 (Detailed response in Comment 3).

b) The MoCATI is a new diagnostic, which is claimed to include the effect of stratification on turbulence. However, as noted in line 122, mixing is turned off above certain Nlim. Physically, the quantity that has such a role in turbulence would be the Richardson number, in which a value greater than the critical Richardson number would be stable (no turbulence). Nevertheless, Richardson number also takes into account the vertical wind shear, not only stratification. Therefore, this modification to the Ellrod-Knox has to be justified.

We rephrased on defining the MoCATI. Instead of being a new turbulence diagnostic as the Ellrod-Knox index, which could be used to forecast CAT, we emphasised that the MoCATI is a modification for turbulent tracer mixing (specifically in global models) . Physically, given the same strength of MoCATI and the Ellrod-Knox index, differences in static stability should lead to a different strength in tracer mixing. We changed in the manuscript from a newly introduced Modified CAT Index (MoCATI) to a newly introduced Modified CAT Index (MoCATI) specifically modified for tracer mixing. and from It is a modification of the Ellrod-Knox index including static stability to It is a modification of the Ellrod-Knox index including static stability to moderate the mixing of tracer under stable conditions. This is also the intention of adding a stability correction for MoCATI. For the stratification correction, we try to avoid using the Richardson number because the shear is already taken into account by the Ellrod-Knox index, therefore only the Brunt-Vaisala frequency is used to avoid considering the shear twice. We agree that physically the Richardson number should be the quantity to switch off turbulence. However, in the stratosphere, under a stable condition, the N2 should be the dominating term of the Richardson number distant from the jets or the vortex. Therefore, we used the N2 for simplicity and also to reduce the potential computational time.

c) The important parameters, like C in equation 1, Nlim in equation 2, tnorm in equation 4, are only stated in the supplement without any description of how the authors came to this particular set of values. Are they tested for a range of values, and this set of values produces the most realistic tracer mixing? Or if the authors estimated reasonable values with some typical time or length scale/some other physical arguments? Or are they taken from other studies?

For the parameters, we performed a series of sensitivity tests with different combinations. The tracer mixing distribution is not that sensitive to the change of the parameters, except for N^2_{lim}. The other parameters mostly only affect the strength of the tracer mixing. It could be attributed to the diffusive advection scheme of EMAC. For C in equation 1, we used 0.025 after performing a PDF, the other values, like 0.1 or 0.01, from other studies, were either too strong or too weak compared to the vertical wind shear and deformation term. For N^2_{lim}, the value

of 6e-4 is taken from the previous studies from Chau et al. [2025], where mixing of tracer is still able to occur under such a stable environment. We also performed sensitivity tests with different N^2_{lim}, ranging from 2.5e-4 to 10e-4, above 6e-4 no longer leads to a difference in the tracer mixing distribution, but only different in the strength of tracer mixing. For the nhours, we took the value used by Ellrod and Knox [2010]. For t_{norm}, it is an empirical value specifically for the current setup that shows an improvement in the ozone representation (A detailed answer is in the comment Equation 4, lines 129-130).
We added in the manuscript a table and paragraph to explain the selected value The CAT submodel allows changes in different parameters via a simple namelist. An example namelist can be found in the electronic supplement. The values used in this studies is summarized in Table 2. The constant C in the DVT of the Ellrod-Knox index is selected after performing an analysis on the frequency distribution (Figure S3), the suggested values of 0.1 from Ellrod and Knox (2010) and 0.01 from Lee et al. (2020) will lead to either too strong or too weak DVT compared to the VWS and DEF. The time interval of DVT (nhours) is selected based on Ellrod and Knox (2010) and the limitation threshold (N^2_{lim}) is selected based on Chau et al. (2025). The t_{norm} which moderates the strength of tracer mixing is an empirical value that shown to improve the ozone representation in EMAC (Figure 2b).

d) The decision to apply the mixing from 500 to 70 hPa (lines 107-108) may exclude some tropopause folds that intrude deep into the troposphere, which are hotspots of STE (e.g. Stohl et al., 2003) and CAT (e.g. Rodriguez Imazio et al., 2022). Will the effect of these events be excluded due to this choice? Given that this parametrisation is the most important component of the study, it deserves a detailed and precise description.

