Answer to comment from referee n. 1 on "Validation of satellite SO2 observations in northern Finland during the Icelandic Holuhraun fissure eruption"
by I. Ialongo et al.

We thank the referee n.1 for his useful and constructive comments. We provide a point-to-point answer here. The comments from the referee are in Italic while the answers from the authors are in Roman.

This paper provides a comparison of results from different SO2 total column products of the OMI and OMPS satellite instruments as well as for ground-based Brewer direct sun measurements at Sodankylä (Finland) during the Holuhraun fissure eruption (Iceland) starting in September 2014. Additionally, SO2 surface concentrations from four air quality stations in Northern Finland are presented, supporting that the volcanic plume was located at low altitudes over Finland.
The paper is well written. However, while the authors suggest to show "validation" results, I would rather consistently use the word "comparison".

We now refer to "comparison" or "evaluation" instead of "validation" as suggested by the referee and we modify the title as:
Comparison of operational satellite SO2 products with ground-based observations in northern Finland during the Icelandic Holuhraun fissure eruption.

It is mainly shown that the satellite as well as the ground-based measurements support the presence of the volcanic plume in the lower part of the atmosphere and that (at least for some observations) the variability of the SO2 columns is similar. While a comprehensive "validation" of operational OMI/OMPS SO2 products and ground-based data generally should be strongly endorsed (the literature on this topic is quite sparse), a main focus of the paper should be on the general uncertainties that come along with the operational processing of the satellite data as well as other problems like e.g. differences in spatial/temporal resolution. These problems are only qualitatively discussed and illustrated by showing results for different a priori assumptions of the SO2 profile for the satellite measurements. Especially for SO2 at low altitudes, the actual profile (plume altitude) has probably the most critical influence on the final column densities.
Generally, it is therefore questionable if the operational satellite products (in this case especially the PBL products) should be used instead of conducting detailed sensitivity studies using radiative transfer model calculations adjusted to the actual set-up. I think this paper has the potential to be published in AMT, but after major revision:

We stress that the focus of this paper is the evaluation of the operational SO2 products at high latitudes. These products are important for several applications, such as hazard mitigation and volcanic eruption alert services (particularly direct-broadcast data). Also, the comparison between different available products is a new aspect presented in this paper.

We now add a separate section (3.3 Analysis of the uncertainties) including the uncertainties of the operational retrievals and the errors expected from using the operational assumptions at high latitudes. We add also two figures (S4 and S5) to the supplementary material including the AMF radiative transfer calculation for different SZA/VZA values and the averaging kernels to describe the effect of the assumed SO2 profile.

Specific comments:
p. 600, l. 8-9:
1.1) How are the "challenging retrieval conditions" actually considered? Especially the OMI BRD PBL product assumes a constant AMF of 0.36 according to an older version of the README file (http://SO2.gsfc.nasa.gov/Documentation/OMSO2Readme.doc). A constant SZA of 30°, VZA=0°, surface albedo=0.05 and surface pressure are assumed. This geometrical assumptions may lead to significant under/overestimation of the SO2 column, especially for high latitudes where the AMF of 0.36 is likely to be strongly overestimated (and therefore the SO2 column to be underestimated – as mentioned by the authors). The presumed settings for the OMPS PCA PBL product are the same, right? As already mentioned, I think it's questionable if these operational products should be used for a "validation" exercise. Some of the parameters indicating "challenging retrieval conditions" are listed in Table 2, but it is not really clear if the associated uncertainties are larger than e.g. the actual satellite ground pixel size or SO2 fit error. How large are the associated errors?

We mean here that we want to evaluate the applicability of the existing products for high latitude conditions (high SZA). We rephrase this sentence in the abstract as:
"The operational satellite SO2 products are evaluated for high latitude conditions (e.g., large solar zenith angle, SZA)."

We also add a discussion about the AMF assumptions in PBL products for high latitude conditions in the new section 3.3 of the revised manuscript (together with an analysis of the uncertainties). The same settings are considered in both BRD and PCA algorithms. Also, we add Fig. S4 in the supplement in order to show how AMF decreases with increasing SZA.
On the other hand, we need to mention that the retrieval under high solar zenith conditions is challenging due to low earthshine radiance received at the satellite, thus larger impacts of instrumental effects (such as stray lights and other spectral artifacts) are expected on the retrieval results. Usually these artifacts lead to biases, which can be quite significant, exceeding those from measurement noises and retrieval errors due to algorithmic assumptions. Consequently improving retrievals with more accurate AMFs or other retrieval assumptions may not improve enough the agreements with ground-based measurements.

