Responses to referee comments on "A study on harmonizing total column ozone assimilation with multiple sensors"

The comments from the two referees are provided below in the bold font style, and the responses are provided in the regular font style. The texts of the referee comments have been copied from original pdf files.

General acknowledgments from co-authors

Having gone back to the paper after some time away after from it, it was clear to us that the organization of the paper, as well as that within some sections, including the Introduction, was lacking. We thank the referees for putting us up to task on this aspect. As consequence, a major re-organization was conducted, accompanied by a reduction in the number of figures (both in the paper and Supplement) and in the text. The amount of details has also been reduced and some phrasings have been modified for improved readability and flow. This may render less obvious the location of implemented changes related specific referee comments. The page and line number of the specific comments of Referee #2 pertain to the earlier submission and not this revised copy.

Anonymous Referee #1

General and specific comments

1. This is a VERY detailed technical report on data assimilation techniques used to harmonize multiple total ozone data sources to minimize biases. The reasons stated for this effort is for the generation of trends over a long period of time (and multiple total ozone sources), for generating good ozone forecasts, and generating good UV Index forecasts. My first comment is that ACP is NOT the place for such an article but rather a more DA oriented journal like AMT or Journal of Applied Meteorology and Climatology.

- A significant motivation in submitting to ACP was the presence, in this journal, of the 2010 paper on "Multi sensor reanalysis of total ozone" dealing with column ozone bias correction and assimilation, in addition to various papers on the evaluation of ozone satellite data and their retrievals, and the MACC reanalysis paper.

2. There is too much detail provided in the early sections of this paper. The descriptions of the assimilation system, satellites, and surface observations can be greatly reduced as should the rest of Section 2.

- As recommended, the descriptions in the original Section 2, as well some results, were reduced. As well, following a comment from Referee #2, major subsections of Section 2 were re-distributed to other related sections. Moreover, the developments (and assimilation results) associated to the updating of background and observation error variances for ozone were removed.

3. I accept the reasons given for assimilating data from multiple sources on page 9. But for operational weather forecasting practicality, it is better to use one total ozone source and monitor the other sources. For a reanalysis effort this also applies, but the bias correction for long term trend (or consistency) plays a more important role.

- A reason for assimilating column ozone from more than one satellite is to better ensure that data is continually available in the event of occasional to permanent interruption of data availability from specific sources. For near–real time assimilation, this implies a need for contingency planning for transitions of bias correction references. This is mentioned in the text. As pointed out by the referee, one does not necessarily need to use data from all instruments either.

The following statements have been added: “One might opt to assimilate data from some sensors and monitor the data from the others through comparisons with the assimilation analyses. As well, while not necessarily negating the need for bias correction, one could always select to assimilate data from sensors with retrieval products having initially smaller biases as compared to other products. The effects of assimilating data with and without bias correction with individual and multiple sensors is also examined.”

The following text in the paper summarizes the impact of multi-sensor assimilation observed on the short-term forecasts: “Furthermore, the ACC demonstrates a more marked improvement in multiple sensor assimilation in the tropical region as compared to OMI-TOMS assimilation alone, which is not seen in the mean differences. The advantage of multiple sensor assimilation is, therefore, more notable in increasing the quality of the pattern and variation of the forecast fields.”

4. It appears to me that more time and effort should be spent on improving the Production and Loss terms to improve the ozone forecasts and maybe rethinking the climatology. The CNTL run shows how badly the total and profile forecasts become over a short period of time. Assimilating OMI or additional ozone sources will improve the analysis but have little effect on the forecasts. Improving the forecasts will also decrease the O-F in the tropics where one would expect them to be small.

- As pointed out and implied by the Referee, improving prediction models is an essential aspect of improving forecasts. Given improved initial conditions from intermittent assimilations, the photochemical forecasts of column ozone and stratospheric ozone over most latitude ranges can remain of good quality for the duration of quite a few days. This is exemplified by the regional changes in column ozone of within about 5% over 2 week periods in Figure 9 of the revised paper in the absence of assimilation. On the other hand, the quality of the ozone amounts near the mesopause, in the mesosphere, and near the ground (when not also in the stratospheric winter pole) could change more rapidly depending on the quality of the photochemistry model. There is also the quality of transport forecasting affecting local ozone concentrations. The authors are not involved in the forecast model development.

The following statement was added: “Also from Fig. 9, we can see that the error of the total column ozone forecast increases by less than 5 % over the course of fifteen days, reflecting the high predictability of ozone medium range forecasts. This limited deterioration would not deter, for example, in properly forecasting the movement of low column ozone regions during this periods and the corresponding changes in clear-sky UV Index.”

5. One of the purposes of this effort is to improve the UV Index forecasts. Clear sky UV Index values are generally determined from total column ozone and solar zenith angle. Percentage errors in the total column ozone reflect nearly an equal (opposite sign) error in the UV Index. So a +/- 2 percent error in total column ozone generates the same amount of error in the UV Index. That is highly acceptable, especially when adding the much larger range of errors due to clouds and aerosols.

- As pointed out by the referee, changes of a few percent in column ozone forecasts imply errors of only a few percent for the UV Index. As well, individual extreme outlier data would usually be identified in the background check phase and not used in assimilation. As the impact of column ozone bias itself on the UV Index would nearly always be at the level of a few percent, the mention of the association of ozone bias correction to the UV Index in the introduction was removed.

6. From the results of additionally assimilating ozone profile information either from the MLS or OMPS-NP, I would hope EEEC considers doing so operationally to improve the ozone profile. This is important radiatively as the ozone profile plays an important role in the temperatures in the stratosphere. This also could improve the ozone forecasts in the high latitudes, especially within ozone depleted regions.

- There is the intention/interest of pursuing the added assimilation of partial column ozone profiles from OMPS-NP and SBUV/2 in near-real time, if not also the OMPS-LP limb profiler. As part of the reduction in size of paper, the assimilation results regarding the vertical structure, and the corresponding use of MLS and ozonesondes, were removed

7. The usage of the OMPS-NM and OMPS-NP versions prior to their "final" version 8 product is unfortunate, but I gather was the only choice for this study. Many of the OMI-OMPS differences will be ameliorated in the version 8. I hope this is made crystal clear in the article so that readers will not get the wrong impression of the quality of the OMPS products in version 8 form.

- As suggested by the referee, the data from the recent retrieval algorithm were not yet available when the study was undertaken. The issue is mentioned in section 2. The following statement has been added in the Conclusions section: "As the quality of the different versions of OMPS retrieved data may differ, one might expect a reduction in bias of the more recent version of the OMPS products based on the SBUV V8.6 retrieval algorithms".

8. The graphics are quite good. One comment is that a solid or dashed line at any 0 (zero) point should be shown as a reference.

- As recommended by the referee, lines at the zero points were added.

9. Another comment is that other color choices for Figure 7 should be considered. Or the graphs be separated showing the GOME results on one and the OMPS & SBUV/2 on the other.

- The six curves of the original top plot have been equally redistributed (3 each) to two plots (now Figure 4 of the revised paper). As the original bottom plot did not add much information, it was been removed.

10. I don't think the "ozone effective temperature" is defined or explained such that the user knows where in the vertical or what layer this "temperature" represents

- The definition of the 'ozone effective temperature' has now been provided in the introduction: "The ozone effective temperature is the average value of the ozone-weighted temperature profile". Its initial definition in the original section 2.2 is also now less abbreviated.

Anonymous Referee #2

General comments

1. The manuscript titled "A study on harmonizing total ozone assimilation with multiple sensors" by Rochon and colleagues attempts to present efforts to assess the impact of assimilating total column ozone datasets from single and multiple satellite data sources with and without bias correction has been examined with a version of the Environment and Climate Change Canada assimilation and forecasting system. While the manuscript presents a wealth of comparisons and analysis, this is not performed in an optimal, easy to follow manner, and results in discouraging the reader as the information that might be of interest is scattered across the text. The crux of the matter, i.e. the improvement [or not] of the forecasts when assimilating [or not] specific satellite datasets starts in Table 4, already in page 34, and in Figure 10, already in page 36, without even counting the numerous Figures in the supplement which makes the reading of this text even more confusing/tiring. I strongly suggest the following to the authors:

Consider shortening your paper significantly, either by omitting steps or by simply braking it down to a two-part paper where in the first [Part I] all material up to and including Section 3.2.2 should go into [as well as the associated supplement material] which would be the "preparation part". The rest [Part II] can be the assessment of the different bias, comparisons, outliers, runs, options a), b), c), etc., as well as the results can go into.

- A major re-organization and a significant reduction of the paper content were conducted. The number of figures has been reduced to ten in the paper and three in the Supplement. This had initially been recommended during the earlier review stage and was to be continued following the discussion phase. There is also no longer any referencing of the few remaining figures in Supplement, with a remaining referencing to Tables S1-S3.

2. Consider adding a "roadmap" to this work right after your introduction. So that people who are interested in specific parts of this work can know which section to follow. For e.g. I suggest you write that : first you will show validation of the satellite sensor to be used as anchor, then you will show the whole bias correction, then you discuss the assimilation system, then the comparisons between the different assimilations you performed, etc etc. etc.

- A re-organization of some sections and some section contents was conducted. The 'roadmap' to the organization of the paper was accordingly revised and made more specific in the introduction.

- This paper was originally intended to be more of an assimilation study. The bias correction aspect eventually became more encompassing. For clarity in the paper, focus is now placed first on total column bias estimation and correction followed by implications of the bias correction and multi-sensor use in assimilation on short-term column ozone forecast. The subsection on the impact of the forecast vertical structure was removed.

The new structure of the paper is now as follows:

1. Introduction
2. Observations
    - Subsections describing the different observation sets

3. Evaluation of OMI-TOMS total column ozone with ground-based data
4. Bias estimation and evaluation using OMI-TOMS as reference
   - Methodologies, results and analysis
5. Assimilation system and results
   - Description of the system and evaluation of the total column ozone short-term forecasts
6. Conclusions

3. The whole "bias discussion" followed so closely by the OMI validation section, etc., is extremely confusing. From the results, one does not follow at all why you have to assimilate also GOME2A [which produces obvious problems] or GOME2B [since you already have two other TOC-providing sensors, OMI and OMPS.]

- As indicated above and below, a re-organization of the paper content has been conducted.

- While it is not essential to assimilate data from GOME-2A (nor all indicated sensors), their use shows the implications when included in assimilation. One could have opted not to assimilate the GOME-2A retrieved product especially, this work shows that its bias can be reduced to provide data of similar quality as other datasets. Text was added to not give the impression that all sensors must be assimilated. See related response above to the general comment 3 from Referee #1.

4. Furthermore, the most important part [in my opinion] is the assimilation of the profiles [partial columns] of OMP and MLS but discussion/results on those is added as an afterthought at the end and not properly analysed. 4.

- The added assimilation of partial column measurements was intended originally to be part of this paper. However, results obtained from the assimilation of these data indicated that some optimization work was required to assimilate these data without inducing biases, this most notably seen in the lower stratosphere and upper troposphere. It was therefore decided to deal with the assimilation of partial columns in a later study. As consequence, the use of OMPS-NP and SBUV-2 in this paper was restricted to the comparisons of their equivalent total columns.

- The assimilation of MLS (and especially the evaluation relative to MLS) was intended to have played a more prominent role if the assimilation of partial columns had been included. The assimilation of MLS was also not being targeted for near-real time assimilation.

- As part of the reduction in size and figures of the paper, and in consideration of the above and below comments, the section on the ozone field vertical structure and use of MLS and ozonesondes has been removed.

5. The conclusions need a massive re-writing so as to give numerics to the work presented. I suggest you re-think the presentation of your results in a bullet-type manner so that actual findings can be easily understood and benefit other scientists.

- The Conclusions section has been revised, with addition of some numerics.

6. In short, I suggest that the authors re-consider their entire strategy, to massively decrease the material in the supplement and change the focus of their paper. Please refer to attachment for further comments.

- Done

Specific comments (copied from the comments imbedded by the referee in a pdf copy of the submitted paper)

P1 L16. Are we to understand that only the OMI dataset was validated in this paper and not the rest of the satellite sensors? if so, please rephrase accordingly.

The data from all the total column ozone satellite sensors indicated earlier in the abstract are evaluated. The later sentence mentioning OMI-TOMS as reference/anchor for the other sources has been moved up in the abstract.

P1 L21. This is a bias correction between what and what? the abstract has to be a stand-alone paragraph, so that someone, by reading it, understand all the major details /findings of the full paper.

While the OMI-TOMS data is indicated as the reference for bias correction (e.g. P1 L16 in original text), the text has been re-organized to make it clearer. The statement "were performed for GOME-2A/B and OMPS-NP" has been added in the sentence to explicitly indicate the datasets being corrected.

P2 L5. I am all in favour of numerical findings in the abstract, however [when read by a non-assimilation expert] these numbers do not mean all that much. Maybe a sentence can be added here, from the conclusions section, discussing the anomaly correlation coefficients in more detail?

It was decided to remove the mention of the anomaly correlation coefficients in the abstract and Introduction, It is mentioned only in the new Section 5, where it is defined, and in the Conclusions.

P3 L4-11. Even though the two reasons might be the world's best trade secret, from this paragraph alone, they do not convince a non-assimiation expert such as myself. Maybe you could add a few more details and/or references.

The Introduction was re-organized, this including the removal of the related paragraph. Accuracy and stability requirements from ESA's Climate Change Initiative have instead been introduced.

P4 L11. "Balis et al., 2007b". Since this appears first maybe this should be 2007a?
P4 L12. ""Balis et al., 2007a". See previous comment

Done.

P4 L13. "Hao et al., 2014". Since in this paper the GOME2 data were assimilated and discussed under the light of the MACC re-analysis [assimilation] maybe you can extract some more information on how the bias correction was treated there, for e.g. Even though of course both observations and assimilation model are different from yours.

Bias correction used in that work is part of the variational assimilation step referred in the introduction (which mentions the same references specified in Hao et al, 2014). It is preferred not to describe the details of variational bias correction in this study.

On the other hand, even though data from the DOAS retrievals were used in Hao et al. (2014), this reference in now used in the new section 4 regarding differences found between GOME-2A and GOME-2B and the GOME-2 instruments and the ground-based data. The added statements are as follows: "For the DOAS retrievals products, Hao et al. (2014) present mean differences with Norther Hemisphere ground-based data varying seasonally between roughly zero and 4 % over the period 2007 to Summer 2013. The seasonal variation for the TOMS retrieval products is also seen here from Figs. 2 and 3 and Tables 2 and 3, with larger differences for GOME-2A of up to about 3 %. They also present differences between the GOME-2A and GOME-2B of less than 1 % covering December 2012 to November 2013 except in the south pole region and in the Southern Hemisphere for about May to September where it reaches at least 2 %. This differs from the usually larger differences seen for the periods of this study with TOMS-based GOME-2 retrieval products. "

P4 L20. It might be a good addition to this paragraph to mention the work of Lerot et al., 2014, in homogenizing satellite TOC observations

https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2013JD020831

as well as the validation of this dataset,

https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015JD023699

and the validation of the extended dataset which also includes the OMI and GOME2B sensors:

https://www.atmos-meas-tech.net/11/1385/2018/

These references have been added. Thank you for indicating them to us.