We tried to apply the parametrization within the UTLS region. For the upper limit of 70 hPa, the environment will be so stable that not much turbulence will occur or the impact on tracers will be negligible. We therefore switched it off, considering the limited computational resources. For the lower limit, considering a global model using T42 (2.8 degree) (spectral) horizontal resolution, deep fold intrusions are unlikely to be properly represented. Such a detailed fold structure can only be found in a much higher horizontal resolution (maybe 0.1 degree).
We now added in the manuscript At pressure levels below 70 hPa, vertical tracer mixing by turbulence is strongly suppressed due to the strong static stability. At pressures higher than 500 hPa, deep tropopause fold intrusions might occur and lead to tracer mixing; however, such structures are usually not represented well in a typical EMAC setup with a coarse horizontal resolution (e.g. T42 and T63). Therefore, the CAT parametrization is switched off for pressure lower than 70 hPa and higher than 500 hPa, considering the limited computational resources and to avoid overlap with the boundary layer parametrization scheme.

3. Third, the authors should present better validation or evaluation of the performance of the new parametrisation scheme (that is based on MoCATI). For those presented in Sect. 3.1, they are mainly made between MoCATI and Ellrod-Knox index that are by definition (equation 2) highly correlated and dependent on each other. One can only see the effect of the additional "correction factor" based on stratification. The analysis done with TI being the reference is also confusing, as TI is yet another highly correlated turbulence diagnostic. Such analysis does not demonstrate the ability of MoCATI to realistically forecast CAT (TI is not a suitable "truth") or show that the additional modification to Ellrod-Knox index is necessary. Therefore, I find Sect. 3.1 redundant, which at most only shows the

internal consistency between MoCATI and Ellrod-Knox index (not surprising from the definition). In my opinion, the authors may directly argue that Ellrod-Knox index is a well-established CAT diagnostic (which only justifies its representation of the occurrence of CAT, not necessarily the strength of mixing), provided that the parametrisation produces realistic tracer mixing. However, concerning the performance of the parametrisation on the mixing of tracer, the only supportive evidence is Figure 7. The annual zonal mean ozone distributions of the simulations are compared to the satellite-based SWOOSH data, with the bias to observation in the MIX experiment being reduced. Nevertheless, the authors did not use it to help justify the validity of the parametrisation, but stated in lines 168-169 that such analysis (i.e. to see if CAT can improve representation of ozone in EMAC) is beyond the scope of the study. The authors should provide better arguments on the validity of the new parametrisation.

