

## ***Interactive comment on “Diagnosing CH<sub>4</sub> models using the equivalent length in the stratosphere” by Zhiting Wang et al.***

**Anonymous Referee #1**

Received and published: 26 June 2017

### *General comments*

The manuscript intends to evaluate stratospheric mixing properties of three different CTMs by comparison with satellite measurements. For that purpose, the study makes use of the so-called equivalent length diagnostic,  $L_e$ , derived from stratospheric methane mixing ratios from MIPAS measurements and model output.

Usually I am pretty much in favor of short papers, provided they are written in a clear and conclusive manner and give all necessary information. Unfortunately, this is not the case here. In my view, the manuscript does not fulfill the basic requirements of scientific writing. From what is written in the text it is not at all clear to me what has been done, why has it been done, what is the goal of the study, what are the results, what is new, how do the methods and results compare to previous studies? For example, section 2

C1

does not include any details of the analyzed model simulations, except the resolution of the models. The presentation of the results in section 3 and 4 lacks a clear thread and line of argumentation. Furthermore, parts of the text are hard to read and would benefit from a better language. Some specific examples are given below.

Besides my criticism concerning the presentation quality I have some major concerns about the approach and applied method. The  $L_e$  diagnostic is used as measure of isentropic mixing, although the authors state in section 4.1 that diabatic processes also have an influence on methane and, therefore,  $L_e$  derived from methane mixing ratios. So in my view the described method is not appropriate to diagnose isentropic mixing and transport barriers. How do you want to make sure that the shown differences between MIPAS and the models are related to shortcomings in the simulated horizontal mixing and not to different representations of diabatic processes? Furthermore, from Wang (2016) I take that the models are not driven by the same meteorological data. As far as I understand the LMDz-PYVAR model is not at all driven by reanalysis data, but by the model's simulated dynamics. If this is the case, it does not make sense to compare individual years of observations with model output, but one would need long time series for a statistical analysis. In general, I have the impression that the whole analysis is mainly based on eyeballing, and I miss a quantitative analysis. This comes back to the question of the main goal of the study: Do you want to do a case study of a specific year or evaluate the model performance in general? Another question is related to the methane chemistry in the models. How is methane oxidation treated in the models? Are there substantial differences or shortcomings that could have an impact in the vertical gradient of methane in the stratosphere?

From the current paper draft I have a very hard time to judge whether there are any new ideas or results presented and whether the applied methods are valid or not. My first impression was to reject the paper. Nevertheless, I would give the authors a second chance to convince me of the relevance of their study, but in my opinion the manuscript clearly needs MAJOR revisions.

C2

### *Specific comments*

- No abbreviations like Jun. or Nov. in the text (e.g. L123).
- L125: replace “, Some...” with “. Some...”
- Sect. 2: As mentioned above I think the description of the models and methods is insufficient. Some examples: How is methane treated in the models (chemistry, emissions)? Which meteorological data are used to drive the transport models? Which years are simulated? 2009-2011? Is the model output treated in the same way as the satellite data, i.e. also interpolated to 1deg x 1deg? How about the vertical resolution? Are the model data sampled at the same vertical levels as the satellite data?
- L78/79: What do mean by this sentence? And which in situ surface measurements are used?
- L99-103: For calculating  $L_e$  the MIPAS data are first binned into a 4deg x 4deg horizontal grid and then interpolated to 1deg x 1deg? Do I understand this correctly? What is the purpose of interpolating the data to a finer grid? You cannot create more information from coarse satellite data by interpolation. How does this affect the calculation of  $L_e$ ? Please clarify. It might be helpful to add a schematic or an example of MIPAS on a specific isentropic surface. Are the model data treated in the same way?
- L109: I think  $\varphi_e$  is supposed to read  $\phi_e$ ?
- L116/117: I do not agree with this statement here. For example, the TM5-4DVAR model shows higher tropical methane mixing ratios at the top than MIPAS.
- L155: hpa -> hPa
- L172/173: I do not understand this sentence. How is vertical mixing related to the fact that methane decreases with altitude?
- L181/182: Here you state that the models do not capture the horizontally and vertically well mixed surf zone between 450-850 K and 60S-30S. How about the quality of the

C3

satellite data? How reliable are MIPAS data at these altitudes?

- L235: Why do the models do a better job in 2009 and 2011? As mentioned above a evaluation based on three years makes only sense when the models are all driven by (the same) meteorological reanalysis data.
- L254: Replace “sink down motion” by “downward motion”
- L265: What is meant by “jet exit region”? The edges of the jet streams?
- Discussion and conclusions: Your discussion of interhemispheric differences in gravity wave activity is not conclusive. Wave activity is expected to be stronger on the northern hemisphere than on the southern hemisphere. The same holds for the discussion of different transport schemes. Sensitivity tests using one CTM with the same transport scheme with different horizontal/vertical resolutions or the other way round would be helpful, but I see that this is out of scope.
- Fig 1: What is here, zonal mean or zonal median CH4 mixing ratios?
- Fig 2: Black thick and thin contour lines are hard to distinguish. I would suggest different colors for CH4 mixing ratios and zonal winds. Furthermore, it would be helpful to use a different, lighter color scale for  $L_e$ . Furthermore, why are there missing values around 1850 K for TM5-4DVAR in February 2010?
- Fig 3/4: Again, black contour lines are hard to read with a dark blue background.
- Supplement: I would like to see all the figures, that are intensively discussed, in the main paper and not in the supplement.
- Difference plots: I would recommend a different color scale, e.g. red-blue, for all figures showing differences. That makes it easier to identify positive and negative differences.