P4 L32. "Levelt et al., 2006". Maybe you can update with:

https://www.atmos-chem-phys.net/18/5699/2018/

Done. Thank you again.

P4 L33. "Callies et al., 2000; Munro et al., 2006". Maybe you can update with

https://www.atmos-meas-tech.net/9/383/2016/

and

https://www.atmos-meas-tech.net/9/1279/2016/

Thank you. The updates to the more recent references has been done.

P5 L30-32. Thank you for this clarification, however the supplement is rather large. If you consider that all those tables and figures are necessary for better understanding/reading of this paper maybe you can separate this work into Part 1 and Part 2. As it stands, it is quite tiring to keep referring to the 20 extra figures of the supplement.

The referencing to Supplemental Figures has been removed and the number of Supplemental figures have been reduced to three.

P6 L3. Where is this operation system used? by Meteo/Environmant Canada?

The phrase "of Environment and Climate Change Canada" has been adapted to qualify the model. As well, the word "operational" has been removed to avoid any implications that ozone assimilation in the work is operational.

P7 L15. How do you account of the differences in overpass time between the different sensors? Similarly, there exist [as was well documented in most validation papers you reference] differences in the TOCs provided by the satellite observations due to the differences in the algorithms, in the spatial resolution of the FOV, due to the stronger degradation of the GOME2A sensor compared to GOME2B, etc. How is this all dealt with?

The description of the bias correction considerations had been provided later is in section 2.4 on bias estimation and correction (and has now been moved to section 4). The differences in spatial resolution is disregarded in our bias correction so any effect on bias correction would be part of residual biases and associated representativeness errors. The representativeness errors are not explicitly quantified and included.

The mention of forecast and observation resolutions has been added in section 4 as "Any bias impact due to differences in spatial resolutions of the instruments and or of the model forecasts would be part of residual biases and associated representativeness errors. Part of the effect of differences in resolution between instruments would be mitigated from bias estimation relying on local averages of differences in space in addition to time."

Differences in algorithms and instrument conditions could be a potential source of bias reflected in the resulting bias estimates. Related statements where mentioned in the original section 2.4 with the sentence "Latitude and time dependences are also introduced to capture other first order retrieval biases, such as potentially related to the applied a priori atmospheric state and its ozone error covariances, and the specification of the ozone absorption coefficients, as well as instrumental changes in time." A much reduced version of this sentence is in the new section 4: "Latitude and time dependences were introduced to capture other data processing biases as well as instrumental changes over time,"

P9 L14. You mean, Koukouli et al., 2012?

Yes, thank you. It has been corrected.

P10 L21. Any validation papers on the NRT GOME2 product you can share?

Am not aware of any papers on the validation of the NRT GOME-2 products relying on the TOMS retrieval approach. The following sentence was added in section 2: "This study provided an opportunity to evaluate the biases of the GOME-2 TOMS products."

Also see the response to P4 L13 for comparisons to GOME-2 products based on the DOAS retrieval in the new section 4.

P10 L22. Any validation on the actual profile of partial ozone columns?

Average differences between SBUV/2 and the version of OMPS-NP were provided in Flynn et al. (2014). This was indicated in the three paragraph of the original section 2.2.3 without specifically providing the differences. Considering the request to notably reduce the description of observations and the use of only total column data in bias correction, the mention of validation of OMPS-NP profile validation has instead been reduced to the sentence "See Flynn et al. (2104) for a description of accuracy and precision of the OMPS-NP V6 products."

P11 L21-24. It is not customary to use numerical findings from a paper which was rejected for publication. I strongly suggest to remove any precise findings from that ACPD 2013 paper from your discussion, since you have the more updated Bai et al., 2016.

It had not been realized that the paper had been rejected. The related sentence and reference have been removed. Bai et al, (2016), already indicated in the paper, summarizes the OMPS-NM evaluation but not that for OMPS-NP as was found in Bai et al. (2013)

P12 L1. This subsection is a bit mixed up. It basically reads as if you wrote the first paragraph, then someone suggested you expand, and you simply added the other paragraphs without making sure that the text flows. You have to re-write this subsection, separating the different verification sources and also explaining exactly how you will use these further on.

Thank you for catching the poor presentation in this section. The intent was to identify all sources in the first paragraph and expand on the description of the different sources in the following paragraphs. The first sentence of the first paragraph is rather long. The text has been notably modified. Headings identifying the separate instruments have been added.

P12 L17-18. How did you choose which Brewer and Dobson stations to use in this work? Fioletov et al., 1999, use a very refined way to assess the representativeness of a ground-based station and its usefullness in validating satellite data. Did you simply take all the data available from WOUDC for the three periods you use?

We considered all direct sun daily data from the WOUDC and from NOAA (for stations not found in the WOUDC) for the periods of interest (there were only 1-2 stations excluded due to an issue with the location - this is indicated in the Supplement tables). In the comparison to OMI, the means and standard deviations of differences about means were generated for each station. Outlier differences were identified for each station as values differing from the mean difference by more than two standard deviations. These were removed from a second calculation of the station mean differences.

The same process was repeated when taking averages over all stations of the station mean differences, this also allowing to identify and exclude outlier station mean differences,

Additional statements have been added to better identify the process of excluding data from the original sets.

P12 L18. Yes, but did you use exactly the same set of GB stations? GB instruments undergo constant calibration and often datasets are improved and reprocessed for WOUDC dissemination.

While considering all WOUDC (and NOAA) data for the evaluation, we did not necessarily use exactly the same stations (see response to P12 L17-18).

P12 L20-21. A factor of 2 to the 4.6 D.U. you mention above? or not?

The standard deviation value was provided as an illustration of the size of standard deviations. The related text was modified/clarified following the changes related to P12 L1. The current sentence referring to the 4.6 DU is "Consistent with the above, an overall precision of 4.6 DU has been obtained by van der A et al. (2010) for Brewer and Dobson direct sun daily averages, excluding outlier data."

P12 L31. I fail to understand why you analyse so much the issue of the effective temperature dependence of the Dobson observations, compared to the Brewer observations, since you are going to assimilate satellite data. This paragraph feels a bit too much information on a side topic, unless there is an actual reason for having to explain it further below.

The ground-based data are used in the evaluation of OMI-TOMS and of the assimilation results. This is now explicitly indicated in the introduction. Based on this, it was considered relevant to correct for the known Dobson bias, where possible, which can be corrected as a function of ozone effective temperature. In this paper, the correction is of most significance for data near or in the south polar region. We slightly reduce the amount of justification text.

P13 L6-7. What exactly does this phrase mean? that Bhartia et al.., 2013 have shown that the SBUV TOCs can be used up to SZAs of 88? if so, please rephrase accordingly.

It has been re-phrased as follows: "Bhartia et al. (2013) has indicated that the total column ozone values resulting from the V8.6 algorithm can be used for solar zenith angles up to 88°."

P13 L8. Are you saying that the MLS profiles have a spatial resolution of 160x160km?!? Surely not. Add validation information on these profiles here.

The value of 160 km refers to the along-track sampling while the effective horizontal resolution, stemming from the limb measurement geometry, is 300-450 km. MLS takes single profile measurements at a spatial interval of about 160 km between profile measurements along the satellite track. The text and use of MLS in the paper was removed as part of the size reduction.

P13 L14-15. As openings to sub-sections are concerned, this is rather dense. Re-phrase accordingly. What exactly are you going to be discussing in this section?

Subsections of section 2 have been re-organized and contents of the earlier section has been notably reduced and modified, and also moved to section 5.

P13 L17-19. Why did you choose to present these details in section 2.4 when it seems that they are necessary to explain section 2.3?

The reason for having had section 2.3 before 2.4 was to support specification of the target bias reduction later mentioned in section 2.4. The original sections 2.1 and 2.3 have been combined, shortened, and moved section 5.

P13 L26. I find this paragraph nearly impossible to follow.

This paragraph was removed as part of the re-organization of section 2.

P14 L13. "adding outlier removal". To which IQR? 1.5 or 3?

Outliers are identified here as differences minus the initial mean difference, i.e.. [diff - mean(diff)]. that are beyond two standard deviations (sigma) of the initial distribution of differences (section 2.4.1). For comparison, the short example at

https://www.thoughtco.com/what-is-the-interquartile-range-rule-3126244

gave a threshold of 19 for the 1.5 IQR. The 2sigma application for this same sample dataset gives a threshold of 16.2 The condition, here, is slightly tighter than the 1.5 IQR case.

The above comments also relevant to the outlier removal in other sections of this paper.

The text at this specific location has been removed as the related text on updating background and observation error standard deviations has been removed as part of the size reduction and of improving readability.

P14 L14-17. Well, to be honest, this paragraph reads as if the aim is to reduce the observation standard deviations when in reality the aim is to quantify properly the observation standard deviations. Please re-write.

An aim of the text regarding the observation standard deviations was to better quantify their values for a later comparison to the estimated biases. The related text has been removed as stated for P14 L13.

P14 L14-17. I also think that the whole issue of the model horizontal resolution vis-a-vis the satellite ground FOV should be discussed more. If not quantified.

The following statements regarding the model horizontal resolution vis-à-vis the satellite ground FOV have been added/modified. In the new section 3: "Any bias impact due to differences in spatial resolutions of the instruments and or of the model forecasts would be part of residual biases and associated representativeness errors. Part of the effect of differences in resolution between instruments would be mitigated from bias estimation relying on local averages of differences in space in addition to time." In the new section 5: "In assimilation, inconsistencies stemming from the differences in resolutions between the model forecasts and the observations would usually be reflected by some corresponding increase of applied observation error variances. This is not explicitly done here."

P14 L31. You have to show this correction. It is imperative. People have spent years trying to homogenise satellite datasets, apart from the van der A work you mention, there is also the work of Coldewey-Egbers et al. [see here https://www.atmos-meas-tech.net/8/3923/2015/] and so on.

For clarity and as part of the reorganization, the bias estimation/correction methodology subsection of the original section 2 describing bias correction has been moved to section 4, modified and somewhat reduced in size, Section 4 (section 3.1 in the original paper) also contains the application of bias correction. The re-organization will hopefully make the content clearer.

Referencing to Coldewey-Edgers et al. has been added in the new section 2.4.1 referring to the various sources mentioned therein describing the use of ground-based data in satellite data validation.

P15 L23-24. Maybe this phrase is incomplete?

Now corrected.

P16 L9. Does the bias correction depend on the bins sizes?

The sizes identify which observations will be used in bias estimation for each bin. The determined bias correction is assigned to the center of the bin, with interpolation of bias corrections for locations between neighboring bin centers. So the size needs to be usually small enough so that differences between bins are not too large while the bins are large enough (if possible) to have enough measurements. There is some flexibility in sizes in consideration of these conditions.

P17 L13. If you accept lat differences within 3 degrees this is more or less 300km, for normal latitudes [i.e. not polar] How does this add up to the 200km collocation criterion you mention previously?

The mention of 3 degrees was more in reference to optimizing the efficiency of the colocation search code and has been removed. The distance limit remains 200 km.

P17 L13. 25 DU on "nominal columns" of 250-350 DU is a 10% offset, isn't that a tad too much?

It was set somewhat arbitrarily, favouring a large value considering the spatial and temporal windows for the colocation. It could have been made smaller. No problematic issues resulting from this choice were identified. The mention of the 25 DU has been removed.

P18 L2. This suggests that you have perfomed this analysis for more months that August 2014? maybe a table with such statistics for the entire season [i.e. jan to dec] would be appropriate here.

This figure was generated as part of the preliminary investigation in selecting the dependence on SZA and latitude. This figure and the related text have been removed as part of the requested paper/figures reduction.

Final results showing the variation with SZA and latitude are provided in section 4. Tables have been added summarizing and accompanying the information in these figures.

P18 L4. I am bit concerned that the comparisons shown here apply to monthly mean values. A difference of -2% for SZA 50 degrees may stem from daily comparisons between 0 and -4% but also from daily comparisons between 10 and -12%. Without the associated STD values it is impossible to say [without weighing them with the N value of course.] Since you will assimilate daily values, shouldn't those statistics be the main focal point?

The satellite data to be assimilated are individual measurements as oppose to daily averages.
The monthly means shown in the original Fig. 2 were for illustration (see also P18 L2) in showing why bias correction was specified to be dependent on latitude and SZA. As indicated in the earlier comment, Figure 2 of the original paper was removed.

The bias correction actually applied in assimilation is adjusted every six hours using a moving two week window. This was indicated in the text and is now also repeated in the assimilation section of paper (section 5).

P19 L20. Shouldn't we see these assimilation runs, first of all? before showing the differences to the OMI dataset, and the ground-based stations?

As the assimilations are most relevant in examining the impact of data with and without biases correction (for individual to multiple sensors), the paper has been organized to place the assimilation component after the inter-comparisons of observation sets. However, the presentation in the text was likely misleading in the regard. The re-organization has hopefully helped in better streamlining this sequence.

P20 L21-22. These comparisons are even better than the individual comparisons reported in the validation reports you have referenced... how do you justify removing stations with mean differences larger than 6 D.U.? are you implying the station is not quality assured? that OMI is not quality assured? why did you not simply discard these stations and use the stations typically studied in the validation studies you reference [Fioletov et al., Balis et al., Labow et al., etc.]

The first three paragraphs below address the 6 DU point and quality assurance. The remaining paragraphs address the better comparisons.

The choice of 6 DU, which is partially subjective, was based on examining the set and range of mean differences in Tables S1 to S3, as well as considering the station locations. Only 3 station mean differences out of ~250 in Tables S1 to S3 indicated values lager than 6 DU. Two of the outliers (beyond 6 DU) are for stations with elevations above 2 km (so a potential effect of representativeness error). For the third, this being the value -21 DU, the series of differences shows a correspondingly large systematic offset present for the Winter 2015 period. This large systematic offset, located within the northern mid-latitude band, highly likely stems from a systematic error of the ground-based data. The mention of the 6 DU threshold has been removed.

There would be quality assurance variations over the stations and for OMI (e.g., as a function of SZA) to some degree. It is likely that data from some stations may not have been as quality assured as others, yielding larger residual calibration errors The quality of the station data would depend on the instrument calibration which can re-done over time. As such, it is not impossible for a station to have different bias levels for different periods. The variation of the mean differences over stations most likely has a component stemming from some station-by-stations variations of quality assurance.

Based on the analysis, the largest quality assurance question for OMI would be regarding the quality of column ozone values of the south polar region. A difficulty here, as pointed also by referee, is the uncertainty of column ozone measurements at high SZA.

