Reviewer (Comments):

Review of " Tropospheric mixing and parametrization of unresolved convection as implemented into the Chemical Lagrangian Model of the Stratosphere (CLaMS)" by Paul Konopka et al.

Recommendation: Publication after revision

The paper is very well organised and written. The topic discussed in this paper, "Improving the state-of-the-art Lagrangian transport model for the stratosphere CLaMS by extending the transport scheme to the troposphere", is in general of high relevance. The reason is, that this would be one important step to enable climate and air quality modelling with the fully Lagrangian CLaMS transport scheme. The simulations with the extended CLaMS transport scheme are validated against satellite observations of the Outgoing Longwave Radiation (OLR) and in-situ observations of CO_2 in the UT/LS by CONTRAIL (Comprehensive Observation Network for TRace gases by AIrLiner observations).

The paper should be submitted after addressing the comments below.

General comments:

The goal to improve the tropospheric tracer transport has been reached with the extended CLaMS transport scheme using a heuristic approach for a better representation of tropospheric mixing and unresolved convection. This result is, as outlined above, very valuable and implicates a high potential for future application of the CLaMS transport scheme. However, to my point of view there are some things missing or at least should be better explained or motivated.

1.) ***Mass conservation in CLaMS***
For this topic, the focus is on the new parameterisation of unresolved convection, because the new tropospheric mixing scheme does not change the redistribution of mass compared to the actual reference version of CLaMS. Although, it is relevant to understand what the adaptive regridding is doing in terms of mass conservation. In this context, I highly appreciate the comments by Ingo Wohltmann and the conclusion by the authors to his comments: "We conclude that we certainly have to improve the explanation of our procedure in the revised version of our paper."
My suggestion would be to use the mass flux residuum of ERA-Interim as shown in Fig.1 in the answer to Ingo Wohltmann's comments motivating their heuristic approach (in section 1 or 2). Additionally, it would be very interesting to see also, how the mass flux residuum looks for the CLaMS reference simulations with standard adaptive regridding and for the control simulations with additional convective uplift (in section 4). The latter should show that the deficit in the range of 700 to 200 hPa has been reduced significantly.

2.) ***Validation of UTLS transport in CLaMS with in-situ CO_2 observations (CONTRAIL)***
CO_2 and especially the propagation of its seasonal cycle from the PBL into the UTLS is highly useful for model transport validation. Here, the authors use as benchmarks the in-situ CO_2 measurements by CONTRAIL aboard of passenger aircraft and the assimilated CO_2 data set provided by CarbonTracker. The latter is mainly constraint by surface measurements (see Table 1 in https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/). Therefore, the CONTRAIL data are the reference for CO_2 in the UTLS and the CarbonTracker data are the reference for the PBL.

My criticism is the way the CONTRAIL data are used for the evaluation in section 4.3. The CONTRAIL CO_2 measurements are zonally and monthly averaged between 2005 and 2008 and interpolated at a latitude-altitude grid with 10° by 1 km resolution and extending between 20°S to 60°N and 5.5 to 12.5 km. In the extratropics, this approach is highly problematic, because the mean (or the interpolated) CO_2 mixing ratios in the grid boxes will be strongly affected by (the irregular and sparse) sampling, especially above 7-8 km, because the individual probed air mass could be tropospheric or stratospheric. This matters for the months when cross-tropopause gradients are large, i.e. February to May in the NH. Also, the seasonal cycle of CO_2 is quite different below and above the extratropical tropopause. My suggestion is to filter out the stratospheric CONTRAIL CO_2 data to avoid this issue. This would not weaken the evaluation, because the CO_2 seasonal cycle in the free and upper troposphere is the relevant diagnostic for the introduced new tropospheric transport scheme in CLaMS. Also the representation of the CO_2 seasonal cycle in the tropical, subtropical (shown exemplarily) and extratropical free and upper troposphere should be discussed in a bit more detail. It is a known issue that modelled tracer transport from the PBL into the extratropical UT is often too weak, especially during summer.

Specific comments:

p.2, l.30: I think you mean here "…above the level of 300 hPa…" and not "…about 300 hPa…"

p.2, l.31-33: Why you use only the nominator – the buoyant production of turbulence – of the gradient Richardson number Ri to parametrise tropospheric instabilities and not Ri itself?

p.3, l.34-38: Just for curiosity, would the actual CLaMS transport scheme accumulate tracer in the PBL, if one uses emissions instead of prescribed PBL mixing rations?

p.4, l.5-7: "*Although we are aware of numerous convective schemes (e.g. Tiedtke (1989); Emanuel (1991)), our approach mainly intends to cover the range of possible variability due to unresolved tropospheric transport*."
It is quite unclear to me what is meant here, see also point 1.) in the general comments.

p.5, Figure 2: It seems that the TIL in the extratropics is (very) weakly pronounced compared to e.g. Birner et al. (2006). Is there an explanation?