This discussion is also added in section 3.3.

1.2) According to the README file, only OMI pixels of the rows 5-55 should be used for the PBL product because of the increased noise (probably because of the large *pixel size and especially viewing zenith angle). Half of the OMI measurements listed in Table 2, however, belong to rows 1-4 and 56-60 (all of them have CF>0.3 anyway). Maybe it would be good to highlight all other measurements or just exclude all measurements at the edge of the swath in Table 2.*

We now highlight the overpass corresponding to small pixels in Table 2. We also separate the large and small pixels explicitly in the additional figures (S2 and S3, for OMI and OMPS PBL products, respectively) in the supplement. We prefer to keep all the pixels in the comparison in order to evaluate also the performance of the different algorithms at the edge of the swath (especially because of the several missing data from the center of the swath due to the row anomaly). As for OMI, also OMPS central pixels (4:33) are considered the most reliable and they are highlighted in bold in Table 2.

2) p. 602, l. 12-13:
It would be good to add some words about other problems (e.g. spatial/temporal resolution, unknown profile, sensitivity, clouds).

By challenging we meant here that: "The opportunities to validate volcanic SO2 satellite products are rare, because only occasionally the volcanic plumes drift over a ground-based station where SO2 measurements are performed."
We add this sentence in the manuscript.
We discuss the challenges in the satellite retrieval in section 3.3 of the revised manuscript.

3) p. 602, l. 12-13:
I think the results from your paper show that this is only the case when you have access to data of a widespread measurement network and not only for just one instrument (as it was also mentioned at the end of Carn and Lopez, 2011).

Please, see point 2

4) p. 602, l. 25-27:
What possible reasons were mentioned in Rix et al. for these "large" discrepancies? Are your measurements affected by the some problems?

We add this sentences: "Part of this difference was due to the fact that the Brewer data were available as daily averages while the GOME-2 measurements represent a snapshot at the time of the overpass. Furthermore, differences can be caused by uncertainties in both satellite and Brewer observations."
We also further discuss the uncertainties on the satellite retrievals in Sect. 3.3 and we add some details on the quality of Brewer measurements in Sect. 2.2.

5) p. 604, l. 22-25 and p. 612 l. 7-11:
What is the advantage of using the OMI DB data here? Although it is interesting to know that the data is in principle available, the DB data for the PBL scenario (like the PBL SP data) is processed using the BRD algorithm and the main difference is claimed to be the sliding median background correction. However, although the resulting differences between the PBL DB and SP data are mostly small, they're up

to about 50% for specific measurements (e.g. 6/9/2014, 9:03, 2.59 vs. 3.86 DU). I think such significant differences for the very same algorithm/instrument clearly point out that you have to put more effort in error analysis and discuss the possible reasons/uncertainties.

We add this text about the differences between DB and SP in section 2.1:"The DB algorithm uses the ``latitude band average'' as residual correction method, while the operational algorithm uses the ``sliding median'' technique, which requires a complete orbit to perform the correction (Yang et al., 2007). Because of these different methods and the observed difference of L1B data between DB and routine processing, differences between SO2 DB and SP products are expected. Assessing the quality of the DB retrievals is also important as they are used for volcanic emission real time services and aviation hazard mitigation (e.g., SACS)."

6) p. 607, l. 5-6:
What exactly is meant by "sufficiently continuous"?

We replace this with: "sufficient amount of".
The remaining days of September showed only sparse Brewer measurements during the day.

7) p. 607, l. 11-13:
The different satellite products are not "sensitive to different altitude regions". The slant columns are just converted into vertical columns by using AMFs for different a priori profiles.

We replace this with: "with different a priori profile assumptions"

8) p. 607, l. 14-15:
It should be already mentioned in Section 2.1 that the DB products are processed by using the BRD algorithm (but uses a different background smoothing correction...what else are the differences?).

Yes, this will be moved to Section 2.1
This is the most critical difference, together with some observed differences in the Level 1 observations.
Please see point 5)

9) p. 607, l. 21-22:
While it was reported that the volcanic plume was close to the ground by several other groups, the satellite SO2 SCD measurements are also likely to be significantly lower due to dilution, as only the average SCD is detected within large satellite ground pixels while the Brewer measurements are local point measurements. Probably the VCDs are closest to the Brewer measurements for the PBL product because of the associated small AMF.