Even when biases for station and OMI values, notable biases of the differences may appear for some stations due to the combinations of highly variable local topography (elevation), the spatial colocation criterion, and the OMI measurement resolution/foot-print on the ground.

It is interesting that the differences produced are smaller than from other references. The following text was added in the new section 3: " The global mean differences, and most regional values, are typically smaller than earlier studies mentioned in the first paragraph. Possible contributors to this might be some differences in time periods, region specifications, ground-based observation sets, or colocation conditions."

Considering this concern, a review of the calculations was conducted. We noted and corrected some transcription errors in Tables S1 to S3, the mislabeling of a station ID and position, and a few duplications of Dobson WOUDC and NOAA station data. The table updates only had a minor impact on the obtained regional and global mean differences.
Additional text has been added discussing the results in the north polar region: "Koukouli et al. (2012) determined standard deviations of the differences of 2.4 and 4.3 % for SZA ranges of 25-70° and above 70°, indicating an increased variability at higher SZAs. Considering the respective periods of this study and of Koukouli et al. (2012), their differences for this region may stem from differences in the range of SZAs. The mean differences for both polar regions are all negative indicating an underestimation of OMI-TOMS column ozone in these regions for these periods relative to ground-based data. which is likely related to high SZAs."

P20 L24. The reason there are larger differences between satellite and ground monitoring in the Arctic and the Antarctic is that there lingering issues with the satellite algorithms over those reasons, apart from any calibration [or lack thereof] applied to the ground-based stations. I do not think that you can justify excluding these statistics in such an ad hoc manner. Please comment/rethink/rephrase.

These outliers did not consist of Arctic stations. The Antarctic stations identified as yielding outlier mean differences, while not included in the global averages of the Table 1 (and stated in the earlier sentence at L21-22)), have been retained in the analysis as indicated with an included discussion. This last point is now indicated in the text at this location.

P21 L6. I think that you should note here that there are known issues with the Filter stations for a number of years now and that official reporting in WOUDC states that these data are to be used frugally.

We were not successful in finding this WOUDC statement. Instead text has been added in this analysis section referring to the statement in the Observations section that "The error standard deviations … are about 1.5 to 2 times larger for filter ozonometers (Fioletov et al., 1999, and references therein; Fioletov et al., 2008)". Fioletov et al., (1999 and 2008) and Staehelin et al., (2003) are provided as references for the three ground-based instruments in the Observations section.

Interestingly, comparisons to filter gave mean differences that were not that much larger than those for Dobsons.

P21 L11-15. In all these works, full seasons have been used for the statistics, with full ranges of SZA observed by the ground and full seasonality effects. I do not consider it product to make a one-to-one comparison in this manner.

The caution of different studies covering different periods has been repeated and, particularly for polar stations, the differences in ranges of SZA has been added. The latter is exemplified in the response for P20 L21-22 as well as the following sentence when discussing results for the Antarctic: "Small differences in observed locations, as well as small differences in solar zenith angles of the colocation pairs at high SZAs, can imply notable differences of observed air masses."

P23 L2. Well, as you know, 1 degree in Antarctica is a large distance and you may be comparing two entirely different air masses for even relatively small SZAs between satellite and ground.

The statement that small differences in observed locations can imply notable differences observed air masses has been added (see P21 L11-15 above).

P43 L8. "Conclusions". Needs re-writing!

The Conclusions section has been revised, this including the addition of numerics.

P46. L3-4. See P11 L21-24

The reference has been removed.

A study on harmonizing total ozone assimilation with multiple sensors

Yves J. Rochon[1], Michael Sitwell[1] and Young-Min Cho[2]

[1]Atmospheric Science and Technology Directorate, Environment and Climate Change Canada, Toronto, M3H 5T4, Canada
[2]Centre for Research in Earth and Space Science, York University, Toronto, M3J 1P3, Canada

Correspondence to: Yves J. Rochon (yves.rochon@canada.ca)

Abstract. Bias estimations and corrections~~The impact~~ of ~~assimilating~~ total column measurements are applied and evaluated with ozone data~~datasets~~ from ~~single and multiple~~ satellite instruments providing near-real time products during Summer 2014 and 2015 and Winter 2015. The developed standalone~~data sources with and without~~ bias correction system can be applied in near-time time chemical data ~~has been examined with a version of the Environment and Climate Change Canada variational~~ assimilation and long-term reanalysis. ~~forecasting system.~~ The instruments to which these bias corrections were applied~~assimilated and evaluated data sources~~ include the Global Ozone Monitoring Experiment-2 instruments on the MetOp-A and MetOp-B satellites (GOME-2A and GOME-2B), the total column ozone mapping instrument of the Ozone Mapping Profiler Suite (OMPS-NM) on the Suomi National Polar-orbiting Partnership (S-NPP) satellite, and the Ozone Monitoring Instrument (OMI) instrument on the Aura research satellite. The OMI dataset based on the TOMS version 8.5 retrieval algorithm was chosen as the reference used in the bias correction of the other satellite-based total column ozone datasets. OMI data was chosen for this purpose instead of ground-based observations due to OMI's significantly better spatial and temporal coverage, as well as interest in near-real time assimilation. Ground-based Brewer and Dobson spectrophotometers, and filter ozonometers, as well as the Solar Backscatter Ultraviolet satellite instrument (SBUV/2), served as independent validation sources of~~for~~ total column ozone data. Regional and global mean differences of the OMI-TOMS data with measurements from the three ground-based instrument types for the three evaluated two month periods were found to be within 1 %, except for the polar regions where~~with~~ the largest differences from the comparatively small dataset in Antarctica exceeded~~exceeding~~ 3 %. Values from SBUV/2 summed partial columns were typically larger than OMI-TOMS on average by 0.6 to 1.2~~±0.7~~ %, with smaller differences than with ground-based over Antarctica. ~~OMI-TOMS was chosen as the reference used in the bias correction instead of the ground-based observations due to OMI's significantly better spatial and temporal coverage and interest in near-real time assimilation.~~ Bias corrections as a function of latitude and solar zenith angle were performed for GOME-2A/B and OMPS-NM~~with a two-week moving window~~ using colocation with OMI-TOMS and three variants of differences with short-term model forecasts. These approaches were~~are~~ shown to yield residual biases of less than 1 %, with the rare exceptions associated with bins with less data. These results were compared to a time-independent bias correction estimation that used colocations as a function of ozone effective temperature and solar zenith angle which, for the time period examined, resulted in larger ~~changes in~~ residual biases ~~as a function of time~~ for bins whose bias varies more in time.

Formatted: Font: (Default) Calibri Light

The impact of assimilating total column ozone data from single and multiple satellite data sources with and without bias correction was examined with a version of the Environment and Climate Change Canada variational assimilation and forecasting system.~~some cases.~~ Assimilation experiments for the July-August 2014 ~~period~~ show a reduction of global ~~and temporal~~ mean biases for short-term forecasts relative to ground-based Brewer and Dobson observations~~data~~ from a maximum of about 2.3 % in the absence of bias correction to less than 0.3 % in size when bias correction is included. Both temporally averaged and time varying mean differences of forecasts with OMI-TOMS were~~are~~ reduced to within 1 % for nearly all cases when bias corrected observations are assimilated for the latitudes where satellite data are present. ~~is present. The impact of bias correction on the standard deviations and anomaly correlation coefficients of forecast differences to OMI-TOMS is noticeable but small compared to the impact of introducing any total column ozone assimilation. The assimilation of total column ozone data can result in some improvement, as well as some deterioration, in the vertical structure of forecasts when comparing to Aura-MLS and ozonesonde profiles. The most significant improvement in the vertical domain from the assimilation of total column ozone alone is seen in the anomaly correlation coefficients in the tropical lower stratosphere, which increases from a minimum of 0.1 to about 0.6. Nonetheless, it is made evident that the quality of the vertical structure is most improved when also assimilating ozone profile data, which only weakly affects the total column short-term forecasts.~~

Formatted: Font: (Default) Calibri Light

Formatted: Font: (Default) Calibri Light

1 Introduction

~~The assimilation of column and stratospheric ozone measurements for ozone-layer forecasting has been conducted mostly as of about twenty five years ago, ultimately culminating with operational ozone-layer and UV-index forecasts (e.g. Lahoz and Quentin, 2010; Inness et al., 2013). This typically involves the application of measurements from single to multiple satellite remote sounding instruments with use of ground-based and other remote sounding data for independent verifications and, occasionally, bias correction.~~

~~The assimilation process consists of introducing information from observations into the model forecasts through the generation of analyses, the statistical blend of the earlier forecast and observations in model space, serving as initial forecast conditions. Traditionally, this blending process assumes that both the model forecasts and observations are statistically unbiased following an initial spin-up time. Unremoved biases or systematic errors of the observations or forecasting model~~

can potentially impact the quality of the analyses and forecasts (e.g., Dee, 2005; Dragani and Dee, 2008). This is important for total column ozone when it comes to monitoring for multi-decadal trends as referenced in van der A et al. (2010), whether this is using measurements with or without data assimilation. Generally, while the effectiveness of bias correction schemes in removing the actual biases is constrained by the limited knowledge of the truth, their impact in reducing relative biases between different assimilation sources and/or correlated variables can potentially be as significant, if not more so, for improved forecasting. An example of this is in multivariate assimilation, where ozone and weather assimilation can be coupled (e.g., Dee et al., 2011; Dee, 2008), in the presence of differing systematic differences between forecasts and observations for different variables.

Total column ozone biases from satellite measurements are typically within a few percent. -Changes of a few percent over time or between instrumentssources are significant in affecting the correct identification of long-term trends. Near-global reductions in column ozone have been -1.8 % per decade from 1980 to the mid-1990s and increases over the past two decades are at only 0.4 to 0.6 % per decade (Steimbrecht et al., 2018). A requirement on the long-term stability of corrected total column ozone observations of 1-3 % per decade was specified by the Ozone_cci project of the European Space Agencies' Climate Change Initiative program in Table 5 of Van Weele et al. (2016). This table also indicates accuracy requirements on but not so much otherwise, possibly unless applying multivariate assimilation, where ozone and weather increments are statistically coupled, with biases large enough to negatively impact the weather fields; the latter assumes a correct vertical distribution of the total column ozone measurements of 2 % for facilitating research on the evolution of the ozone layer from radiative forcing and 3 % for studies on short-term, seasonal, and interannual variability. As an example, for an accuracy requirement of 2 % and measurement precisions between 1.0 and 1.7 %, biases need to be no larger than about 1.7change and would extend to 1.0 %, respectively. The comparison of model biases. Total column ozone data from different instruments allows for the identification of the level of agreement between datasets, potentially under various conditions, and can highlight cases and bias correction is included in this ozone assimilation study considering comparatively short two month periods for two reasons other than the future potential extension to long-term trend monitoring and or multivariate assimilation. One is the conventional preference, that any residual measurement biases of the data affecting the analyses be generally less than their random errors. Second, it provides a mechanism to identify and correct bias occurrences which exceed typical levels even if not frequent or continuous. Both will contribute to the quality of UV index forecasting at least for clear sky conditions with small to large relative biases. As. The developed bias correction infrastructure can be applied to other constituents and data types as needed.

Total column ozone bias estimation for observations can be performed in different ways as a function of recognized dependency factors such as the solar zenith angle, latitude, and seasonal variation among others, potentially with related bias error covariances. Seasonal and related latitudinal changes in biases may result from limitations in retrieval algorithms such as, for example, not accounting well enough for the variation of the ozone layer temperatures in specifying the ozone absorption coefficients. Differences and limitations in accounting for clouds and surface albedos may also contribute to errors in total column ozone (e.g. Antón et al., 2009a)., sources that provide accurate and stable long-term datasets can potentially be The bias parameterization may range from being spatially and temporally global to more local. As well, bias estimation and removal can be performed prior to generating the analyses during assimilation (as done in van der A et al. 2010 and 2015 and in this work for ozone) or during the objective analysis step (e.g., Innes et al., 2013; Dee and Uppala, 2009; Dee, 2005). The latter requires that the objective analysis setup is able to incorporate this task and directly provides the potential of reducing systematic differences between the model's short-term forecasts and the observations. This approach is being applied operationally by the European Centre for Medium-Range Weather Forecasts' (ECMWF) Integrated Forecast System (IFS, 2015) for both weather and ozone using a variational bias correction scheme. For ozone assimilation with the IFS, partial column profiles from the Solar Backscatter Ultraviolet instrument (SBUV/2; e.g., Bhartia et al., 2013; McPeters et al., 2013) on satellites of the National Oceanic and Atmospheric Administration (NOAA) are used to provide corrections for other sources. as an anchor for the other assimilated ozone datasets. The anchor is the observation dataset serving as reference, or serving to establish the reference, about which bias estimation and correction are performed. This system is also applied with the Monitoring Atmospheric Composition and Climate project (MACC; e.g. Inness et al., 2013; and MACC-II Final Report, 2014) with profile measurements from both SBUV/2 and the Microwave Limb Sounder (MLS) on the Aura satellite serving as anchor. An intermediate version consists of applying two objective analysis steps: The first step assimilates the anchor data, followed by a separate bias estimation and correction for the initially unassimilated data that considers the differences between observations and analyses from the first step.

The validation of satellite remote sounding products usuallyoften includes a comparison to ground-based measurements, which provide a long-term reference record. For satellite instruments measuring column ozone, this typically consists of comparisons to Brewer and Dobson spectrophotometers, and potentially filter ozonometers. The main advantage of ground-based versus satellite total column ozone measurements is that they can view the sun directly as oppose to relying on the backscatter of solar radiation, reducing the complexity and error sources of retrievals. The final resulting systematic errors of the calibrated ground-based total column ozone daily averages for well-calibrated and maintained Brewer and Dobson instruments are no larger than ~in the neighbourhood of roughly 1.5-2 %, % or better, excluding sites with outlier characteristics (considering Fioletov et al., 1999, 2005, and 2008). Much of the ground-based total column ozone data may be available soon after the measurements, with the original calibration usually being sufficient. For exceptional cases where the original calibration may have been faulty, a final calibration for the ground-based total column ozone may lag by one to two years from near-real time. Previous studiesStudies have examined the dependence of the differences between the satellite-based and ground-based total column ozone measurements on latitude, solar zenith angle, viewing zenith angle, seasonal dependence, cloud cover, reflectivity, and the ozone effective temperature, as well as other factors, for instruments such as for the Ozone Monitoring Instrument (OMI; Balis et al., 2007a; Viatte et al., 2011; Koukouli et al., 2012; Bai et al., 2016; Balis et al., 2007b; Viatte et al., 2011), the Global Ozone Monitoring Experiment-2 (GOME-2; BalisKoukouli et al., 2007b2012; Hao et al., 2014; Antón et al., 2009b, 2011; Balis et al., 2007a; Loyola et al., 2011; Koukouli et al., 2012 and 2015; Lerot et el.,, 2014; Hao et al., 2014; Garane et al., 2018), and the Ozone Mapping Profiler Suite (OMPS; Bia et al., 2013, 2016; Flynn et al., 2014), as well as studying their long-term stability (van der A et al., 2010 and 2015). In this paper, an observation dataset that serves as a reference in a bias estimation is referred to as the *anchor*. Reanalysis studies covering many years, such as van der A et al. (2010 and 2015), have directly used ground-based data as the anchor. A limitation in the use of ground-based observations as an anchorreference in bias estimation is that these observations are only available for certain locations, leaving many areas uncovered, especially in the Southern Hemisphere and over oceans. For the Southern HemisphereIn such cases, the applied bias parameterization may not necessarily capture as much of the absolute spatial or instrument-to-instrument relative bias variations of the bias as compared toas using observations from a satellite-borne instrument that coverscovering a larger domain. If such a satellite-based anchorreference is employed, it should ideally be in good agreement with the ground-based measurements. Considering the limited projected lifetimeslifetime and possible deteriorationsdeterioration or failures of satellite-based instruments, transitions to new references would also need to be requiredenvisaged in an operational settingsettings.