Our original intention is to demonstrate that the MoCATI (which is designed primarily for tracer mixing) shows a comparable performance to the Ellrod-Knox index, and we think the additional stability term is necessary in order to avoid over-mixing of tracer in the stratosphere. We agree that the analysis does not demonstrate the ability of MoCATI to forecast CAT realistically, but as we mentioned above, we treat MoCATI more as a modification specifically for tracer mixing, instead of forecasting CAT. Therefore, the analysis is more to demonstrate that the stability modification does not destroy the original Ellrod-Knox index. We agree that the comparison with Ellrod-Knox Index and ROC curves analysis leads to confusion, and it has now been moved to the supplement.
In terms of the necessity of the stability modification, we also performed simulations with simply the Ellrod-Knox index as a mixing coefficient. We compared this to observations (Swoosh) and found that this "overmixed" the tracers. This led to an underestimation of ozone in the stratosphere compared to SWOOSH, providing further observational indication, that our modification is applicable for global model resolution.
In terms of reducing ozone bias compared with the SWOOSH data, we agree that it improved the ozone bias in a climatological sense. We originally tried not to overstate it, considering we do not have a thorough analysis on the ozone representation on an event based and it is also not the intention of this study.
We now include the comparison of SWOOSH with the results without stability modification as well in section 3.2.1 and include an analysis of a normalized RMSE of the simulations.
We now added in the manuscript An extra simulation (QCTM-MIX-ENI) is conducted to show the necessity of the stability modification of MoCATI by using simply the Ellrod-Knox index as the mixing coefficient (the other model setup is identical). Figure 6c shows the annual mean difference of ozone between QCTM-MIX-ENI and SWOOSH. Without the stability modification, ozone is "overmixed" and leads to underestimation in many regions, especially the Northern hemisphere and the tropics. In order to evaluate whether CAT could improve the representation of ozone climatologically in EMAC, we calculated the normalized root mean square error (RMSE) of all 3 simulations (Table 2). QCTM-MIX shows the lowest normalized RMSE in most of the region except for the southern hemisphere. QCTM-MIX-ERI shows the lowest RMSE in the southern hemisphere, but shows a larger RMSE in other regions compared to QCTM-NOMIX. These results indicate that QCTM-MIX shows an improvement in all regions compared to QCTM-NOMIX, while QCTM-MIX-ENI improves the most in the southern hemisphere, but worsens the representation in other regions compared to QCTM-NOMIX. These results show that with the new mixing parametrization of tracers based on the static stability correction on the Ellrod-Knox index (MoCATI), the representation of ozone in the EMAC model is improved compared to SWOOSH in a climatological sense.

4. This comment concerns the major conclusions in Sect. 3.2.2 and 3.2.3. Since only one set of parameters for the CAT parametrisation is used to generate the MIX simulation, how confident are the authors in the changes to the tracer distribution or radiative effect? For instance, if the mixing is actually even stronger in reality (than that imposed by the parametrisation), would the distribution of tracer and the radiative effect be significantly different?

Yes, the radiative effect will be significantly different depending on the strength of the mixing. We performed the same simulation with t_{norm} a magnitude larger (weaker mixing), it still shows a similar cooling effect at the TOA, but significantly weaker, since the modification on tracers is also weaker. Therefore, we would expect a stronger radiative effect if the mixing is even stronger in reality.

5. Also, water vapour is stated as playing an important role in radiation, but it is left out in the parametrisation due to technical issues. A major conclusion that is highlighted would be the –0.2 W/m2 radiative effect at TOA. Will this number be changed (or even turn into a warming effect) if water vapour is included? If so, would the authors consider reminding readers of this limitation, like in line 14 "The global average radiative effect is about –0.2W/m2" and lines 253-254 "The simulation results show that the CAT mixing of tracers leads to a radiative cooling ... 208.9 mW/m2.

We are expecting a warming effect if we include water vapour. We now modified the manuscript from The global average radiative effect is about -0.2 W/m^2 to The global average radiative effect is about -0.2 W/m^2 without considering water vapour and from The simulation results show that the CAT mixing of tracers leads to a radiative cooling on the global mean radiation fluxes at the top-of-the-atmosphere (TOA) of about 208.9 mW/m^2. to The simulation results show that the CAT mixing of tracers leads to a radiative cooling on the global mean radiation fluxes at the top-of-the-atmosphere (TOA) of about 208.9 mW/m^2 without taking water vapour into account.
We also modified in the conclusion from The TOA is expected to be 0.2W/m2 cooler. to The TOA is expected to be 0.2W/m2 cooler without considering water vapour.

6. Concerning the above major comments, I suggest the authors consider having a subsection to discuss the limitations of this study.

