Our responses to review comments (repeated in *italics*) are given below in **red**.

Response to Reviewer 2

This paper presents a comprehensive model intercomparison of the impact of bromine containing VSLS on the stratospheric bromine loading. This is a good initiative and the outcome of this intercomparison will be important of assessing the impact of bromine on stratospheric ozone. It is in particular noteworthy that a lot of observations are employed to assess the quality of the model results.

We thank the Reviewer for his/her comments on our manuscript. We are pleased that he/she finds the work important and acknowledges the large body of atmospheric observations used throughout.

I have some points (see below), where I think the discussion in the paper can be clarified and improved. The impact of particular model features (e.g., the convective schemes employed) on the different results could be brought out more clearly. The reader ultimately will be interested in what the problematic model features are, because these are the features that need improvement in the further developments of such models. This point cold be brought across in the paper in a better way. In summary, I think that a revised version of the paper, taking into account the points raised in the reviews will be a valuable contribution to ACP.

We have addressed the review comments and believe the paper has been strengthened accordingly. We are pleased the Reviewer thinks this work will make a valuable contribution to ACP.

Detailed comments

Five out of the 11 participating models are nudged to or driven by ERA-Interim. While ERA-Interim is a good choice, this fact will lead to the multi-model mean being biased to an ERA-Interim world. I suggest to bring this point across more clearly. Does this fact have any implications for the conclusions of this model intercomparison?

It is indeed the case that most models use ERA-Interim met fields, but as you can see other models (ACTM, NIES or the free running models) do not produce MAPE very different from those produced by the ERA-interim driven model. In fact there is quite good agreement between the major reanalysis products from ECMWF, JMA and NCEP. This was very recently highlighted by Harada et al., (2016) and is clearly shown in Figure 4 of that paper: see: http://jmsj.metsoc.jp/EOR/2016-015.pdf. Thus we do not believe at this stage that our model spread is hugely biased towards ERA-interim. In the revised manuscript we have commented on this, as requested, in the first paragraph of Section 2.3.

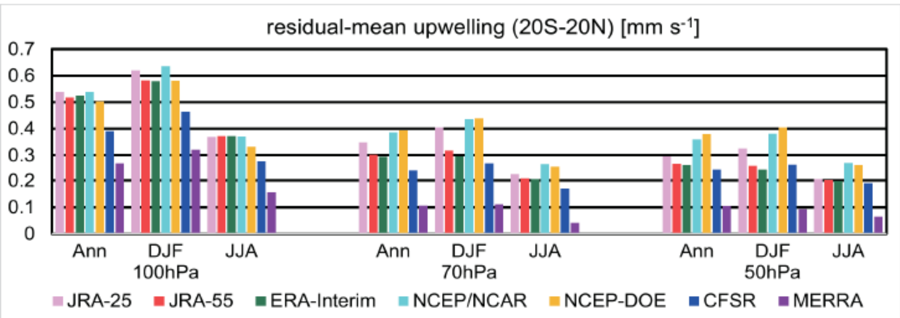

Figure 4. Annual and seasonal mean residual-mean upwelling within 20° of the equator, averaged from 1979 to 2012.

Figure from Harada et al., JMSJ, 2016

Another model feature, which is important for tropospheric transport of VSLS is the convective parametrisation used in the model (see for example Rybka, H. and Tost, H.: Uncertainties in future climate predictions due to convection parameterisations, Atmos. Chem. Phys., 14, 5561-5576, doi:10.5194/acp-14-5561-2014, 2014, and references therein). I suggest more discussion of this point in the paper. Also, the information of the convective scheme used in the different models should be included in Table 2. Perhaps some of the model differences and some of the model similarities can be attributed to using a particular convective parametrisation or a particular meteorology?

As per the Reviewer's suggestion, we have extended Table 2 to include the convection and boundary layer mixing schemes used by each model. We did not find any systematic differences in our results related to the choice of convection scheme or input meteorology. We have now made this point in the abstract:

"Overall, our results do not show systematic differences between models specific to the choice of reanalysis meteorology, rather clear differences are seen related to differences in the implementation of transport processes in models".

Related to the above, and following comments from Reviewer #1, we have made clear that differences between CTM setup and the implementation of transport processes is important, as all models would claim to be simulating the above physical processes. Further, we now define what we mean when referring to "transport" differences early on in the manuscript (last paragraph of Introduction):

"... we define *transport* differences between models as the effects of boundary layer mixing, convection and advection, and the implementation of these processes. Note, the project was not designed to separate clearly the contributions of each transport component in the large model ensemble, but can be inferred as the boundary layer mixing affects tracer concentrations mainly near the surface, convection controls tracer transport to the upper troposphere and advection mainly distributes tracers horizontally (e.g. Patra et al. 2009)"

Finally, we have expanded the discussion of the role of convection in the manuscript. In addition to new text placed in Section 3.1.2 (suggested by Reviewer #1), we have added discussion in Section 3.4 relating to model differences shown in Figures 14 and 15. We also cite the *Rybka and Tost* paper.