The focus of this study is specific to the assimilation of retrieved total column ozone observations from satellite remote sounding instruments using the Environment and Climate Change Canada (ECCC) weather assimilation system adapted for constituent assimilation. This work is being performed toward improving ECCC UV Index forecasts (Tereszchuk et al., 2018) and toward contributing to the monitoring of the ozone layer and to improving temperature forecasts through the radiative impact of ozone analyses (e.g., de Grandpré et al., 2009). While not unique to this work (e.g., Innes et al. 2013; van der A. et al., 2010 and 2015), a major aspect of this undertaking is the harmonizing of the observations from multiple sensors through bias correction.

In this paper, we examine the effects of assimilating different individual or combined total column ozone satellite remote sounding observations with and without bias correction on short-term forecasts of ozone. These assimilations will be univariate ozone assimilations and utilize operational ECCC weather analyses. Six-hour forecasts are used when directly comparing forecasts from different experiments. The assimilated data sources used in this study are OMI aboard the Aura research satellite (Levelt et al., 2006), the GOME-2 instruments on the European MetOp-A and MetOp-B satellites (Callies et al., 2000; Munro et al., 2006), and the total column measuring instrument of OMPS (Dittman et al., 2002a,b; Flynn et al., 2006) on the Suomi National Polar-orbiting Partnership (S-NPP) satellite. Total column ozone bias estimation for observations can be performed in different ways and depend on different factors, such as the solar zenith angle (SZA), latitude, and season, among others. Seasonal and related latitudinal changes in biases may result from limitations in retrieval algorithms. For example, the retrieval algorithm might not adequately account for the temperature dependence of the ozone absorption coefficients. Differences and limitations in accounting for clouds and surface albedos may also contribute to errors in total column ozone (e.g.,. In view of interest in near-real time assimilation, criteria applied in this paper for the selection of an anchor for total column ozone bias correction of the Antón et al., 2009a). Bias parameterizations may range from being spatially and temporally global to more local.

The harmonization of different datasets through bias correction can be applied instruments include the availability of large near-real time daily datasets with good spatial coverage over multiple years, and having differences from well-calibrated ground-based data being stable over time and close to zero on average. Other studies, such as van der A et al. (2010 and 2015) for standalone analyses, reanalyses, and in near-real time data assimilation. The assimilation process consists of introducing information from observations into model forecasts through the generation of analyses, the statistical blend of earlier forecast and observations, which serve as the initial conditions for subsequent forecasts. The assimilation of column and stratospheric ozone measurements for ozone-layer forecasting has been conducted mostly as of about twenty five years ago, ultimately culminating with operational ozone-layer and UV-index forecasts (e.g., Lahoz and Errera, 2010; Inness et al., 2013). This typically involves the application of measurements from single to multiple satellite remote sounding instruments with the use of ground-based and other remote sounding data for independent verifications and, occasionally, bias correction.

Traditionally, the assimilation process assumes that both the model forecasts and observations are statistically unbiased following an initial spin-up time (unless biases are estimated within the analysis step). Unremoved biases or systematic errors in the observations or forecasting model can potentially impact the quality of the analyses and forecasts (e.g., Dee, 2005; Dragani and Dee, 2008). This is important for total column ozone when it comes to monitoring for multi-decadal trends, as referenced in van der A et al. (2010), for both trends inferred from just the observations themselves or from their use within a data assimilation system. Generally, while the effectiveness of bias correction schemes in removing biases is constrained by limited knowledge of the truth, their impact in reducing relative biases between different assimilated observations and/or correlated fields can potentially be just as significant for improved forecasting. An example of the later is in multivariate assimilation, where ozone and meteorological assimilation can be coupled (e.g., Dee, 2008; Dee et al., 2011).

Ideally, the anchor used within a bias correction scheme should be accurate~~ covering many years~~, have a wide range of coverage in both space and time, and for near-real time applications be available within a few hours or less after measurements are taken. The summed~~instead directly used ground-based data. Summed~~ partial columns from SBUV/2 satellite instruments have been recommended as an~~the~~ anchor for long-term studies (Labow et al. 2013). This is due to ~~) considering~~ the long-term ~~time~~ coverage provided by the series of SBUV/2 instruments, combined with the low variations in ~~variation over~~ time of the~~overall~~ differences between these instruments and~~from~~ ground-based data (~~remaining~~ usually within ±1 %, but~~with~~ reduced ~~differences~~ for recent years).~~. As such, SBUV/2 data, which also satisfy the above mentionned criteria, could serve as anchor for bias correction.~~ Labow et al. (2013) also show differences over time of SBUV/2 with OMI data remaining within about 1 and 2 % for the Northern Hemisphere (based on the Total Ozone Mapping Spectrometer (TOMS) version 8.5 total column retrieval algorithm, an enhancement of the version 8 algorithm described by ~~(~~Bhartia and Wellemeyer, 2002).~~) remaining between about 1 and 2 %.~~ McPeters et al. (2015), showing similar magnitudes and stability of differences in time, concluded~~conclude~~ that OMI-TOMS data could~~can~~ be used in trend studies. The merging of OMI with SBUV/2 and earlier TOMS instrument data for this purpose was performed by Chehade et al. (2014). ~~Since OMI is a mapping instrument, thus allowing for more colocations with measurements from other instruments over a few days, the OMI-TOMS dataset was instead selected as the anchor in this work. Verifications are performed with the OMI dataset and other independent measurements. In turn, the OMI-TOMS dataset is compared to the data from Brewers, Dobsons, filter ozonometers, and SBUV/2 over the examined time period.~~

The focus of this study is bias estimation and correction of column ozone for multiple satellite sensors, towards eventual use in near-real time data assimilation. The bias estimation and correction methods developed in this study may be integrated into an assimilation scheme, and so can be applied in near-real time, and could be utilized for other constituents. In this paper, we evaluate several different bias estimation schemes used to correct observations of column ozone from satellite-borne instruments. Many of these methods utilize colocated observation sets for bias estimation. From this consideration, OMI-TOMS was chosen as the anchor for bias estimation and correction, as its dense spatial coverage allows for more colocations with measurements from other instruments. As part of this work, the OMI-TOMS column ozone data were evaluated using ground-based Brewers, Dobsons, and filter ozonometers observations, as well as compared to SBUV/2 column ozone, for the limited time periods of interest in this study. For these datasets, a target maximum residual bias of 1 % following bias corrections was selected. This satisfies the column ozone 2 % accuracy requirement from European Space Agencies' Climate Change Initiate program (Van Weele et al., 2016) for random error levels of up to 1.7 %.

In this paper, we examine several bias correction methods that use a discrete binning in latitude and solar zenith angle that, unlike a functional parameterization, allows for arbitrary nonlinear dependencies. In addition, an alternative estimation involving the dependency on the ozone effective temperature (the mean temperature weighted by the ozone profile), as employed in van der A et al. (2010 and 2015), was explored. However, as discussed later in the paper, dependencies on factors such as changes in cloud cover and viewing zenith angle, were not examined.

Following bias estimation, data assimilations of column ozone observations from individual and multiple satellite instruments were conducted with and without bias correction. The impacts on the resulting six-hour forecasts were then assessed. The assimilations were conducted with the Environment and Climate Change Canada (ECCC) meteorological assimilation system adapted for constituent assimilation. These assimilations were univariate ozone assimilations and utilized operational ECCC meteorological analyses. The data sources assimilated in this study and correspondingly involved in the bias estimation analysis are the GOME-2 instruments on the European MetOp-A and MetOp-B satellites (Munro et al., 2016; Hassinen et al., 2016), the total column measuring instruments of OMPS (Dittman et al., 2002a,b; Flynn et al., 2006) on the Suomi National Polar-orbiting Partnership (S-NPP) satellite, and OMI aboard the Aura research satellite (Levelt et al., 2018).

This paper is organized as follows: Section 2 describes the utilized ozone observations covering July-August 2014 and 2015 and January-February 2015. Following a general quality assessment of the OMI data based on available literature, Section 3 evaluates the OMI column ozone data for these periods against ground-based measurements. Having assessed the quality the OMI data for these specific periods, Section 4 describes and applies three different bias estimation approaches with the column ozone measurements of different satellite instruments relative to OMI. The impact of column ozone assimilation on six-hour forecasts for individual and multiple sensors with and without bias corrections is examined in Section 5 for July-August 2014 using comparisons to both OMI-TOMS and ground-based data. A time varying bias dependent on both latitude and solar zenith angle is estimated that uses the differences with colocated total column ozone OMI measurements. Other dependencies not included in this study, such as changes in cloud cover and viewing zenith angle, would contribute to the resultant observation error standard deviations estimated in this work and in contributions to the far wings of the observation error probability distribution, and for some cases might result in outliers whose influence would be screened out or reduced through the assimilation quality control. A discrete binning in latitude and solar zenith angle will be used instead of a functional parametrization to allow for arbitrary nonlinear dependencies. These bias estimates will be compared to bias estimates that use differences between observations and short-term forecasts (innovations) in addition or as an alternative to colocation differences. The dependence on the ozone effective temperature is briefly evaluated in comparison to the temporal dependence with latitude.

A description of the assimilation and forecasting system, the observations, and the bias correction approaches is provided in Section 2. Section 3 consists of an evaluation of the OMI-TOMS data over the different two month time periods studied in this paper, the examination of bias estimations and corrections with a few approaches, and the analysis of relative impacts of the different assimilation scenarios. Conclusions are provided in Section 64. The Supplemental material document forto this paper provides additional figures and tables supporting and complementing the discussed and presented results; only, the tablesfigure and table numbers of which are directly referenced inidentified beginning with the paperletter 'S'.

2 Experimental setup and methodology

2.1 The assimilation system

Global ozone assimilation experiments are performed with the operational Global Environmental Multiscale (GEM; Côté et al., 1998a and 1998b; Charron et al., 2012; Zadra et al., 2014a,b; Girard et al., 2014) numerical weather prediction (NWP) model coupled to a linearized ozone model (LINOZ) (McLinden et al., 2000; de Grandpré et al., 2016) using incremental three-dimensional variational (3D-Var) assimilation with first guess at appropriate time (FGAT; Fisher and Andersson, 2001). The 3D-FGAT assimilation with GEM-LINOZ uses components of the ECCC Ensemble -Variational data assimilation system (Buehner et al., 2013 and 2015) adapted by the authors and P. Du (ECCC) for constituent assimilation and being run without ensembles. The LINOZ model uses pre-computed coefficients generated as monthly mean climatologies for calculating the ozone production and sink contributions at each time steps throughout the stratosphere. A relaxation toward the GEM ozone climatology (Fortuin and Kelder,, 1998) is imposed below 400 hPa to constrain deviations away from the climatology with a relaxation time scale of 2 days.

The GEM model is executed with a 7.5 min. time step for a uniform 1024×800 longitude-latitude grid and a Charney-Phillips vertically staggered grid (Girard et al., 2014; Charney and Phillips, 1953) with 80 thermodynamic levels extending from the surface to 0.1 hPa. The horizontal grid corresponds to a resolution of ~0.23° in latitude and ~0.37° in longitude, representing a 25 km resolution at latitude 49°. The vertical resolution in the upper-troposphere/lower-stratosphere (UTLS) region is in the range of 0.3 to 0.6 km with the resolution gradually changing to ~1.6 km at 10 hPa

and 3 km at 1 hPa. Successive short-term three to nine hour forecasts were generated from analyses provided for 00, 06, 12, 18 UTC synoptic times.

The analyses are a composite of the already available ECCC operational weather analysis and the ozone analyses generated separately from assimilations in the study. The same process of updating the weather fields every six hours is also applied for the reference case without ozone assimilation.

Incremental assimilation is conducted at each six-hour interval with T280 spectral space background error covariances, an FGAT time resolution of 45 min., and increments at the model vertical levels on a 600x300 Gaussian grid with final increments interpolated to the model grid. As part of the background error covariances, the third order autoregressive correlation model (TOAR; Gaspari and Cohn, 1995) was applied in specifying spectral space globally homogeneous, isotropic and separable horizontal and vertical error correlations. Initial vertical correlations were calculated in physical space on the model vertical grid using monthly six-hour time differences of the free running GEM-LINOZ model as proxy to six-hour forecast errors (Polavarapu et al., 2005; Jackson et al., 2008; Section 5.5 in Bannister, 2008). The resulting static monthly error correlations are only rough approximations as compared to more realistic non-homogeneous, non-isotropic and state dependent error correlations which could be estimated through ensembles; the latter is not done and not essential for this study. Fits to the TOAR correlation model were applied to the correlations for each level to give localized non-negative correlations (see Figs. S1 and S2 in the Supplement). Smoothing was applied to the vertical level-dependent set of derived e-folding error correlation lengths of the TOAR functions, defined as distances for correlations of e^{-1}. The vertical error correlation lengths are similar to those implied by the correlation length scales shown in Massart et al. (2009; accounting for approximate scaling from d[ln p] to km). The vertical error correlation lengths above 100 hPa correspond to half width at half maximum values (multiplying correlation lengths by ~1.2) that nearly equal the model vertical resolutions with values ranging from ~0.5 km at 100 hPa to ~3.5 km at 1 hPa. The horizontal error correlation lengths are ~125 km near the surface and increase from ~165 km at 100 hPa to just under 750 km at 1 hPa. A summary of the applied background and observation error variances is provided in Section 2.3.

In addition to the preliminary removal of data with large observation minus forecast differences (OmF; innovations) during background check, variational quality control (Andersson and Järvinen, 1999; e.g. Gauthier et al., 2003) is applied to the observations as part of the objective analysis step after the fifth iteration of the minimization. This reduces the weight of the observations associated to large increments at each minimization step in accordance to the size of the differences between the observations themselves and the latest analysis estimates from the most recent minimization iteration.