We have now added a new subsection "Limitations of the current CAT approach" and stated Despite the improvements shown in Section 3.1, several limitations regarding the current CAT mixing scheme should be noted. The current approach of the CAT mixing scheme assumes the tracer mixing flux is linearly proportional to the magnitude of the CAT index, while the strength of turbulence and turbulence diagnostics are only piecewise linear (Sharman et al., 2006). It should also be noted that this parametrization is specifically tuned for the EMAC model under T42L90MA resolution, which is one of the most widely used resolutions, especially for long-term climatic studies including the stratosphere. For example, the namelist parameter C in the DVT is resolution-dependent and t_{norm} is an empirical value as well. Application of the CAT submodel to other model setups of EMAC or other climate models will require retuning. Also, water vapour is not treated as a chemical tracer in EMAC but as a thermodynamic quantity (specific humidity) and is therefore excluded in the current version of CAT in order to maintain identical model dynamics for comparison between the simulations under QCTM mode.

Minor comments

Lines 37-38: For the citation of Kelvin-Helmholtz instability being related to vertical wind shear, I think it is more appropriate to also include other studies. E.g. the classics like Richardson (1920), or Dutton and Panofsky (1970).

Included now in the manuscript.

Line 39: Dutton and Panofsky (1970) would be a suitable reference to link Kelvin-Helmholtz instability to the formation of CAT.

Included now in the manuscript.

Lines 73-74: Could the authors comment on the vertical resolution near the tropopause? Since vertical wind shear is deemed important, the vertical resolution is of relevance to a reasonable representation of the vertical wind shear.

The following line is added in the manuscript: The vertical resolution in the UTLS is approximately 400 to 500 m.

Lines 80-83: Could the authors provide some simple description of the different data sets? (RC1-base-07, Ref-C1, CCMI, AMIP-II)? Like the statement of "historical hindcast" for CCMI-2022 Ref-D1.

The description is add in the manuscript The chemistry setup is adapted from the RC1-base-07 of Jöckel et al. (2016), which is the Ref-C1 of the CCMI (Eyring et al., 2013), a free-running historical hindcast. and The sea surface temperature and sea ice concentration are prescribed by the AMIP-II data set (Taylor et al., 2000), the monthly means computed from model output, which agree with the observations.

Lines 115-117: The weighting constant C is likely to be resolution-dependent. In the supplement (Fig. S1), C has a magnitude of 0.1. How is this value set? Did the authors perform calibration for EMAC?

We performed a similar approach as Lee et al. [2020] by constructing probability density functions (PDF) for the DVT with different C to find the suitable scaling factor with a similar magnitude as the Ellrod index. The example namelist in the supplement S1 is not the value used in the simulation; we have now updated it. The C value used is 0.025. We compared the PDF of C = 0.1, 0.025 and 0.01. C = 0.1 is too strong compared to the Ellrod index and will dominate the whole Ellrod-Knox index. C = 0.01 is too weak, especially at the stronger values which identify as the occurrence of turbulence. The C is also found to be not sensitive to the mixing of tracer considering the diffusive advection scheme in EMAC.
We have now stated in the manuscript The constant C in the DVT of the Ellrod-Knox index is selected after performing an analysis on the frequency distribution (Figure S3), the suggested

values of 0.1 from Ellrod and Knox (2010) and 0.01 from Lee et al. (2020) will lead to either too strong or too weak DVT compared to the VWS and DEF..
We have now included the frequency distribution as Figure S3 in the supplement.

Lines 120-122: For the limitation threshold, did the authors perform analysis to support their choice of 6 × 10-4 s-1 (Fig. S1)?

We chose the N^2_{lim} based on the previous study [Chau et al., 2025], where mixing of tracer is still found in the region where the N^2 is below 6 × 10-4 s-1. We have also performed a sensitivity study with the limitation threshold, with the values above 6 × 10-4 s-1, it only affects the strength of mixing, but not changing the distribution of the mixing.
We added in the manuscript the limitation threshold (N^2_{lim}) is selected based on Chau et al. [2025].

Figure 1 and equation 3: In Figure 1, there are arrows in both directions. Is the flux from level n+1 to level n similarly parametrised? If so, it will give the net flux being (χn − χn+1) × CAT index (positive upward), which resembles the typical K-theory or flux-gradient theory for turbulent fluxes. Is this the case?