Birner, T., D. Sankey, and T. G. Shepherd (2006), The tropopause inversion layer in models and analyses, Geophys. Res. Lett., 33(14), doi:Doi 10.1029/2006gl026549.

p.7, l.23-26: How the free parameter Nc, the critical limit for static stability, below which enhanced tropospheric mixing is triggered in the new parameterisation, is estimated? Which Nc value has been used for the CLaMS control simulations? How sensitive are the results to the choice of Nc?

p.8, l.1-2: I understand, that it is technically easier and better comparable to the previous CLaMS version, if the step of the additional tropospheric mixing is executed after the adaptive regridding, but does this also makes sense from a physical point of view? Would the result be different with the inverse transport operator: First step: mixing due to vertical instability and second step: mixing due to strong horizontal wind shear (deformation)?

p.9, Figure 4: A similar question: Why you implemented the vertical displacement in your new parameterisation for unresolved deep convection after the horizontal displacement? Is this realistic for deep convection?

p.9, l.11: Same question as for Nc: How sensitive are the results to the choice of the criteria $N^2_m < 0$ triggering convective events in the parameterisation?

p.11, l.1-2: "*It means that the mixing procedure is able to adjust a certain increase or decrease in the number of air parcels, but this amount should be below ± 10%.*"
This has to be better explained, see point 1.) in the general comments.

p.11, l.1-2: Again, how sensitive are the results of the control simulations to the choice of criteria that only ΔTheta > 35 K triggers convection? Would it not be better to use instead of a fix value the criteria Theta+ΔTheta > upper level of the PBL? This would mean that only convective events are considered that lift the air parcel out of the PBL, where the CO_2 mixing ratios of the air parcel will be overwritten by the prescribed value in the next time step anyway.

p.12, l.17-20: "*It should be emphasized…*"
What would this mean for simulations using emission fluxes rather than prescribed surface mixing ratios (see also my question p.3, l.34-38 above)?

p.13, Table 1: The critical values of the dry and moist Brunt Vaisala frequency used for the simulations should be added here. Also, it would be easier to use different names for both simulations FULL_EXT (e.g. FULL_EXT_0.3 and _0.7).

p.14, Figure 8: Please specify the simulations:
Top: FULL_EXT_0.3 or FULL_EXT_0.7?
Bottom: REF or REF-6h?

p.15, l.4: Boucher et al. (2009) is not really a good citation for stratospheric CO_2 and methane oxidation. The chemistry of methane in the middle atmosphere was to my knowledge first considered by Bates and Nicolet (1950). Early measurements of stratospheric CO_2 and CH_4 profiles date back to the 1960s and 70s, e.g. by Ehhalt.

Bates, R. D., and M. Nicolet (1950), Atmospheric Hydrogen, Publications of the Astronomical Society of the Pacific, 62(365), 106.

Ehhalt, D.H., The atmospheric cycle of methane. Tellus: 26, 58, 1974.

Ehhalt, D.H., L.E. Heidt, R.H. Lueb, and E.A. Martell, Concentrations of CH4, CO, CO2, H2, H2O and N2O in the upper stratosphere. J Atmos Sci: 32, 163, 1975.

p.15, l.16: How the reference simulations has been initialised on 1.1.2000? With an empty atmospheric domain – no CO_2? If so, the spin-up of only 5 years might be a bit too short for UTLS CO_2 analysis.

p.15, l.16: "*… all other control runs were started using the output of the reference run for the initial distribution.*"
This is a bit unclear to me. Does this mean that the control runs have started with the output of the reference run from 31.12.2004? If so, the results of the control runs for 2005, at least for the LS, is influenced by the prescribed distribution of the reference simulation. In the UT, this should only influence the first months of 2015, but still, this is not optimal for the comparison.

p.15, l.25: "*…on, the former being in better agreement with CarbonTracker.*"

I think, this statement is a bit misleading, because CO_2 assimilated from CarbonTracker data set is not the reference for the middle troposphere or UTLS, see also general comment point 2.).

p.17, Figure 10: Should be improved, see general comment point 2.).

p.17, Figure 10, Legend: According to Table 1, the name of the data set in the legend should be TROP_MIX and not VERT_MIX.

p.17, Figure 10, Caption: The altitude range in (b) is 10.5 km and not 15.5 km.

p.17, l.3-5: For the extratropics the gridding of the aircraft data has to be done in tropopause related coordinates or has to be filtered for tropospheric data, see general comment point 2.). Otherwise the averages are strongly biased by the sampling statistic of tropospheric and stratospheric air in the individual bins.

p.17, l.5-6: Has CLaMS simulations been sampled along the flight track?

Sec.4.4, p.18, l.10-11: This should be simply demonstrated by comparing REF vs. REF-6h.

Sec.4.4, p.18, l.16-18: This might be true, but I cannot really see the differences in AoA between Fig 11b and c. To demonstrate this, a difference plot would be necessary.