The assimilation experiments are conducted for July-August 2014, with a start date of 28 June 2014, 18 UTC, with and without bias correction for individual and combined datasets. The initial weather conditions for 28 June are from ECCC operational weather analyses and the initial ozone field is an analysis from an earlier assimilation.

2.2.2 Observations

In this section, we give a brief description of the column ozone observations made by the satellite-based instruments that are assimilated as well as the verification sources. The sources of the assimilated observations offer near-real time (NRT) column ozone products available for operational use as well as other products with additional or different processing resulting in a lag from NRT. The datasets selected for assimilation are being considered by ECCC for potential future NRT operational assimilation for contributing to both ozone-layer and UV index forecasting. All assimilated data sources in the study, namely the OMI, GOME-2, and OMPS-NM (total column Nadir Mapper) use optical solar backscatter of ultraviolet radiation in the nadir or near-nadir and provide data only during daytime. The data for GOME-2, OMPS, and SBUV/2 are being acquired in BUFR format from the National Environmental Satellite, Data, and Information Service (NESDIS/NOAA). For OMI we have used the standard science data column ozone products which are close to, but can differ slightly from, the NRT data (OMI NRT Data User's Guide, 2010; Durbin et al., 2010). The OMI data is obtained from the Earth Observing System Data and Information System of the National Aeronautics and Space Administration (EOSDIS/NASA). These differences between the two datasets are examined in Section 3.1. Thinning (reducing the pixel map density on the ground) to a resolution of ~1° is applied to all total column measurements in consideration of the background horizontal error correlation lengths and not including observation error correlations and is performed after the colocation-based bias estimation. The different sources of total column ozone observations are incorporated separately and not as an average when generating the analysis increments from multiple observation sources. The alternative of statistically averaging satellite observations from the different sources (e.g., van der A et al., 2010) In this section, we give a brief description of the column ozone observations involved in the implementation and evaluation of bias correction as well as the observations used for the validation of short-term forecasts. Observational data set were obtained for the periods of July-August of 2014 and 2015, and January-February 2015. The main data sources of interest are those specifically intended to provide satellite-based column ozone allowing near-real time (NRT) assimilation. These consist of OMI, GOME-2, and OMPS-NM (total column Nadir Mapper) instruments that rely on optical solar backscatter of ultraviolet radiation in the nadir or near-nadir and provide data only during daytime. Ground-based Brewer, Dobson, and ozonometer filter instruments and additional satellite-based data from OMPS-NP (partial column Nadir Profiler) and SBUV/2 are included for evaluation and validation purposes.

is not done here and would decrease computational cost. As different nadir mapping satellite instruments relying on solar backscatter radiation typically view similar horizontal regions over a few hours with measurements of sufficient precision for assimilation, one might argue that assimilating multiple sources is not necessary. On the other hand, assimilating data from different sources ensures that data availability gaps occurring in time and space from one instrument can be offset by the presence of other instruments and ensures temporal continuity following the interruption or end of an instrument's operation. Using multiple instruments can also serve in mutual near-real time monitoring of the different sources when it comes to temporal changes in biases and in better ensuring an automated transition to a new and acceptable reference source once the original anchor is no longer available.

~~The mean number of observations in a six-hour time period for the total column ozone measurement instruments under consideration are in the range of roughly 30 to 200 thousand depending on the source, with a reduction by factors of 5 to 20 from thinning (Fig. S3 of the Supplement). Through their much larger numbers (e.g. each by a factor of ~20 or more), the thinned OMI, GOME-2, and OMPS-NM total column measurement sets result in a stronger influence on the total ozone analyses and short-term forecasts as compared to profilers such as OMPS-NP and MLS, the later used here mainly for verifications. Adding the assimilation of profilers, while not done here except for MLS in one experiment, would serve to constrain the ozone field vertical structure.~~

~~Independent verifications of short-term forecasts are performed using both column and profile measurements from satellite, in-situ, and ground-based instruments. These sources are briefly identified following a description of the assimilated data sources.~~

~~While assimilation experiments are conducted only for the July-August 2014 period, bias corrections and evaluations are also done for July-August and/or January-February 2015 to illustrate possible differences and similarities in time and between seasons and successive years.~~

~~2.~~2.1 OMI

The Ozone Monitoring Instrument (OMI) aboard the Aura research satellite has been in operation since August 2004. The instrument stems from a collaboration between the Netherlands Agency for Aerospace Programmes (NIVR), now called the Netherlands Space Office (NSO), and the Finnish Meteorological Institute (FMI). The OMI instrument provides a cross-track width of about 2600 km on the ground and total column ozone mapping at a spatial resolution of 13 km along, and~~x~~ 24 km across, the orbit ground track at nadir (e.g.,~~.~~ Bhartia and Wellemeyer, 2002; OMI Data User's Guide, 2012). Some strips of the OMI measurement tracks were removed due to~~are missing as a result from flagging of~~ the row anomaly of the OMI instrument, which for the time period under consideration, effects 23 of the 60 rows.[1]

Two different ~~level 2~~ total column ozone products are derived from the OMI radiances, one processed by NASA based on the Total Ozone Mapping Spectrometer (TOMS) ~~version 8 (V8)~~ total column retrieval algorithm (versions 8 and 8.5) and the other made by the Royal Netherlands Meteorological Institute (KNMI) using the Differential Optical Absorption Spectroscopy (DOAS) algorithm. The OMI-TOMS algorithms~~V8 algorithm~~ (Bhartia and Wellemeyer, 2002) ~~uses~~ principally utilizes only two different wavelengths, one with strong and one with weak ozone absorption, to estimate the total column ozone and surface reflectivity. In the DOAS algorithm (Veefkind and de Haan, 2002; Veefkind et al., 2006), first the slant column density is retrieved from a spectral least squares fit to the measured ratio between the Earth radiance to solar irradiance using 25 wavelengths spanning 331 to 337 nm. The slant column density is then converted to a vertical column using the air mass factor.

[1] See http://projects.knmi.nl/omi/research/product/rowanomaly-background.php for more information and updates regarding the OMI row anomaly.

Overall, the two different retrievals agree to a high degree, with the global average falling within 3 % of one another, with the largest differences occurring for cloudy conditions and in the polar regions (Kroon et al., 2008).

This study employs the OMS-TOMS V8. Long-term term stability, relatively little solar zenith angle and latitude dependence, and small overall biases (averaging to roughly 1.5 standard science% or better) have been observed in the OMI-TOMS retrieved data when compared to ground-based total column ozone products which are close to, but can differ slightly from, the OMI-TOMS NRT data (OMI NRT Data User's Guide, 2010; Durbinmeasurements as well as being relatively stable over time (Koukouli et al., 2012; Balis et al., 2007b; McPeters et al., 2008; in addition to Labow et al., 2013, and McPeters et al,, 2015, referenced in the introduction). OMI-TOMS total column ozone shows no to little (<1 %; but up to ~2% for cloud top pressure) dependency on cloud fraction, reflectivity, or cloud top pressure, depending on the reference when compared to ground-based Brewer and or Dobsons (Bak et al., 2015; Bai et al., 2015; Balis et al., 2007b; Antón et al., 2009a; Antón and Loyola, 2011; Koukouli et al., 2010).2013) with increasing scatter for more overcast conditions (e.g. Balis et al., 2007b); ground-based column ozone data would also have errors associated to cloud cover.

The paper by van der A et al (2010 and 2015) indicate negligible variation with viewing zenith angle but a dependence on effective ozone temperature which reflects some dependence on latitude and time (e.g., Koukouli et al., 2016) comparable in magnitude to other satellite instruments. However, van der A. et al. (2015) also include a second order correction as a function of solar zenith angle (SZA) which is small as compared to most other instruments, and about the same or smaller than SBUV/2 instruments, but still non-negligible. We choose the OMI-TOMS total column ozone product as the anchor in our bias correction scheme, as already indicated in the introduction, and will examine the dependence with effective ozone temperature in Section 3. While not done here, one could generally adjust satellite-based anchors according to constant offsets, and long-term trends if needed, of differences with ground-based datasets, as done by van der A et al. (2010, 2015) for example.

The OMI-TOMS data have estimated root-mean squared errors of 1-2 % (OMI Data User's Guide, 2012). Comparisons of the standard OMI-TOMS data to ground-based measurements over the application periods for this work are provided in Sections 3.1. The corrections of van der A et al. (2015) for OMI-TOMS of a constant offset of ~3.3 DU plus a linear function of effective ozone temperature has not been applied here, nor the ~1,5 % offset determined in other studies (e.g. McPeters et al., 2015; Bak et al., 2015; Labow et al., 2013). There is also a second order non-negligible correction as a function SZA in van der A et al. (2105), even though small compared to most other instruments, which is not present in van der A. et al. (2010). The implications of these corrections for the periods in the study are considered in Section 3.1.

The OMI NRT Data User's Guide (2010) and Durbin et al. (2010) indicate a daily maximum percentage difference of 2.6 % between the for the column ozone OMI-TOMS NRT products as compared to the OMI-TOMS standard science and NRT products, with a weekly average maximum difference of 1.4 %. Further comparisons by the authors showTime mean differences generally between 0.02-0.04%of the standard and NRT product versions as a function of latitude and solar zenith angle were calculated for July-August 2016 and January-February 2017. The OMI-TOMS column ozone has estimated root-mean-squared errors of 1-2 % (OMI Data User's Guide, 2012). as the OMI-NRT observations were no longer available for the 2014-2015 time periods when this work was undertaken. All time mean differences over the entire range of bins for both periods were determined to be within roughly 0.02-0.04% (Fig. S4), no larger than the 0.1 DU storage accuracy for the data files used to store the total column ozone observations, with the majority of individual differences being below the storage accuracy. While some individual point differences can be larger, time mean differences suggest that the near-real time and standard OMI-TOMS products would generate nearly equal bias estimates when serving as anchor.

2.2.2 GOME-2

Global Ozone Monitoring Experiment-2 (GOME-2) instruments are on the MetOp-A (GOME-2A) and MetOp-B (GOME-2B) polar orbiting satellites, launched in October 2006 and September 2012, respectively, and are operated by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT). As of July 15 2013, GOME-2A has been operating with a swath width of 960 km and a 40 km × 40 km spatial resolution, while GOME-2B has a larger swath width of 1920 km and a 40 km × 80 km spatial resolution (e.g.,., GOME-2 ATBD, 2015; ATBD stands for the Algorithm Theoretical Basis Document-).

Total column ozone retrievals are available from EUMETSAT relying on the DOAS approach (Loyola et al., 2011) and from the National Environmental Satellite, Data, and Information Service (NOAA/NESDIS/NOAA) with retrievalsthe retrieval based on the TOMS V8 algorithm (e.g., Zhand and Kasheta, 2009). The DOAS total column ozone products are indicated to have estimated accuracies of better than 3.6-4.3 % (for clear to cloudy conditions) and 6.4-7.2 % for SZA below and above 80°, with precisions of under 2.4-3.3 % and 4.9-5.9 % (GOME-2 ATBD, 2015; GOME User Manual, 2012; GOME-2 ATBD, 2015).). The GOME-2operational NRT products used here, as well as those form OMPS and SBUV/2, were acquired from NOAA/NESDIS/NOAA and stem from the TOMS approach. ThisTheir uncertainty characteristics estimated in this study are provided an opportunity to evaluate the biases of the GOME-2 TOMS products. in Sections 2.3 and 3.2.

2.2.3 OMPS

The Ozone Mapping Profiler Suite (OMPS) on the Suomi National Polar-orbiting Partnership (S-NPP) satellite, launched October 2011, consists of a combined nadir mapper (OMPS-NM) and nadir profiler (OMPS-NP),) and a separate limb profiler (OMPS-LP), which provide total column, partial column profile, and limb profile products, respectively. A second suite was placed onboardon board the Joint Polar Satellite System JPSS-1 satellite (Zhou et al., 2016), renamed NOAA-20 and launched in November 2017. The retrieved data used in this study are from the OMPS S-NPP nadir measurements and are considered to be at a provisional product maturity level. They do not include improvements from the various corrections, calibration adjustments, and retrieval algorithm updates performed since the original near-real time acquisition for the July-August 2014 period (personal communication from L. Flynn, NOAA, 2016). Only the nadir mapper data is assimilated, with the summed partial columns of the nadir profiler also evaluated during bias correction. The nadir mapper has a cross-track width of about 2800 km and a 50 km × 50 km resolution at nadir. The OMPS-NM retrievals, summarized by Flynn et al. (2014), were made at the NOAA Interface Data Processing Segment using the ratio of the measured Earth radiances to solar irradiances at multiple triplets of wavelengths. Flynn et al. (2014) provide total column ozone accuracy and precision requirements of ~3.5-4 % and ~2 % for SZA up to 80° and found average biases of -2 to -4 % with respect to the OMI-TOMS and the Solar Backscatter Ultraviolet SBUV/2 satellite instrument products.

The OMPS-NP provides profiles with a resultant 250 km × 250 km field of view. Only their summed partial column values are applied in the paper, and this only for comparison to the other datasets regarding total column biases. Profiles used were obtained with an implementation of the Version 6 SBUV/2 instrument algorithm (Bhartia et al., 1996) with the a priori profiles derived from the OMPS-NM. The OMPS-NM and OMPS-NP ozone retrievals from the SBUV V8.6 retrieval algorithms (Bhartia et al., 2013; as referred by Bai et al., 2016) became available after the completion of the assimilation experiments conducted for this work. The OMPS-NP V6 profiles consist of 12 layers with the profile layer bottoms (in hPa) of 1013 (surface), 253, 127, 63.3, 31.7, 15.8, 7.92, 3.96, 1.98, 0.99, 0.495, 0.247. The provisional OMPS-NP V6 product has the lowest layer extending to sea level instead of the surface. As a correction, the partial column values of the lowest layer were reduced by scaling according to the reduction in the pressure layer thickness of the lowest layer when moving the lower boundary from the sea level to the actual surface. Ozone profile accuracy and precision requirements are both 5-10 % in the middle to upper stratosphere for SZA up to 80° with larger uncertainty at lower and higher altitudes (Flynn et al., 2014). Based on the same reference, the OMPS-NP layer data would generally be consistent with corresponding SBUV/2 data processed with the V6 algorithm. The OMPS-NP data in the South Atlantic Anomaly region are recommended to be discarded (personal communications from L. Flynn) while total column measurements in this region are retained.