"The high altitude model-model differences in $CHBr_3$, highlighted in Figures 14 and 15, are attributed predominately to differences in the treatment of convection. Previous studies have shown that (i) convective updraft mass fluxes, including the vertical extent of deep convection (relevant for bromine SGI from VSLS), vary significantly depending on the implementation of convection in a given model (e.g. Feng et al., 2011) and (ii) that significantly different short-lived tracer distributions are predicted from different models using different convective parameterisations (e.g. Hoyle et al., 2011). Such parameterisations are often complex, relying on assumptions regarding detrainment levels, trigger thresholds for shallow, mid-level and/or deep convection, and vary in their approach to computing updraft (and downdraft) mass fluxes. Furthermore, the vertical transport of model tracers is also sensitive to interactions of the convective parameterisation with the boundary layer mixing scheme (also parameterised) (Rybka and Tost, 2014). On the above basis and considering that the TransCom-VSLS models implement these processes in different ways (Table 2), it was not possible to detangle transport effects within the scope of this project. However, no systematic similarities/differences between models according to input meteorology were apparent".

I also have reservations about the concept of a "preferred" tracer. I think this means that the emission inventory somehow interacts with the transport scheme of the model to produce reasonable results at higher altitudes. But this means that the higher altitude agreement could be right for the wrong reason. I know it is demanding a lot from models, but of course one would *expect to design independently the best emission inventory and the best (vertical) transport to obtain the best agreement with measurements. Obviously this model intercomparison cannot achieve this goal, but I think the discussion of these issues could be improved.*

The overarching goal of the work was to calculate a climatological multi-model best estimate of stratospheric bromine SGI from $CHBr_3$ and CH_2Br_2. It was essential, of course, for this estimate to be based on simulations that provide the best possible model-measurement agreement at the surface. Given good surface agreement, the models' transport of $CHBr_3$/CH_2Br_2 from the surface to higher altitudes, against that observed, has been tested. The fact that models do not necessarily agree as to which emission inventory "performs best" at all surface sites (against measurements) we believe is an important finding of this work. It has implications for model studies attempting to quantify the global flux of VSLS to the atmosphere and, in particular, for studies attempting to reconcile such estimates obtained from different models. We have added a sentence to the revised manuscript in Section 3.1.2 (end of 3rd paragraph), where the above is discussed, to make the latter point more clear:

"Ultimately, attempts to reconcile estimates of global VSLS emissions, obtained from different modelling studies, need to consider the influence of inter-model differences, as discussed above."

The discussion of "preferred tracer" has also been clarified in Section 3.2 in response to a comment from Reviewer #1.

Finally, the impact of ENSO activity on the stratospheric bromine loading is unclear. What is the message of the paper here? The paper states that there is a strong correlation of SGI with ENSO (e.g. abstract), but that there is no correlation of ENSO (MEI) with the bromine loading in the LS (e.g. conclusions). But SGI is important for the bromine loading in the LS. This points needs to be clarified and better discussed in the paper.

OK, we have clarified this. Our results show that (i) SGI is enhanced over the East Pacific during strong El Niño conditions (e.g. in 1997/1998 as can be seen in Figure 17). Related to this (ii) SGI is strongly correlated to MEI over the East Pacific (where significant SST warming occurs under El Niño conditions) but (iii) averaged over the whole tropical domain, there is little correlation between SGI and ENSO. The latter point is because of the zonal structure in SST anomalies (and therefore convective activity) associated with ENSO activity. Essentially, the effect of warming and intensified convection in some areas (i.e. East Pacific) on stratospheric Br SGI can be cancelled out by the cooler SSTs in other tropical regions. Aschmann et al. (2011) performed a detailed analysis of how bromine SGI is affected by ENSO and indeed reported this complex zonal structure. We have now clarified these points at the end of Section 3.5 of the revised manuscript.

Minor issues

- *Title: I am not sure if "TransCom-VSLS" should be in the title; the name of the project will not be relevant on a timescale of years, when the paper will still be read.*

Since the concept and experimental method are chosen from previous TransCom experiments, it gives a link to the paper's evolution. Thus we would prefer to keep this term in the title.

- *l. 7: I do not think that model estimates should be used to "constrain" measurements.*

We are referring here to using models to help constrain the current SGI range.

- *line 20: change 'optimal' to 'best'*

OK. We have done this.