Yes. The difference would be how the mixing coefficient is calculated. We now added in the manuscript The mixing scheme of CAT follows a typical K-theory for turbulence fluxes and uses a turbulence diagnostic (CAT Index) to serve as a mixing coefficient.

Equation 4, lines 129-130: Did the authors perform analysis to choose a particular value for t_{norm}? Which reference data is used to determine if it really "moderates the strength of mixing"?

Yes. We performed simulations with different t_{norm}. The current used t_{norm} is an empirical value shown to improve the ozone representation in EMAC compared to SWOOSH. A magnitude larger (weaker mixing) or smaller (stronger mixing) t_{norm} either leads to insignificant improvement or worsens the representation. With a different vertical and horizontal resolution, a different tuning on the empirical t_{norm} might be required.
We added in the manuscript The t_{norm} which moderates the strength of tracer mixing is an empirical value that shown to improve the ozone representation in EMAC (Figure 2b). and also in the limitation subsection the namelist parameter C in the DVT is resolution-dependent and t_{norm} is an empirical value as well. Application of the CAT submodel to other model setups of EMAC or other climate models will require retuning.

Figure 2: Please check if the unit is correct. The Ellrod-Knox index in ERA5 reanalysis covers the order of magnitude from 10-7 s-2 to barely reaching 10-5 s-2 (Lee et al., 2023) at 250 hPa. As the output of EMAC is coarser than ERA5, I expect a smaller value for Ellrod-Knox index (since the derivatives are most likely smoothened). However, the plot is showing values mainly in the 10-6 to 10-5 s-2 values.

In figure 2, it is the output calculated by the ERA5 reanalysis, using the suggested value of C = 0.1 in the original literature. For the calculation in EMAC, we used C = 0.025 as mentioned and explained above. We recalculated the Ellrod-Knox index and MoCATI with C = 0.01 for the original figure 2 using ERA5 data, we obtained a similar plot with smaller values, the strongest region still lies within 10^{-5}, but considering it shows a specific mixing event and rest of the regions are within the lowest bound of the colorbar with the values of 0.15 $\times 10^{-5}$, which mean it is below the value of 1.5 $\times 10^{-6}$. (The original figure 2 now is moved to the supplement as figure S4.)

Lines 153-155: Could the authors clarify how the ROC analysis is performed? Am I correct that if TI exceeded a certain threshold (T1,2,3,4,5 in Figure 5), it is treated as the "ground truth" turbulence, then different values of the decision threshold of Ellrod-Knox index or MoCATI are used to verify against these TI-based "truth" to obtain the ROC curves? Apart from being unclear in this sentence, I am not convinced by using a highly related diagnostic (i.e. TI, on which Ellrod-Knox index is built) as the "truth" for CAT and verifying the performance of MoCATI or Ellrod-Knox index. Also, as mentioned in major comment 3, similar performance between MoCATI and Ellrod-Knox is expected, by definition. I therefore do not understand the inclusion of this analysis in the manuscript.

As we mentioned in comment 3, we treat MoCATI more as a modification specifically for tracer mixing, instead of forecasting CAT, and tried to demonstrate that the stratification correction does not destroy the performance of Ellrod-Knox index. To avoid confusion, we now move the ROC analysis into the supplement.

Lines 156-157: The effect of static stability is introduced into MoCATI, but whether it is realistic or not is not shown (being similar to Ellrod-Knox index does not support this statement).

As we mentioned in comment 3, we treat MoCATI as a modification for tracer mixing.

Lines 160-169: What is the major message for this subsection? To me, it shows that the introduction of the parametrisation seemingly alleviates the bias and improves the representation of ozone in EMAC (when compared to observations). However, it is stated at the end that this is not in the scope of this study. It is a bit confusing, could the authors clarify?