WhileThe OMPS-NM retrievals, summarized by Flynn et al. (2014), were made at the NOAA Interface Data Processing Segment using OMPS data used are not from the ratio of latest applied total column retrieval algorithms, the measured Earth radiances to solar irradiances at multiple tripletsresults obtained in Section 3 can serve in comparing to the overall quality of more recent dataset versions. The evaluation of wavelengths. Bai et al. (2013) for the OMPS-NPThe nadir mapper has a cross-track width of about 2800 km and a 50 km × 50 km resolution at nadir. Flynn et al. (2014) provides total column ozone accuracyobtained based on the TOMS Version 7 algorithm and covering January 2012 to February 2013 gives global mean differences with ground based data of 0.21 % for Brewer measurements and precision requirements of ~3.5-4 % and ~2 %, respectively, for SZA up to 80° and found average biases of -20.86 % for Dobson measurements, each with standard deviations close to -4 % with respect to the OMI-TOMS and the Solar Backscatter Ultraviolet SBUV/2 satellite instrument products.

3 %. The results of the evaluations from Bai et al. (2015, 2016) for the more recent OMPS-NM total column ozone products based on the SBUV V8 and V8.6 retrieval algorithms, respectively, are consistent with Bai et al. (2013). Bai et al. (2015) indicate global mean differences of OMPS-NM with ground-based data of 0.59 % for Brewer measurements and 1.09 % for Dobson measurements, with standard deviations close to 3 % and for the same period as Bai et al. (2013). As a reference, Bai et al. (2016) provide , as reference, a distribution of OMPS-NM minus OMI-TOMS values with a mean of 7.6 DU (~2.5 % for a total column of 300 DU) and a standard deviation of 5.8 DU at the Tsukuba station (36.1° N, 140.1° E) covering the period

of 2012 to early 2015., with OMI being closer to the Dobson ground-based data (accounting for the seasonal variation of the differences).

OMPS-NP profiles, each with a 250 km × 250 km field of view on the ground, were provided from2.2.4 Independent verification sources

Other than an implementation of the Version 6 SBUV/2 instrument algorithm (Bhartia et al., 1996) with the a priori profiles derived from the OMPS-NM. This version of the OMPS-NP data provide profiles on 12 layers. See Flynn et al. (2104) for a description of the accuracy and precision of the OMPS-NP V6 products. While only the nadir mapper data were assimilated in Section 5, both the nadir mapper and the summed partial columns of the nadir profiler were evaluated during bias correction.

2.4 Independent verification sources

Ground-based and satellite-based column ozone data serve as independent verifications of the OMI-TOMS measurements, with the former also used for validation of the forecasts resulting data assimilation. These data are described below.

2.4.1 Ground-based data

The ground-based datadata which serves as anchor, the verification data sources consist mostly of Brewer, Dobson,Dobsons and filter ozonometer total column ozone measurements (Fioletov et al., 1999 and 2008; Staehelin et al., 2003)and ozonesonde profiles from the World Ozone and Ultraviolet Radiation Data Center (WOUDC) and, secondly, of; for total column ozone measurements see Staehelin et al., 2003; Fioletov et al., 1999 and 2008; for ozone sondes, see references specified in Dupuy et al., 2009), Brewer and Dobson measurements from the Global Monitoring Division of the NOAA Earth System Research Laboratory (see Coldewey-Egbert et al., (2015) for various references on the validation of , total column ozone from the Solar Backscatter Ultraviolet instrument (SBUV/2) on the NOAA 19 satellite data with ground-based Brewers and Dobsons). Only direct sun, clear-sky daily daytime averages from these instruments were used.(McPeters et al., 2013; Bhartia et al.,, 2013; Flynn, 2007), and ozone profiles from the Microwave Limb Sounder (MLS) on the Aura satellite (Waters et al., 2006; Livesey et al., 2006, 2013; Froidevaux et al., 2008; e.g., Inness et al., 2013; Adams et al., 2014), The error standard deviations for Brewer and Dobson direct sun data are no larger than ~1.5 to 2.0 % for well-calibrated and well-maintained instruments Brewer and Dobson direct sun data and about 1.5 to 2 times larger for filter ozonometers (based on Fioletov et al., 1999, and references therein; Fioletov et al., 2008). Consistent), 5 % for ozonesondes (e.g. SPARC, 1998; Smit and Sträter, 2004), and as provided with the data for MLS (Livesey et al., 2013). The main disadvantage of filter ozonometers, relative to Brewers and Dobsons, is that the two spectral filters, i.e., the two wavelengths, used in ozone retrieval make them sensitive to atmospheric aerosols. Vertical piecewise averaging was applied to the ozonesonde profiles for final resolutions of 250 Pa up to an altitude of 25 km and 50 Pa for altitudes above 35 km, with the above, ana linearly varying resolution between 25 km and 35 km.

The applied ground-based Brewer and Dobson spectrophotometer and filter ozonometer total ozone data are direct sun (clear-sky) daily daytime averages. An overall precision of 4.6 DU has been obtained by van der A et al. (2010) for Brewer

Formatted: Default Paragraph Font, English (United States)

Formatted: Default Paragraph Font, English (United States)

Formatted: Default Paragraph Font, Font: Not Bold, English (United States)

Formatted: pb_toc_link, English (United Kingdom) and Dobson direct sun daily averages, excluding outlier data.~~, which is consistent with standard deviations being less than 2 % based on Fioletov et al. (1999 and 2008); the precision for filter ozonometers appear larger by roughly less than a factor of two.~~ As in van der A et al. (2010) and Koukouli et al. (2016), the Dobson ozone values were adjusted following the correction of Komhyr et al. (1993; see also van Roozendael et al., 1998) as a function of ozone effective temperature (-0.13 % K^{-1} about 227 K; The ozone effective temperature is the average value of the ozone-weighted temperature profile.).~~).~~ This correction is not applied to Brewer data in this study following van der A et al. (2010) based on Kerr (2002). The results of Redondas et al. (2014) support neglecting the small sensitivity to ozone effective temperature for Brewer measurements but~~as well as~~ accounting for the larger sensitivity for Dobson values. Avoiding~~A consequence of avoiding~~ the correction for Dobsons results in~~latter is~~ a seasonal dependence of the Dobson total column ozone errors. The calculated seasonal variations of differences of OMI-TOMS and OMI-DOAS with Brewer and Dobson instruments in Balis et al. (2007a~~2007b~~) further support neglecting corrections to the Brewer data (~~,~~ if we exclude consideration of results at the equator and in Antarctica which rely only on one station each).~~,~~ while favouring including the corrections for Dobsons. Leaving Dobson measurements uncorrected would introduce~~ showing~~ a 1-2 % seasonal~~.5 % annual~~ variation (for~~of the differences. As another~~ example, Bai et al. (2016) showed~~show~~ a fairly consistent seasonal variation of the differences with the Tsukuba Dobson measurements with an amplitude of about 2 % for both OMPS-NM and OMI-TOMS).~~.~~

2.4.2 SBUV/2

Data from the Solar Backscatter Ultraviolet instrument (SBUV/2) were used for verification purposes.~~ ~~ The ozone data from SBUV/2 for the period of interest are from the NOAA 19 satellite (Flynn, 2007; Bhartia et al., 2013; McPeters et al., 2013).~~launched in February 2009.~~ Two versions of the total column ozone data are used here: The first is~~. These are~~ from the SBUV V8.6 profile retrieval using wavelengths in the range of 250 to 310 nm (Bhartia et al., 2013; summarized by McPeters et al., 2013; see also Flynn, 2007) for which the total column ozone is the sum of the partial column layers, and second is from the SBUV V8 total column retrieval using two wavelengths between 310 and 331 nm (Flynn, 2007; Flynn et al.,~~,~~ 2009). The ozone measurements cover 170 km × 170 km field of views at the ground and have separations along the satellite orbit tracks of about 170 km. Labow et al. (2013) found the agreement between total column ozone data of SBUV instruments from~~form~~ the summed partial columns and the Northern Hemisphere~~northern hemisphere~~ ground-based data to be better than 1 %. Bhartia et al. (2013) has indicated~~indicates~~ that the total column ozone values resulting from~~from the summed partial column profiles allow extending~~ the V8.6 algorithm can be used for~~total column ozone retrievals to a~~ solar zenith angles up to~~angle of~~ 88°.

3 Evaluation of OMI-TOMS total column ozone with ground-based data

Differences between OMI-TOMS and ground-based Brewer and Dobsons data have shown long-term term stability and relatively little solar zenith angle and latitude dependence (Balis et al., 2007a; Koukouli et al., 2012; Labow et al., 2013;

Formatted: pb_toc_link, Font: Bold, English (United

Formatted: Font: (Default) Times New Roman, Bold, English (United Kingdom)

McPeters et al., 2008 and 2015). Comparisons of OMI-TOMS V8.5 total column ozone with Northern Hemisphere ground-based data by Labow et al. (2013), and McPeters et al. (2015) based on multiple years indicate an average underestimation of OMI-TOMS of about 1.5 %. Figure 2 of McPeters et al. (2015) shows variations of weekly mean differences about the long-term average underestimation mostly within about ±1 %. With OMI-TOMS V8, McPeters et al. (2008) found positive average differences with Northern Hemisphere Brewers and Dobsons covering 2005 and 2006 of 0.4 % with a stations-to-station standard deviation of 0.6 %. Also, OMI-TOMS total column ozone data show no to little dependency on cloud fraction, reflectivity, or cloud top pressure (<1 %; but up to ~2% for cloud top pressure) (Balis et al., 2007b; The MLS ozone profiles are from version 3.4 of the processing algorithm (Livesey et al., 2013) and are used without their averaging kernels. The along-track sampling and effective resolutions are ~160 km and, in the stratosphere, 300-450 km. The vertical levels of the retrieval solutions are in pressure with 12 levels per decade in the stratosphere, this corresponding to a vertical resolution of ~1.3 km, with the effective vertical resolution being ~2.5 km. The recommended usable vertical range for the ozone profiles is 261 hPa to 0.02 hPa with highest precisions of 2-3 % and accuracies of ~5 % in the range ~2-46 hPa.

2.3 Applied background and observation error variances

All the presented assimilation cases, except one, will have used an originally prescribed set of ozone observation and background error variances. An updated set of error variances considered to be closer to optimality, at least for the observations, was also derived with implications shown for one assimilation experiment. These updated variances were derived using the observations to which the colocation bias correction described in Section 2.4.1 was applied. Their derivation and the results are presented in this section, in addition to the originally applied variances, to support the target for reduction in total column observation bias indicated in Section 2.4. For this updated set, a single iteration of the Desroziers et al. (2005) approach applied to total ozone served as an initial step to adjust the error variances following a first set of assimilations. Separate additional metrics were then applied to complete the ozone background and observations error variance adjustments. The resultant new error variances are considered more representative of the actual observation error variances and so a better guide for imposing target conditions on observation bias reductions. As will be briefly shown, the overall diagnostics presented in Section 3 on the quality of the short-term ozone forecasts are largely consistent between the two sets of assimilations and so the results from the former can be taken as qualitatively applicable to the latter. Both are summarized for completeness.

The original error standard deviations assigned to all total column measurements are set to 2 % for all assimilated sources. For the updated set of statistics, the observation error standard deviations were re-estimated following the assimilation with all satellite total column datasets using two main steps: First, a single iteration of the Desroziers et al. (2015) approach was performed with the adjustments parameterized as a function of SZA. The variances of the differences between colocated observations (see Section 2.4.1) between each instrument and OMI were then computed and compared to the expected sum of the individual variances, which assumes the errors on the two different observation sets are not correlated. The variances of the differences between OMPS-NM and OMI were used to obtain an estimate of the error variance of OMI, where for this estimate the error variances of OMPS-NM were replaced with the error variances of OMI scaled by the ratio of error variances between OMPS-NM and OMI as diagnosed by the Desroziers approach. The error variances of OMI were then used in the comparison of the variances of differences between collocated observations for the remaining instruments (GOME-2A/B, OMPS-NM) to estimate their error variances. The final updated error standard deviations are provided in Fig. 1. These error standard deviations, determined and smoothed as a function of solar zenith angle for each total column measurement source, are in the range of 1-2 % for SZA less than 70° and 1.5-3.5 % for higher SZAs, with the smaller values for both OMI and OMPS-NM. The obtained values are on the lower end of, or smaller than, the instrument error random uncertainties indicated in Section 2.2. On the other hand, global root-mean-square error levels obtained by van der A et al. (2010) for GOME-2A (DOAS) and OMI-TOMS from comparisons with ground based measurements were estimated to be 4.74 and 4.59 Dobson Units (DU), respectively. These translate to values in the range of ~1.3-1.9 % for column measurements between 250 and 350 DU. While not used for the applied updated values, tightening the colocation requirements for the differences with OMI from the criteria of Section 2.4.1 by factors of 4 and 2 respectively for the maximum separations in distance and time, and adding outlier removal, further reduces the final observation error standard deviations by roughly 10 % each, for a total reduction of 20 %. While colocation differences do not reflect model representativeness errors to the observation standard deviations, their contributions may not be that significant considering the respective sizes of model horizontal resolutions and the observation pixels on the ground. The actual standard deviations may therefore be smaller overall, say by 10-20 %, than those used as final values.

The original background error standard deviations as a function of latitude and vertical level are from a single iteration of the Desroziers et al. (2005) approach with an assimilation of MLS ozone data applied following originally assigned ozone background error standard deviations of 5 % in the stratosphere and upper troposphere. The resulting latitude varying background error standard deviations at sample vertical levels of 1, 10, 50, and 300 hPa are in the ranges ~6-12 %, ~3-5 %, ~5-15 %, and ~15-24 % respectively. Constant extrapolation in absolute value uncertainty was imposed for lower tropospheric levels with resultant percentage values being ~15-16 % of the Fortuin and Kelder (1998) climatology in volume mixing ratio. An update of the values assigned in the stratosphere and upper troposphere accompanied the update of the observation error standard deviations described above. This was done by scaling the original background error standard deviation field as a function of vertical level above 300 hPa by latitude dependent total ozone error standard deviation scaling factors. A linear tapering in ln(η) of the adjustment was performed for lower altitudes to preserve the original surface values applicable for assimilation of surface observations even though surface observations were not used in this study. Total ozone background error standard deviation scaling factors were derived as a function of latitude from an assimilation of all four total column data sources, namely GOME-2A, GOME-2B, OMPS-NM and OMI, the first three being bias corrected relative to OMI. The latitude dependence of the scaling was set from a single iteration of the Desroziers et al. (2005) approach. A global absolute scaling was then determined by repeating the assimilation (with the updated observation error variances) a few times and successively scaling the background error variances to ultimately give a minimization chi-square (twice the cost-function value divided by the number of observations) in the neighbourhood of unity (Fig. S5). The total ozone error standard deviations were thus reduced from the first set of values in the range of 1.8-3.4 % to 0.5-0.8 % (Fig. 2). These new values are typically factors of 2 or more smaller than the observation standard deviations estimated as a function solar zenith angles. Contributors to these small values would be the large number of total column observations combined with their error standard deviations and, to some degree, having well predictable short-term column ozone forecasts. While still seeming rather small, these were applied as is. Using the cost-function for this purpose is valid for uncorrelated observation and background error variances, correct specification of spatial error correlations, and negligible residual biases under all conditions at least between observations and forecasts, which are not strictly true. Nonetheless, the change from the original to the updated set of error variances results in a distribution of the statistically normalized innovations closer to the normal Gaussian probability density function with a reduction in maximum probability density from 0.72 to 0.35. The final probability density function shows more outliers than from a normal distribution (Fig. S6). Unaccounted observation systematic error sources such as may be associated with the treatment for cloud cover and differing viewing zenith angles may be contributing to outliers. The distributions for the four total column datasets give maxima from 0.32 to 0.41 as compared to the overall and normal distribution maxima of 0.35 and 0.40. These results do not imply that the ratios of observation to background error variances are correct. If the observations error standard deviations were potentially smaller by a scaling factor of say 0.8, retaining the same total prescribed error variances of the observation minus forecast differences would imply an increase of the background error standard deviations by a factor of roughly about 1.6 for regions of lower SZAs, to at least 2.0 for SZAs of ~70° and above (for latitudes most often near the southern pole for this period). This would result in background error standard deviation close in size to the OMI-TOMS and OMPS-NM error variances. Interestingly, the overall ratios of observation to background error variances would then be closer to those from the original variances,

While the background error variances would be partially dependent on the observation set being assimilated, they have not been accordingly adjusted in the work when changing the number of sources being assimilated.

