Review van Gijsel et al, AMTD, 7, 2014

General comments

The identification of parameters that influence the data quality of atmospheric trace gas retrievals is a central question in many validation analyses. But it poses intrinsic challenges due to the multi-dimensional and non-linear nature of the problem. The authors of "*Using self organising maps to explore ozone profile validation results – SCIAMACHY limb compared to ground-based lidar observations*" propose a machine learning method, based on self-organising maps (SOMs), to uncover patterns in the quality of the retrieved profile data. They explain their method step-by-step, along with illustrations for the particular case of SCIAMACHY limb ozone profile data which is compared to lidar observations. The paper goes beyond a mere description of the method, as it continues with an interpretation of the results that lead to several conclusions on the variables that influence SCIAMACHY limb ozone profile data quality.

The proposed SOM-based method is novel and generally applicable to any comparison study of two vertical trace gas profile data sets, and therefore relevant for a wide audience in atmospheric sciences. In addition, the authors succeeded in providing a concise yet pedagogical description of their method, which should be accessible for readers not acquainted with machine learning methods. The scientific relevance and presentation quality are hence very high. However, the scientific quality should be improved before accepting this paper. The results shown for the illustrated case are not sufficient to motivate the author's conclusions on the relationship of various parameters with SCIAMACHY ozone data quality. Even stronger, I am concerned that the interpretation is at least partially wrong, as I explain in the specific comments below. When higher-dimensional and/or non-linear problems are tackled with sophisticated and powerful methods the user risks over-interpreting the results if no basic cross-checks are made, e.g. using traditional methods. Since this paper may open the door for many readers to start using SOMs as well, it is imperative that the authors provide an exemplary illustration not only of the method, but also of the subsequent interpretation of the results. The latter is not the case, according to me. I therefore recommend a major revision, in which the authors either provide further evidence in support of their claims, or refrain from the discussion of the interpretation of the SCIAMACHY results.

Specific comments

My main objection is the interpretation of the results in Section 3.4 (p4384 l25 – p4385 l12). Since these are repeated quite prominently in the abstract (p4375, L10-16) and in the conclusions (p 4388, L25 – p 4389, L13) these form an important part of the paper. In find that the authors interpret the altitude-dependent correlation of the explanatory variables (Figure 7) in a careless way. They state that with increasing altitude the quality of SCIAMACHY limb ozone (between 18 km to 36 km) depends mostly on longitude (*lon*), then solar zenith angle (*SZA*), then latitude (*lat*) and solar azimuth angle (*SAA*). The patterns in Figure 7 for these four parameters are similar, and in addition, it is quite clear from Figure 6 that they all correlate very well with the lidar station. The latter is also pointed out by the authors to some extent, in Section 3.3 (p4384 L2-6). According to me, there is therefore a more natural explanation for the patterns in Figure 7, namely that the codebook vectors correlate

mostly with lidar station (which can unfortunately not be computed). In this case, the patterns in Figure 7 are a result of the correlation of the four EVs with the station's location. To sort this out, it would be instructive to show

- Correlation plot of codebook vector versus (lat,lon,SZA,SAA) at various altitude levels, where each marker is coloured according to the dominant lidar site.
- At the beginning of the analysis, a plot of the altitude-dependence of the (median/mean +- spread) normalized differences at each of the seven lidar sites.
- A correlation plot of each pair of lat, lon, SZA and SAA would help, with markers coloured by lidar site.

If these plots do not provide support for the author's interpretation, the corresponding parts of the paper need revision. In a similar way, Figure 10 may be overinterpreted (Section 3.4 and the conclusions), and a simpler fundamental explanation is possible.

A second important concern is the representativeness of the SOM training sample. The definition of the training sample is a very important step in any machine-learning based-method. This should be stressed out more in the paper, and I would welcome some more motivations behind the choices made by the authors. I suspect that the comparison sample is not homogeneously distributed over the lidar stations. Figure 6 e.g. shows that three sites dominate most neurons (Hohenpeissenberg, then Mauna Loa, and finally Lauder). Is it possible that the bulk of the data in the training sample corresponds to these stations? Could you e.g. add the number of pairs per site in Table 1? I do not see how one can arrive at representative conclusions with a non-representative input sample. Coming back to the remarks in the previous paragraph: it is possible that this is at the heart of the patterns seen in Figure 7. How can you be sure that SCIAMACHY limb ozone data quality truly depends on lat, lon, SZA and SAA? The training sample could be made more representative by restricting the same number of pairs per lidar site. And even then, it is not a trivial task for the specific example given the limited collocation statistics of satellite-ground comparisons. The method therefore seems to have more potential for satellite inter-comparisons where, in some cases, much higher collocation statistics can be gathered and more representative training samples can be obtained.

Technical corrections

P4375

- L17: exploring → exploration
- L21: coupling → merging

P4376

- L5: inter-cluster → inter-class (?)
- L21: remove "used"

P4377

- L1: increased → increases
- L5: proposed → considered
- L15: Sect. 2 → Section 2

- L21: remove “proposed”

P4378

- L1: SCIAMACHY “limb”
- L3: launched “into space”
- L17: remove “which is”

P4379

- L11: increasing → degrading

P4380

- L2: are → is
- L4-8: how exactly is the mapping done? How is the “distance” defined between a difference profile and a codebook vector?

P4381

- L10: not at all → not all
- L14-15: put “with the matching metadata” between brackets

P4386

- L15-16: remove phrase “The codebook vectors … by the indices”. It is a repeat of previous paragraph.

P4387

- Last paragraph: double check the naming of cluster 1, 2 and 3. Do they really correspond to what we see in Figure 10? It appears to me that it is the other way around.

P4388

- L15: After that the role → After that, the role

Table 1

- Please add number of collocations per site.

Figure 1

- L2: trained used a → trained on a
- L3: the SOM consists of “a” grid formed

Figure 5

- Add unit [%] in caption. Please clarify in the main text or in the caption whether the orientation is identical as for Figure 6 and Figure 9. I assume they are, but the aspect ratio suggests the opposite.

Figure 7

- Reduce the number of labels on the Y-axis.

Figure 9

- Clarify that orientation is as in Fig 5 & 6.

Figure 10

- Is the naming of the clusters the same as in the main text?