- *l. 36: Isn't 6 month a bit long for very short lived?*

We agree that intuitively it does seem long for a “very short-lived” compound. However, this is the definition used in previous WMO Ozone reports. VSLS local lifetimes at the surface can vary substantially in space and time, though the <6 months rule is broadly accurate.

• *l 51: ‘recent’ twice in this sentence*

OK. A “recent” has been removed.

• *l. 52: try nmathrm{VSLS} to avoid italics in VSLS. (Similar for MAPE (l. 345) below).*

We thank the Reviewer for this suggestion. The italics have been removed from “VSLS” in Br_y^{VSLS} throughout the manuscript.

• *l. 59: ‘owing to’ instead of ‘due to’*

OK, we have replaced this.

• *l. 76: I think you mean Tissier and Legras here*

Yes, that has been corrected.

• *l. 78: do you mean “broadly similar” here?*

Yes, we have corrected this.

• *l. 100: what do you mean by “climate modes” – more explanation here.*

We are referring to modes of climate variability, specifically ENSO in this case. We have been more explicit in the revised manuscript and refer directly to ENSO rather than “climate modes”.

• *Figure 1: This figure is not really discussed in the paper. Which message does it communicate? I suggest removing the figure from the paper.*

We feel that Figure 1 provides a visual overview of the experimental design for the reader. Otherwise it is extremely difficult to show the flow of work and the design of experiments, with several emission scenarios, to a new reader. In this figure we also show how the model output is used for calculating SGI. Some of these concepts are new to the TransCom initiative, thus we would prefer to keep it.

• *l. 144: is a bottom-up . . .*

We have changed “bottom-up” to “a bottom-up”.

• *line 179: this means that the multi-model mean is highly influenced by CTMs driven by ERA-Interim data – correct?*

Yes. We have now commented on this in the manuscript. Please see our above response to the first detailed review comment on page 1 of this document.

• *l. 211: instead of ‘see also’ you could perhaps state for which information which paper should be consulted.*

We have removed the reference following “see also” as this was not needed.

• *l 301: what is the reason that ‘clear outliers’ are found? Are these models with obvious errors?*

The text here is discussing Figure 3. Generally there are few outliers and these outliers are limited to specific sites (for example, B3DCTM and STAG at SMO). We have added some text directly following the sentence containing "clear outliers" in Section 3.1.1 commenting on potential causes.

"The cause of the outliers at a given site are likely in part related to the model sampling error, including distance of a model grid from the measurement site and resolution (as was shown for CO_2 in Patra et al., 2008). These instances are rare for VSLS but can be seen in B3DCTM's output in Figure 3 for $CHBr_3$ at SMO. B3DCTM ran at a relatively coarse horizontal resolution (3.75°) and with less (40) vertical layers compared to most other models. Note, it also has the simplest implementation of boundary layer mixing (Table 2). This above behaviour is also seen at SMO but to a lesser extent for CH_2Br_2, for which the seasonal cycle is smaller (see below). The STAG model also produces distinctly different features in the seasonal cycle of both species at some sites (prominently at CGO, SMO and HFM). We attribute these deviations to STAG's parameterisation of boundary layer mixing, noting that differences for $CHBr_3$ are greater at KUM than at MLO – two sites in very close proximity but with the latter elevated at ~3000 metres above sea level (i.e. above the boundary layer)."

Later in the paragraph: "The NIES-TM model does not show major differences from other models for $CHBr_3$, but outliers for CH_2Cl_2 at Southern Hemispheric sites (SMO to SPO) are apparent. We are unable to assign any specific reason for the inter-species differences seen for this model".

• *l. 329: use r for the correlation coefficient*

OK, we have italicized "r" throughout the text.

• *l. 366: why does convection influence "near-surface" abundances of VSLS?*

Convection lofts tracer mass away from the boundary layer. We already point to several references which show this to be the case. For example, see plot of convective updraft mass fluxes in Figure 3 of Feng et al. (2011) which shows the vertical extent of convection.

• *l. 414: I think it is problematic that models have a preferred tracer: doesn't this imply that results could be right for the wrong reason?*

This comment is addressed in the detailed comments section above.

• *l. 425: Where is the reproduction of the c-shape shown? This seems an important issue.*

It is clearly visible for SHIVA and HIPPO-1. We have now been explicit as to where we are referring: panel (a), 2nd and 3rd row of Figure 10.