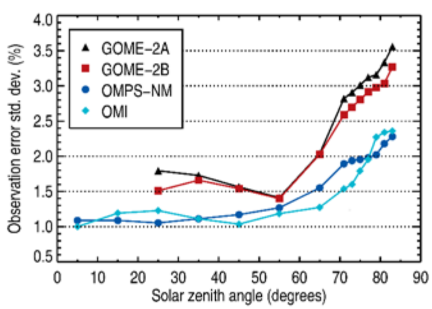


Figure 1. Updated set of applied observation errror standard deviation estimates for GOME-2A, GOME-2B, OMPS-NM and OMI total column measurements as a function of solar zenith angle band centers, for the combined period of July and August 2014. A 2% constant served in specifying the first set of applied error standard deviation estimates.

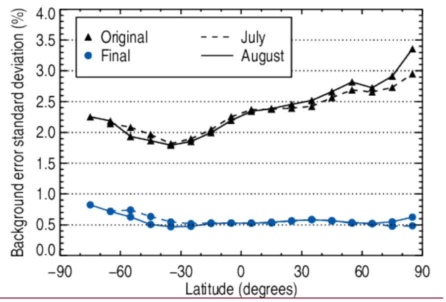


Figure 2. Original and updated (final) set of total column background error standard deviations obtained from projecting the applied latitude and vertically dependent background ozone error variances and the globally homogeneous background vertical error correlations to the total ozone observation space as a function of latitude for July and August.

2.4 Bias estimation and correction

Antón et al., 2009a; Antón and Loyola, 2011; Koukouli et al., 2012; Bak et al., 2015; Bai et al., 2015). The papers by van der A et al (2010 and 2015) indicate negligible variation with viewing zenith angle. For these reasons, and the near-global daily spatial coverage of its measurements, the OMI-TOMS total column ozone product was selected as the anchor in the applied bias correction schemes described in Section 4.

To further examine the acceptability of using OMI-TOMS as a reference for bias correction, a mean differences comparison of OMI-TOMS V8.5 with near-colocated ground-based data at available sites over the periods of study was conducted. The colocation requirements are the same as those specified in Section 4.1 for the inter-comparison of satellite sensors. Summary results are shown in Table 1 and Fig. 1Observation biases can be examined as a function of various factors. In this study, the bias correction applied in the assimilation experiments use bias estimates for discrete SZA/latitude bins as a function of time. Different bias estimation methods based on observation colocations and observation differences with forecasts will be examined. One of these other methods is to use the dependence on the ozone effective temperature instead of latitude and time (e,g., van der A et al., 2010). Solar zenith angle dependence is specifically included considering the varying sensitivities between the different instruments as shown in Koukouli et al. (2012). Latitude and time dependences are also introduced to capture other first order retrieval biases, such as potentially related to the applied a priori atmospheric state and its ozone error covariances, and the specification of the ozone absorption coefficients, as well as instrumental changes in time. While the dependency on other factors such as cloud cover and viewing zenith angle can vary with the instrument and retrieval algorithm, they are not included here as predictors. Their impact would therefore be reflected in the estimated standard deviations derived for observations and outlier errors referred in Section in 2.3. The bias correction target is to reduce residual biases as a function of SZA and latitude relative to OMI-TOMS generally to within 1 %, considering total ozone measurement random error standard deviations being no smaller than ~1 % (Section 2.3). An evaluation of OMI-TOMS total column ozone as the anchor for the periods in this study is provided in Section 3.1 relative to ground-based data, with a supplementary evaluation using SBUV/2 in Section 3.2.1.

Most of the bias estimation results are presented and discussed in Section 3.2. The bias estimates and behaviours obtained in this study do not necessarily reflect the quality of the retrieval data for other periods. Some of the differences that are seen may be attributed to the colocations being approximate (where applicable) or to one or more of the other factors not taken into account here, such as bias as a function of the viewing zenith angle, which is not examined as it was not considered as significant for instruments of interest based on the available literature. Variation as a function of ozone effective temperature (average ozone profile weighted temperature) as was applied for the van der A et al. (2010, 2015) reanalyses is separately examined in Section 3.2.2 as substitution for variations in latitude and time over weeks to months. The dependence of bias as a function of column ozone amount as another possible alternative is potentially subject to small systematic contamination from correlated uncertainties between the column ozone values and the differences (Fig. S7 and the followed related analytical development in the Supplement) and so is not applied for bias estimation.

2.4.1 Colocation approach

As indicated in Section 2.2.1 and in the introduction, the OMI-TOMS column ozone data was selected as anchor in bias estimation and correction. Separate bias estimations are conducted for each distinct instrument-platform pair, with the bias estimates assigned from the mean differences. The estimations can be performed either by direct comparison of observations at colocated locations or through the intermediate use of model forecasts (or analyses). While both are examined, the direct colocation approach without forecasts serves as the baseline. Here, the initial observation differences between colocated observations for use in bias estimation are taken as acceptable for points within 200 km and ±12h, with a latitude difference of no more than 3°, total column differences within 25 DU, and solar zenith angle differences smaller than 5° for SZA under 70° and smaller than 2° for SZA between 70° and 90°. A similar bias correction was applied to the equivalent total ozone of the OMPS-NP partial column profiles for comparison to OMPS-NM. The latitude and solar zenith angle bins have a size of 5° each for total column ozone measurements, and 10° each for partial column ozone profiles, except at larger solar zenith angles where bin sizes are reduced to 2° for 70-90°; larger bin sizes could have been used except likely at high solar zenith angles. In any case, only data with SZA under 84° are used in assimilation considering the larger uncertainties at higher SZA. The smaller bins at high SZA stem from the stronger gradients in the differences between instruments. The larger bin sizes for partial column ozone profiles are in consideration of the smaller density of profile measurements. The resultant bias corrections are assigned to the midpoint of each bin with a two dimensional piecewise linear interpolation applied to points at intermediate SZA and latitude values; data that would require corrections from extrapolation are instead discarded.

To provide preliminary insight on bias estimates, Fig. 3 shows the monthly and hemispheric mean colocation differences with OMI-TOMS for GOME-2 and OMPS-NM as function of SZA with separate curves for the northern and southern hemispheres in August 2014 (see also Figs. S8 and S9). The differences between hemispheres suggest that including latitudinal variation would be more essential in some cases than others. The differences for OMPS-NM are consistent with the average biases of -2 to -4 % with respect to OMI-TOMS and SBUV/2 found by Flynn et al. (2014). The discontinuity in time mean differences with GOME-2A/B appearing at 70° in SZA may be associated to the switch in the wavelength for reflectivity retrieval between lower and higher SZA from 331.3 nm to 360.1 nm (Table 1.13 from Zhand and Kasheta, 2009). As such, no interpolation is applied over the SZA value of 70° for GOME-2A/B. The figure shows a better agreement in monthly mean differences with OMI-TOMS for GOME-2B with magnitudes under 1 % for SZA below 70° and a maximum of ~1.6 % or ~5 DU above 70°. Largest differences are found for GOME-2A and the provisional OMPS-NM data reaching at least ~4 %, or ~10-12 DU, at some SZA values. The larger changes with SZA shown in the Southern Hemisphere for high SZA, e.g. above 80° for GOME-2, may also be found in the Northern Hemisphere depending on the month. In Section 3.1, it is shown that trends and variation over time of the differences within a month can be as significant for some latitude/SZA regions. The correlation of differences in latitude and time with effective ozone temperature is also characterized in Section 3.2.2.

Mean differences for each latitude/SZA bin are generated for individual six-hour intervals with, as a precaution, the removal of outliers beyond two standard deviations about the mean when there are at least 100 points per bin. Instead of monthly mean bias estimation, a moving window using the previous two weeks of data is applied to better capture variations in time. The six-hour mean differences over the two-week moving window are weighted in time with a Gaussian weighting function with a half width at half maximum of 4.7 days. The six-hour mean differences are generated starting two weeks prior to the start of assimilations to provide data over the full window at the start of the assimilation. Another two standard deviation outlier removal is applied, this time according to the variability of the six-hour mean differences over the two-week period. A minimum of 25 total contributing differences originating from at least four six-hour intervals is imposed for valid bias estimates for each bin.

2.4.2 Bias estimation involving differences with forecasts

An alternative bias estimation approach utilizes the differences of the original retrieved observation data with short-term forecasts with the same binning in latitude and solar zenith angle over a two-week moving window. These bias estimates can be obtained by considering the OMI total columns differences with forecasts (i.e. OmF), with or without colocation requirements, or simply without any direct use of OMI. The FGAT short-hour forecasts F would have been influenced by the assimilation of all bias corrected data used in an experiment while the observations O denote retrieved observations prior to bias correction. For each bin, the bias estimate could be obtained from moving time series window of

a) $\langle (O - F) - (O - F)_{ref} \rangle$ with the same colocation requirements as Section 2.4.1,
b) $\langle O - F \rangle - \langle O - F \rangle_{ref}$ without the above colocation requirements, or

c) $\langle O - F \rangle$

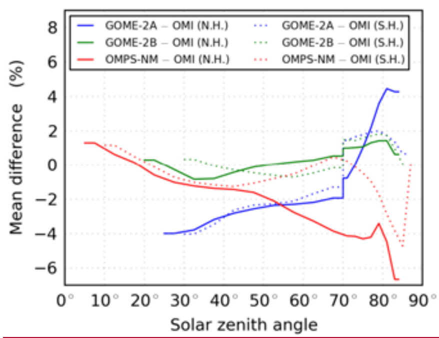


Figure 3. Mean total column ozone differences between GOME-2A/B, OMPS-NM, and colocated OMI-TOMS data as a function of solar zenith angle for the Northern and Southern Hemispheres for August 2014. Differences were computed from observation colocations. For the GOME-2 instruments, separate continuous difference regions are specified about 70° (see text in Section 2.4.1 for additional information). Constant difference extrapolation is applied from bin midpoints at the edges of distinct regions.

with the subscript 'ref' denoting differences for observations of the anchor set, this being OMI here. All three options with short-term forecasts are applied total column ozone measurements for comparison. Just as the colocation approach, these cases require the mean innovation differences being available for the SZA, latitude, and time bins. In this work, options (b) and (c) become successive fallback approaches to (a) in the absence of collocated anchor measurements for a bin, with option (b) automatically reducing to option (c) in the absence of the OMI or anchor data. ~~Option (a) provides the potential benefit of accounting for spatial differences between paired colocation points, while options (b) and (c) bring the potential advantage of bias correction in the absence of sufficiently close colocation pairs. Option (c) provides a bias correction option for times and locations where the reference is not available. For option (c), innovations would be of more benefit when the forecasts more strongly reflect the influence of the anchor data from previous analyses than that of the model and initial condition errors. The implications of having used or not used the anchor data in the assimilation are presented in Section 3.3.~~

~~3 Results~~

~~Assimilation runs covering July-August 2014 have been performed for individual and combined sets of total column ozone satellite data sources with and without bias correction based on the colocation approach. The resulting short-term forecasts from the assimilation runs are compared to each other and to the absence of assimilation.~~

Following an evaluation of OMI-TOMS for the two-month periods in Section 3.1, bias estimates are determined for the other total column datasets using the two general types of differences described in Section 2.4, one based only on colocations and the other on differences with forecasts. The dependence on the ozone effective temperature as done in van der A et al. (2010) is examined separately as a substitution for latitude and time dependence. The bias corrections applied in assimilations are dependent on latitude and solar zenith angle with a two-week moving window. Bias estimation and correction using the observation colocation approach by itself can be done either prior to or as part of assimilation runs, the former being applied here. Mean differences with forecasts would normally be determined and applied for bias estimation during the assimilation and forecasting cycle. For convenience, here we instead use differences for spatially thinned observation sets, thinned to ~1° sampling, with six-hour forecasts of separate assimilation and forecasting sets. The estimated biases for different datasets and from different approaches are presented and discussed first in Section 3.2. Both the July-August and January-February periods are considered for a comparison of bias estimates between seasons within a yearly cycle. The related impact of total ozone assimilations on short-term forecasts is assessed in Section 3.3.

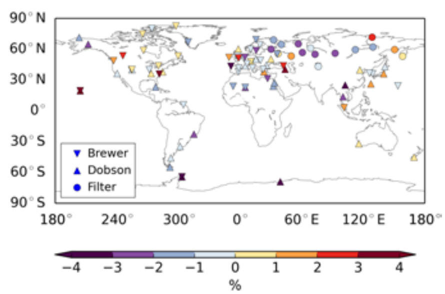


Figure 4. Mean total column ozone differences between OMI-TOMS and Brewer, Dobson, and filter ozonometer measurements over July-August 2014.

3.1 Evaluation of OMI-TOMS total column ozone with ground-based data

A mean differences comparison of OMI-TOMS with near-colocated ground-based data at available sites over the periods of study was conducted to support the selection of OMI-TOMS as anchor justified earlier through references identified in Section 2.2.1. Summary results are shown in Table 1 and Fig. 4 (see also Tables S1 to S3). Bimonthly mean differences over regions, globally, and for the individual stations were produced for the three periods of Table 1 based on totals of 5352 Brewer, 40 Dobson, and 20 filter ozonometer stations.locations; Fig. 14 shows the station locations and mean differences for the July-August 2014 period. The sizes of the global mean differences over the different periods are in the approximate ranges of 0.0 to -0.1 % for Brewer, -0.2 to 0.4 % for Dobson, and -0.8 to -0.7 % for filter ozonometer instruments. The differences for filters

Formatted: Font: +Body (Times New Roman) are however confined to upper northern latitudes. These global and regional averages exclude stations with mean differences larger than two standard deviations of the initial mean differences, corresponding to between 3 and 4 %; this outlier removal process was also applied to each station in determining the mean differences at the ~~% and obtained after removing~~ stations. ~~with mean differences larger than 6 DU.~~. The total number of ~~these~~ outlier stations per time period ranges from 0 to 5 (Tables S1 to S3), ~~with~~ some of which are~~these being~~ stations at high elevation or in Antarctica. While excluded from contributing to the global averages~~The numbers of statistical outliers are small in comparison to the much larger percentages~~ of Table 1,~~sites showing outlier or suspect characteristics over the 5-year period examined by Fioletov et al. (2008). Considering the few 2-3 stations in Antarctica and some consistency between station values,~~ the outlier station mean differences were not~~are~~ included as part of the regional mean differences for 60-90° S in Table 1, and the Antarctic outlier stations were retained for further evaluation later in this section.

