**General Comments:**

The paper by Dhomse and Chipperfield describes two new stratospheric data sets (TCOM-CH4 and TCOM-N2O), which were generated by a combination of TOMCAT model data and occultation measurements using a machine learning approach.
These data sets are unique in a sense that different satellite instruments were used to generate the merged long-term stratospheric data set.
The data are in general useful – as also written in the paper – for the evaluation of models and as a-priori information for retrievals.
Both data sets are publicly available for download (has been checked).

The paper is well written and contains all relevant information for data users except for some points addressed below.

I have the following general comments/questions:

1. The provided data sets are zonal averages. However, the zonal averaging is not explained in the method. Is it done before or after training/application of the correction term?
Please clarify in the paper.

2. It seems that in the machine learning approach the occultation measurements are considered as 'truth' for the training (and also later for testing/validation).
How are uncertainties of the measurement data considered?
How large are these, and how do they compare to the differences seen during testing/validation?
Retrieval errors are mentioned in the paper, but only in a general way without any quantitative assessment.
It is also unclear, if/how the vertical resolution of the measurements / averaging kernels etc. is considered in the method.
This should be addressed in the paper.

3. How relevant is the time dependence in the regression model?
Would it be possible to use this correction also for times not covered by the measurements (it seems so as the evaluation period is after the training period)?
What do you think are the limitations?

**Specific Comments:**

1. p. 5, l. 121:
Please explain SWOOSH.

2. p. 5, l. 130:
Note: 1 km is probably the vertical sampling of the ACE profiles; the vertical resolution depends on the averaging kernels.

3. p. 5, l. 133:
Limiting retrieval errors to <100% is a quite coarse filter. Since the correction method does not seem to consider measurement errors this may be an issue at least at higher altitudes.

4. p.5. l. 136ff:
*A 10° latitudinal overlap between the bins is allowed...*

   From the definition of the bins, it seems that the overlap is 20 degrees?

   Furthermore, the explanation *...to include possible extreme variations in the training data set* is unclear – why is the variation in the non-extended bins not sufficient?

As I understand, the following sentence *Estimated differences for overlapping grids are averaged...* actually does not refer to this step of the method (calculation of differences) but to later merging of the data in step 6.

These sentences should be re-formulated to clarify the above.

5. p. 5, eq. 1:
   Please specify what exactly is meant with 'time' in this equation. Is it the absolute measurement time (e.g. UTC time) or local time? In general, units should be specified for all quantities (e.g. does CH4 refer to number density or a volume mixing ratio?).

6. p. 6, l. 176:
   *which might introduce homogeneities*
   Do you mean inhomogeneities?

7. p. 7, l. 1ff:
   Are the feature importances normalised or not?

8. p. 7, l. 196/197:
   It seems that these sentences (esp. the value 18 km) refers to CH4 - please clarify in the text.

9. p. 7, l 199/200:
   *Hence attributing a single variable or a single processes is not possible.*
   The formulation is unclear (attributing to what?).
   In fact a similar statement is at the end of this paragraph.
   Please reformulate.

10. p. 8, l. 234/235:
    Please add a bit of summary information about Figs. S5 to S8, e.g. if results are similar for the other latitude bins or not.

11. p. 8, l. 236ff and Fig. 3:
    The vertical variation of the TCON-N2O profiles seems to be larger than in the original data (both observation and model), especially for the evaluation data set.
    Please explain.

12. p. 8, l. 245:
    Also, please add some general information about Figs. S9 to S12.

13. p. 8, l. 253:
    What exactly is meant with *unusual data points in 2004*?
    Do you refer to the larger values at 40 km?
    Please clarify.

14. p. 9, l. 2:
    *largest corrections are observed in the lower stratosphere*
    This is the case for the absolute corrections, not the relative ones.

15. p. 9, l. 259ff:
    Especially regarding N2O, how large is the error of the measurements at high altitudes / low concentrations? How reliable are the measurements at high altitudes?

16. p. 9, l. 261–263:
    *As the ACE-FTS retrieval algorithm uses multiple micro-windows, there may be a seasonal shift in averaging kernels causing fluctuations in the retrieved profiles.*
    Why do multiple micro-windows cause a seasonal shift?
    Do you mean that different micro-windows are used during different seasons?
    Please clarify.

17. p. 9, l. 263/264:
*As we use only positive data points for XGBoost training...*
Due to measurement uncertainties the occultation profiles may contain negative data points. If you only use positive data for training this might results in a bias.
Please clarify.

18. p. 9, l. 274/275:
Why do you show in Fig. 6 daily means for TCOM but monthly means for SPARC data? Wouldn't it be better to use for the comparisons in both cases the same averaging time interval?
Please explain.

19. p. 9, l. 274ff:
What is the difference between the ACE-FTS CH4 data set used in this study and the corresponding SPARC data set?
Please explain.

20. p. 10, l. 310ff:
Only a suggestion:
Maybe the comparisons between TOMCAT and TCOM should be described before the validation with independent data sets as they define the expected accuracy of the TCOM data.

21. p. 11, l. 338 and Fig. 8:
*for some years CH4 differences are clearly distinguishable.*
Please be more specific here.
Do you mean the occasionally high values in the tropics?
Actually, a lot of differences at 45 km seem to be below the lower range of the colour scale (-15%).
How representative are relative values at high latitudes where concentrations are low?
What is the reference for the relative values (TCOM or TOMCAT)?

22. p. 12, l. 370:
*A possible explanation would be strengthening of the stratospheric circulation...*
What do you want to explain here? A trend or a non-existent / negligible trend (as mentioned in the previous sentence)?
Please clarify.

**Technical Corrections:**

1. In general, please check the text for missing 'the' in the sentences.

2. p. 5, l. 125:
occulatation → occultation

3. p. 5, eq. 1:
I suggest that instead of the written quantity names (like 'temperature') variables should be used in this equation. Note that a sequence of italic letters in an equation could be (formally) misinterpreted as products of single variables.

4. p. 8, l. 229:
to -0.05 → to -0.05 ppm.

5. Figs. 4 and 5:
Please specify the CH4 and N2O unit in the figure or the caption.

6. Suggestion regarding the data sets:
   It would be good to have the unit of zonal mean CH4 and N2O not only in the global attributes of the data sets but also (or instead) in the attributes of the corresponding zonal mean variables.