We agree that the parametrization improves the representation of ozone in EMAC in a climatological sense, we now also include an analysis of the RMSE.
We now added in the manuscript In order to evaluate whether CAT could improve the representation of ozone climatologically in EMAC, we calculated the normalized root mean square error (RMSE) of all 3 simulations (Table 2). QCTM-MIX shows the lowest normalized RMSE in most of the region except for the southern hemisphere. QCTM-MIX-ERI shows the lowest RMSE in the southern hemisphere, but shows a larger RMSE in other regions compared to QCTM-NOMIX. These results indicate that QCTM-MIX shows an improvement in all regions compared to QCTM-NOMIX, while QCTM-MIX-ENI improves the most in the southern hemisphere, but worsens the representation in other regions compared to QCTM-NOMIX. These results show that with the new mixing parametrization of tracers based on the static stability correction on the Ellrod-Knox index (MoCATI), the representation of ozone in the EMAC model is improved compared to SWOOSH in a climatological sense.

Lines 184-186: The presence of mountains may trigger mountain wave turbulence, which is sometimes treated as part of CAT. Therefore, the presence of mountains may promote the excitation of gravity waves and locally stronger vertical wind shear. Could the authors elaborate on this statement of having fewer mountains resulting in stronger vertical wind gradients?

The gravity waves are not resolved in the EMAC model; they are parametrized and only properly considered in the stratosphere for gravity wave breaking and driving of the Brewer-Dobson circulation and only have a minor influence on the wind velocity in EMAC and it is an independent process(submodel), and also they do not act on the tracers.
To avoid confusion, we changed in the manuscript from The changes in both gases are stronger in the southern hemisphere, potentially due to the stronger jet streams and fewer mountains, leading to a stronger vertical wind gradient. to The changes in both gases are stronger in the southern hemisphere, potentially due to the stronger jet streams, leading to a stronger vertical wind gradient.

Line 207, equation R3: This equation seems to me not balanced, please check.

Corrected.

Lines 232-234: It is said that "O3 is relatively more sensitive to NOx (less sensitive to VOCs) than without the CAT mixing". Is this concluded from the examination of FNR? FNR results are only mentioned later in the paragraph, so please consider rearranging the order.

This is first concluded by the results of the alpha, and the FNR shows a consistent result with it. Therefore, we mentioned later in the manuscript that It shows consistent results with the CH3O2, the increasing FNR indicates the O3 chemical regime is becoming relatively more NOx sensitive (less VOCs sensitive).

Lines 230 and 235: Is there any reason why 800 hPa level also responds significantly?

The 800 hPa changes could be attributed to the changed oxidation capacity due to the changes in HOx; while the changes in NOx is insignificant at that level, the more efficient methane oxidation leads to more HCHO and hence more HO2.

Line 243: Could the authors explain if the values of 7.31 to 7.24 years of lifetime for CH4 are averaged vertically? I cannot match these values to the "global" line in Figure 11c and therefore ask.

The values of 7.31 to 7.24 years are the global averages of all latitudes, longitudes, and altitudes. The lines are the averages at each vertical level. The lower troposphere and above the stratosphere have low values. The y-axis is in hPa, which does not represent the true ratio in terms of volume in the atmosphere. The current plot minimizes the region above the stratosphere. For example, the altitude above the stratosphere will occupy more space in a plot with altitude(m) as y-axis.

Line 260: The reference to Figure 7a should probably be 7b or 7c?

Corrected.

Lines 276-289: When comparing the results to Riese et al. (2012), what is the role of water vapour? Since water vapour is not included in the CAT mixing parametrisation in EMAC, would that be even more important than the fact that the turbulent parametrisation schemes are different between CLaMS and EMAC?