Instead of this sentence the dilution effect is added here:

"Furthermore, one must note that the satellite retrievals are expected to be lower than the Brewer values due to dilution, as the average SO2 columns derived within the relatively large satellite pixel are compared to the local point measurements from ground-based observations."

10) p. 607, l. 23-26:
As mentioned before, due to the constant AMF of 0.36, the PBL product is probably not well suited for a "validation" (especially for large SZA/VZA).

We now discuss the error associated to the SO2 vertical column due to the AMF assumption in Section 3.3 of the revised manuscript.
We avoid anyway the word "validation" in the revised paper.
See also point 1.1).

11) p. 608/609 and Fig. 2:
Please add error bars. How large are the uncertainties? Are all satellite measurements exceeding the detection limit? Maybe it would be good to show local satellite maps of the PBL product scaled to the maximum detected SO2 VCD to get a feeling for background noise and uncertainties.
The different products for the cloudy measurements are hard to distinguish. Maybe it would be good to think about using other symbols.

The satellite retrieval uncertainties are discussed in a separate section (3.3 in the revised manuscript), as mentioned before. The information about the detection limit is also introduced in the same section, after the calculation of the precision.
We separated the observations based on the cloud/SZA information in Fig S2 and S3 in the supplementary material in the revised manuscript (as also suggested by referee #2). These figures include all overpasses within 60 km from Sodankylä and give also an idea about the background noise around Sodankylä.
We keep this figure 2 as is, in order to keep the retrievals derived for different SO2 profile assumptions summarized in the same plot.

12) p. 608, l. 1-8:
The PCA PBL result is significantly lower than the BRD PBL result. What are possible reasons?

Because the PCA algorithm uses the entire spectrum in the SO2 fitting to reduce interferences from instrumental or geophysical effects in general it was found PCA SO2 results to be smaller than BRD, particularly for high latitudes.
This comment is added to the text.

13) p. 608, l. 9-17:
While the measurement conditions are quite similar for the previous day, the data for the satellite and ground measurements agree less. What are the possible reasons?

We would say that the data agree more or less in the same way in the sense that satellite data are smaller than Brewer. Actually the SO2 column from PCA is in better agreement to the ground-based data than the 5 September and also very close to the BRD value.

14) p. 608, l. 18-23:
The largest data gap for the Brewer measurements appears right where the largest SO2 CD is expected? Is there a specific reason for that or is it just coincidence?

The data gap around 9-10 UTC is just connected with the variable cloud cover during that day. Direct sun measurements are not shown in case of high air mass values (after 14:20 UT).

15) p. 608, l. 24-29:
The PCA PBL and BRD PBL product show a large discrepancy. Again, the PCA value is much smaller and close to zero. What are the possible reasons?

Please, see point 12).

16) Section 3.3:
While it is interesting to see the data for the air quality stations, I don't really see how this adds something to the "validation" process. No Brewer measurements are presented (and available) and the main conclusion is that the volcanic plume was located at low altitudes over Finland.

Despite satellite vertical columns and ground-based surface concentrations are not quantitatively comparable, the observed spatio-temporal link between high SO2 concentration values at surface and large total columns from satellite adds confidence in satellite-based observations for volcanic emission monitoring also at surface levels. In particular, satellite instruments show their capability to detect the position of the volcanic plume as compared to independent ground-based observations.

We will add this comment in section 3.4 of the revised manuscript.

17) p. 612, l. 23-25:
Is this only the case for the PBL products (because of the constant AMF) or for all satellite products?

We refer here to PBL products. On the other hand, the LF algorithm accounts for the actual observation conditions and has no inherent bias under high solar/viewing zenith angles.
We add this comment to the text too.

Minor comments:

18) p. 601, l. 22: observation systems

This is replaced with "services".

19) p. 604, l. 4: Please add:

Carn, S. A., Yang, K., Prata, A. J. and Krotkov, N. A.: Extending the long-term record of volcanic SO2 emissions with the Ozone Mapping and Profiler Suite nadir mapper, Geophysical Research Letters, DOI: 10.1002/2014GL062437, 2015

Yes, this paper came out after the manuscript was submitted. Now it is added in the reference list.

20) p. 604, l. 22: products

corrected

21) p. 605, l. 23-24: Maybe good to mention that Kagoshima is located right next to a very active volcano (Sakurajima).

This was added

22) p. 606, l. 22-24: Probably most important: There's no sun light during winter at high latitudes

This was added as:
"Furthermore, no observations are available during the deepest winter time because of the reduced sun light hours at high latitudes."