• *l. 435: The concept of a 'preferred' tracer means that the emission inventory somehow interacts with the models transport scheme to produce reasonable results at higher altitudes – correct? Can you describe in more detail here, what 'worse agreement' means?*

The models, with good agreement at the surface (by way of their preferred tracer), produce a sound simulation of the transport of $CHBr_3$/CH_2Br_2 from the surface to higher altitudes (as evidenced in Figures 10 and 11). The point we are making is that if the model-measurement agreement at the <u>surface</u> is degraded (e.g. in the simulations using the non-preferred tracer), then the absolute model-measurement agreement at higher altitudes is also worsened. We have reworded the text here for clarity. The sentence now reads:

"For a given model, simulations using the non-preferred tracers (i.e. with different $CHBr_3$/CH_2Br_2 emission inventories, not shown), generally lead to worse model-measurement agreement in the TTL. This is not surprising as model-measurement agreement at the surface is poorer in those simulations."

• *l. 485: is CO really short-lived?*

CO has a global lifetime of several months and therefore is similar to some VSLS. In order that CO and VSLS are not confused, we have reworked this sentence.

• *l. 492: state the lifetime in months/weeks*

OK, we now state that CH_2Cl_2 has a local lifetime > 1 year in the TTL.

• *l. 527: you might want to add here also Tissier and Legras 2015; Vogel et al. 2014*

OK. We have done this.

• *l 560: Clarify which best estimate is meant here, TransCom or WMO.*

WMO. We have added the citation to Carpenter and Reimann to clarify.

• *l. 593-595: The last sentence states that the VSLS loading in the LS is not correlated to MEI. But the sentences above state that bromine SGI is sensitive to modes such as MEI. Isn't this a contradiction? I think more discussion is require here.*

See earlier and also later comment. We have clarified this section of text and we are saying that the correlation is related to a particular region (the tropical E Pacific).

• *l. 598: change to: these processes*

OK, we have added "processes".

• *l. 599: change 'a range' to 'a number'*

OK, we have done this.

• *l. 614-618: Is the point here that the seasonal cycle is not dependent on the emission inventory, but the absolute model-measurement agreement is? How can this be the case. Please clarify. (See also abstract).*

Yes, that is correct and is the case because at most sites the seasonal $CHBr_3/CH_2Br_2$ abundance is determined from seasonality in the chemical loss rate (same for all models and same between model simulations with different emissions). We have clarified this point. The sentence now reads:

"At most sites, (i) the simulated seasonal cycle of these VSLS is not particularly sensitive to the choice of emission inventory, and (ii) the observed cycle is reproduced well simply from seasonality in the chemical loss (a notable exception is at Mace Head, Ireland)."

Of course the <u>absolute</u> model-measurement agreement will be sensitive to the emission fluxes as they vary in strength.

• *l. 626: change optimal to best*

OK, we have done this.

• *l. 634: what exactly is meant by 'online calculations'?*

An "online" emission calculation here refers to one in which emissions are calculated by taking into account the interaction between the atmospheric state and the ocean. Online calculations consider the actual seawater concentration of VSLS to derive air-sea concentration gradients and calculate fluxes. This is different to the approach mostly used to date whereby climatological emissions are prescribed. We have clarified this in the text.

• *l. 648: But the 'higher altitudes' are most relevant for the transport of VSLS into the stratosphere – correct?*

Yes. The model-measurement agreement during ATTREX is discussed in Section 3.3. We have added a sentence in the Summary and Conclusions noting that most models fall within 1 standard deviation of the observed mean at the tropopause.

• *l. 663: You mean the SGI range by Carpenter and Reiman, add the citation for clarification.*

OK, we have done this.

• *l. 670-672: This is astonishing, isn't it? I suggest somewhat more discussion on this point.*

This comment was answered earlier. Essentially, it may not be too surprising as changes to SSTs (and convective activity) associated with ENSO is zonally very asymmetric, with warming in some regions and cooling in others. The warming of East Pacific SSTs under El Niño conditions leads to enhanced SGI <u>over this region</u>, but when SGI is averaged over the whole of the tropics, this effect is dampened/cancelled. We have clarified these points at the end of Section 3.5

• *l. 676: change 'changes to' to 'changes of'*

We feel that "changes to emissions" reads better than "changes of emissions".

• *l. 678: change 'increased' to 'increase of the'*

OK, we have done this.

• *l. 679: distinguished from what?*

From the present day loading. To clarify we have changed "distinguished" to "determined".

• *l. 689: why is R Hommel not abbreviated?*

We did not abbreviate R Hommel because the abbreviation would be the same as the earlier, but different "RH".

• *Fig. 1: not sure if this figure is necessary*

We feel that Figure 1 provides a visual overview of the experimental design for the reader. We would prefer to keep it. Perhaps the Editor can comment on this.

• *Fig. 2: Continents in light grey would look better than in black.*

OK, we will update.

• *References: There are some references that need to be updated; ACP vs ACPD, Werner et al., 2016 etc.*

We have updated the reference list.