Riese et al. (2012) use Chemical Lagrangian Model of the STratosphere (CLaMS), which is a CTM. Thus, the water vapour does not change the circulation, while EMAC is a CCM, with the water vapour coupling to the dynamics through radiation and diabatic heating. In order to run EMAC as a QCTM, the feedback of water vapour on the dynamics needs to be excluded. We have also tested an extended version of CAT, which is now in preparation for publication, with mixing the water vapour under GCM mode with nudged meteorology. This still leads to an opposite result with the Riese et al. (2012). This is because the mixing parametrization of CAT in our approach primarily accounts for vertical mixing , while CLaMS does account for stirring and quasi-isentropic mixing, which dominates in the stratosphere .

Lines 314-317: Could the authors give a more detailed discussion on the possible effects of water vapour? Since it is mentioned to be of similar importance to ozone in terms of radiative effect, it would be great to inform the readers more about water vapour.

We note that a detailed analysis of water vapour is subject to a study on its own (in. prep.). We consider the reviewers comment and changed in the manuscript from Even though water vapour is not taken into account in the current version of CAT to obtain consistent dynamics in the QCTM mode, it is planned to also consider the potential radiative impact of water vapour by extending the CAT submodel and include the mixing of water vapour and temperature. The water vapour is expected to increase in the UTLS considering its gradient. to Even though water vapour is not taken into account in the current version of CAT to obtain consistent dynamics in the QCTM mode, it is planned to also consider the potential radiative impact of water vapour by extending the CAT submodel and include the mixing of water vapour and temperature. The water vapour is expected to increase in the UTLS considering its gradient. The moist tropospheric air will be expected to be mixed upward with the dry stratospheric air and may significantly change the radiative budget of the atmosphere.

Lines 317-318: For different turbulence diagnostics, a calibration is necessary (concerning the mapping of their magnitude to the strength of turbulence, as mentioned in major comment 2a). Also, it may be interesting to explore the "sensitivity" of chemistry and radiation to the uncertainty of turbulence diagnostics (as each of them is designed to target different processes or derived empirically by different methods or data).

We agree that different turbulence diagnostics need extra calibration.

We have changed in the manuscript from The CAT submodel is also expected to be extended to include and combine more turbulence diagnostics in the future, considering the pros and cons of each diagnostic. to The CAT submodel is also have the option to include or combine more turbulence diagnostics (e.g. Graphical Turbulence Guidance from Sharman et al. (2006)) in the future, considering the pros and cons of each diagnostic. However, potential recalibration would be required.

Technical corrections
The entire manuscript:
Please improve the consistency of the terminology used in the manuscript. For instance, EMAC is referred to as "chemistry climate model" in the title, but also "climate chemistry model" subsequently (also whether or not to hyphenate). Also, "Southern Hemisphere" is sometimes with upper and sometimes with lower case letters. Please make sure the acronyms are introduced before they are used. For instance, MESSy is only introduced in line 69, but already used in line 58. Also, "UT" is only used once, without introducing it.

Corrected

Line 300: typo of "ceO3"

corrected

References

C. H. Chau, P. Hoor, and H. Tost. Simulated mixing in the utls by small-scale turbulence using multi-scale chemistry-climate model meco(n). *Atmospheric Chemistry and Physics*, 25(20): 13123–13140, 2025. doi: 10.5194/acp-25-13123-2025. URL https://acp.copernicus.org/articles/25/13123/2025/.

Gary P. Ellrod and John A. Knox. Improvements to an operational clear-air turbulence diagnostic index by addition of a divergence trend term. *Weather and Forecasting*, 25(2):789 –

798, 2010. doi: 10.1175/2009WAF2222290.1. URL https://journals.ametsoc.org/view/journals/wefo/25/2/2009waf2222290_1.xml.

Dan-Bi Lee, Hye-Yeong Chun, and Jung-Hoon Kim. Evaluation of multimodel-based ensemble forecasts for clear-air turbulence. *Weather and Forecasting*, 35(2):507 – 521, 2020. doi: 10.1175/WAF-D-19-0155.1. URL https://journals.ametsoc.org/view/journals/wefo/35/2/waf-d-19-0155.1.xml.