The global mean differences, and most regional values, are typically smaller than earlier studies mentioned in the first paragraph. Possible contributors to this might be differences in time periods, region specifications, ground-based observation sets, or colocation conditions.

The regional mean differences are within 1 %,~~% in size,~~ with the exceptions being Antarctica for both Brewer and Dobson instruments and the north polar region for Dobson and filter ozonometers instruments. Table 1 shows small positive biases of less than 0.7 % over the region encompassing Canada, the continental United States, and Greenland, as compared to small negative biases of up to -0.4 % over Europe and Northern Africa. ~~and Greenland, as compared to small negative biases up to -0.4 % over Europe and Northern Africa. The global and most regional mean differences are notably smaller than the roughly 1.5 % underestimation identified by Bak et al. (2015), McPeters et al. (2015), and Labow et al. (2013). These studies have used multiple years in their analyses. McPeters et al. (2013) instead found positive average differences with northern hemisphere Brewers and Dobsons covering 2005 and 2006 of 0.6 % and 0.8 %, respectively, however, with uncertainties of 1.1 % and 1.5 %. The mean differences for both polar regions are all negative suggesting an underestimation of OMI-TOMS column ozone in these regions for these periods.~~ The mean differences for the north polar region of -0.3 to -0.6 % for Brewers are under~~areunder~~ the 1 % target, while the mean differences for Dobsons are -1.2 to -1.6 % and -1.4 to -1.1 % for filter ozonometers. The results for Dobsons and filters are similar despite error levels for the filter instruments being about 1.5 to 2 times larger (Section 2.4.1) and the small datasets. The values over the three seasons are in good agreement despite the small to moderate (≤361) number of colocations. The~~%. These fall within the range of the~~ mean differences covering 2007-2010 from Koukouli et al. (2012) for ~~(2012) at -1.5±2.4 % for Brewers and -0.5±3.0 % for Dobsons. The adjustments of Dobson values applied in this study based on the ozone effective temperatures, which was not included in Koukouli et al. (2012), reduced the sizes of the mean differences in~~ this region have the same sign, with values of -1.5 % for Brewers and -0.5 % for Dobsons. ~~by less than 1 %.~~ The average solar zenith angles for stations in the north polar region, while higher than for the middle latitude region, were less than 70° for all instruments and periods except for some Brewer instruments during the January-February 2015 period reaching at most ~76°. Koukouli et al. (2012) determined standard deviations of the differences of 2.4 and 4.3 % for SZA ranges of 25-70° and above 70°, respectively, indicating an increased variability at higher SZAs. Considering the respective As such, stray light would not significantly impact the total column ozone values (e.g., Moeini et al., 2018; Evans et al., 2009).

More severe underestimations at 3-6 % occur during July-August in Antarctica, this associated to SZAs close to or greater than 80° and possibly a strong latitudinal gradient associated to the winter South Pole polar vortex. While the small size of the dataset for the limited 1-3 stations in this region restricts the statistical significance of these results, the level of consistency between the instruments and sites suggest it being worthwhile to consider this data and so were retained in Table 1 for the rows of the 60-90° S region. The largest adjustments made to the July-August Dobson data occurred for the two Antarctica stations of Marambio (64.23° S, 56.62° W) and Syowa (69.01° S, 39.58° E) due to the low ozone effective temperatures, near 200 K, with increases in mean differences from near zero to -3 to -4 % (see also Koukouli et al, 2016). This brought the differences for Dobsons closer to those for the Brewer at the Marambio station. The differences for the Dobson at Ushuaia (54.85° S, 63.31° W) at the southern tip of Argentina are also negative in the range of -0.8 to -1.6 % over the three periods of this study and of Koukouli et al. (2012), their differences for this region . Other factors may stem from differences in the range of SZAs. The mean differences for both polar regions are all negative indicating an underestimation of OMI-TOMS column ozone in these regions for these periods relative to also affect the ground-based values in this region. Bernhard et al. (2005) noted an underestimation potentially exceeding 2 % for SZAs larger than 80°, reaching 4% in the ozone hole region for a SZA of 85°, that could result from the standard Dobson retrieval method assuming the ozone layer being at a specific heightdata, which is likely. A related, to adjustment would increase the differences with OMI-TOMS. As well, ground-based measurements at very high SZAs.

Formatted: Font: (Default) Times New Roman

Table 1. Regional and global relative mean differences (%) of total column ozone between OMI-TOMS and the specified ground-based instument types over July-August 2014/2015 and January-February 2015. -The averaging excludes stations having outlier station mean diffferences for each period (see Supplement tables S1 to S3 and the text of Section 3.-1) except for the two rows for the latitude region 60-90° S as described in the text. The standard deviations (S.D.) are for the inter-station variation of the station mean differences about the regional or global mean differences. Unavailable S.D. values for available mean differences imply the presence of only one station. The Dobson total column ozone measurements for the two July-August periods were adjusted as a function of the ozone effective temperature (see Section 2.2.4); those for the January-February period were not adjusted in the absence of the ozone effective temperature for the period. The impacts of the Dobson July-August period corrections on the global mean differences were reductions between 0.0 and 0.4 %.

Instrument type	Region	Regional and global mean differences (%) [# of colocations]					
		July-Aug. 2014		July-Aug. 2015		Jan.-Feb. 2015	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Brewer	Latitude range: 60-90° N	-0.3 [258]	0.8	-0.6 [361]	1.0	-0.6 [9]	1.3
	Latitude range: 30-60° N	0.1 [17730 [1816]	1.4	0.6 [13841 [1422]	1.6	0.0 [8651 [886]	1.1
	Latitude range: 30° S - 30° N	0.4 [296]	1.9	-0.5 [165]	0.7	-0.32 [314]	1.3
	Latitude range: 30-60° S	-	-	0.1 [38] -	0.0-	0.2 [55] -	0.0-
	Latitude range: 60-90° S	--5.59 [13]*	-	-	-	-2.5 [152]#	2.0
	North America and Greenland	0.7 [669]	1.1	0.82 [1020]	1.57	0.3 [492495]	1.10
	Europe and Africa	-0.3 [13461282]	1.4	-0.3 [742780]	1.2	-0.5 [4543 [427]	1.1
	East Asia and Other	-0.6 [3121 [419]	1.69	-1.0.8 [186 [148]	0.98	-0.2 [3983 [388]	1.45
	Global	0.1 [23270 [2370]	1.4	-0.31 [1948]	1.5	-0.1 [13441310]	1.2
Dobson	Latitude range: 60-90° N	-1.4 [396 [64]	1.53	-1.2 [29]	0.0	-	-
	Latitude range: 30-60° N	0.3 [3310 [421]	0.87	0.6 [3015 [328]	1.3	0.8 [1677 [187]	1.0
	Latitude range: 30° S - 30° N	--0.3 [2402 [270]	2.47	-1.0 [188.8 [200]	1.34	1.0.1 [120 [149]	1.42.2
	Latitude range: 30-60° S	-0.5 [1504 [167]	0.9	-1.0 [111]	0.4	--0.0 [1363 [171]	1.30
	Latitude range: 60-90° S	-3.3 [6]+	0.1	-4.3 [2]^	-	-0.0 [1024 [134]$	1.7
	North America and Greenland	-0.5 [1256 [167]	0.76	-0.6 [578 [84]	1.10.9	0.3 [531 [73]	0.5
	Europe and Africa	-0.6 [3277 [400]	1.43	0.2 [293]	1.65	0.7 [135]	1.1
	East Asia and Other	0.3 [3147 [361]	1.82.0	-0.6 [2795 [291]	1.23	0.1 [3374 [433]	1.47
	Global	-0.2 [7661 [928]	1.56	-0.2 [629668]	1.4	0.3 [5254 [641]	1.25
filter ozonometer	Latitude range: 60-90° N	-1.4 [47]	0.8	-1.01 [16]	1.0	-	-
	Latitude range: 30-60° N	-0.3 [54]	1.6	-0.5 [62]	2.0	-0.7 [7]	1.2
	Global	-0.8 [101]	1.4	-0.67 [78]	1.8	-0.7 [7]	1.2

Formatted Table

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: Check spelling and grammar, Pattern: 15% (Auto Foreground, White Background)

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

Formatted: French (Canada), Check spelling and grammar

[*] Outlier mean difference from the Marambio station. [#] Includes the Amundsen-Scott, Marambio and outlier Zhongshan stations. [+] Includes Marambio and Syowa stations. [^] Outlier Syowa station only. [$] Amundsen-Scott, Marambio and Syowa stations.

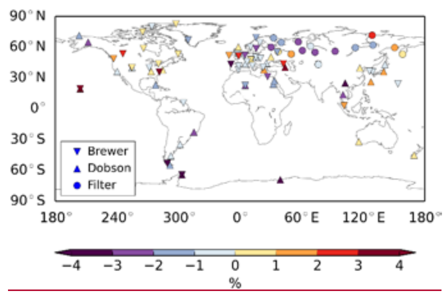


Figure 1. Mean total column ozone differences (%) between OMI-TOMS and Brewer, Dobson, and filter ozonometer measurements over July-August 2014. The colours blue to purple denote negative differences and the colours yellow to red refer to positive differences.

More severe underestimations of OMI-TOMS relative to ground-based observations of 3-6 % occur during July-August in Antarctica, which is associated with SZAs close to or greater than 80° and possibly a strong latitudinal gradient associated to the winter South Pole polar vortex. While the small size of the dataset of 1-3 stations in this region restricts the statistical significance of these results, the level of consistency between the instruments and sites suggest that it is worthwhile to consider this data and so were retained in Table 1 for the rows of the 60-90° S region. The following two paragraphs present reasons that may contribute to either increasing or decreasing the differences in this region.

require an unobscured horizon which may not always be possible throughout the day; often antennas, other buildings, etc., partially block the sun, thus producing low ozone values. Bernhard et al. (2005) noted an underestimation potentially exceeding 2 % for SZAs larger than 80°, reaching 4% in the ozone hole region for a SZA of 85°, that could result from the standard Dobson retrieval method assuming the ozone layer being at a specific height (however, adjusting for this would increase the differences with OMI-TOMS). Another factor that would increase the differences with OMI-TOMS for measurements at high solar zenith angles, especially Dobsons, is stray light (e.g., Moeini et al., 2018; Evans et al., 2009),) which results in an underestimation of the total column ozone up to at least 5-7 %. The stray light sensitivity also depends on the total column ozone itself, with the effect being smaller under ozone hole conditions than over normal conditions. The Brewer measurements in Antarctica are from double-monochromatic instruments and so only slightly sensitive to stray light as compared to the Dobsons.

At high SZAs, in the vicinity of the polar vortex, the horizontal differences in location between the station and the average of the observed ozone would be sensitive to the strong horizontal gradients in total column ozone. Small differences in observed locations, as well as small differences in solar zenith angles of the colocation pairs at high SZAs, can imply notable differences of observed air masses. For example, approximatelyApproximately accounting for a latitudinal displacement of slightly more than 1° resulted, for example, in reducing the July-August 2014 mean difference from the Brewer at Marambio from the -5.9 % in Table 1 to -2.7 %. -2.7 %. As another point of consideration, Figure 3 of Balis et al. (2007b) did not show significant dependence on solar zenith angles for OMI-TOMS relative to Brewers for SZA larger than 20° except above 80° with an overestimation of about 4±2 % instead of an underestimation. This difference though could be related to differences in time periods, stations, and dataset sizes. The discussion offor the polar regions and high solar zenith angles is extended in Section 4section 3.2.1 with a comparison to SBUV/2 total column ozone data and a note regarding the assimilation with Aura-MLS in section 3.3.1.

While the OMI-TOMS data could be underestimatingunderestimates of total column ozone in the polar regions for these periods, there is some uncertainty as to the actual OMI-TOMS bias considering factors that could affect the reliability of the comparison with the ground-based data at high solar zenith angles for Antarctica, this even beyond the low number of ground-based observations. van der A et al. (2015) included an adjustment to OMI-TOMS total column ozone data based on the ozone effective temperature in addition to a constant offset of 3.3 DU, which usedusing a comparison to Brewer and adjusted Dobson data. This, which would increase the OMI-TOMS total column ozone in Antarctica by about +10.5 DU for the two July-August periods. Including: including also the second-order dependence on SZA of that paper would reduce thisreduces the change to +9 DU. This adjustmentIts application would have improved the agreement in the 60-90° S latitude band of Table 1, while adding to the mean differences in the other regions by less than ~1 %,% except possibly in the January-February- 60-90° N region.

Excluding the uncertainty in quantifying corrections in the south polar region, the low**3.2 Bias estimation**

3.2.1 Colocation approach

Time mean differences of the OMI-TOMS V8.5 data with the ground-based data for most regions supports not having to adjust the data before serving as anchor in the bias estimationthe various instruments, GOME-2A, GOME-2B, OMPS-NM, OMPS-NP, and correction of the other satellite sources SBUV/2, with OMI-TOMS for the limited period covered in this study. While not done here, a correction specifically at high SZAs based on differences for the north polar and the 30-60°S regions could be envisaged.

4 Bias estimation and evaluation using OMI-TOMS as reference

Observation biases can be examined as a function of various factors. In this study, the bias correction applied in the assimilation experiments used bias estimates for discrete SZA/latitude bins as a function of time. Different bias estimation methods based on observation colocations and observation differences with forecasts will be examined. Solar zenith angle dependence is specifically included considering the varying sensitivities between the different instruments as shown in Koukouli et al. (2012). Latitude and time dependences were introduced to capture other data processing biases as well as instrumental changes over time. The alternative method of using the dependence on the ozone effective temperature instead of latitude and time (e.g., van der A et al., 2010) was also explored. Any bias impact due to differences in spatial resolutions of the instruments or model forecasts would be part of the residual biases and associated representativeness errors. Part of the effect of differences in resolution between instruments would be mitigated from bias estimation relying on local averages of differences in space in addition to time. While the dependency on other factors such as cloud cover and viewing zenith angle can vary with the instrument and retrieval algorithm, they are not included here as predictors. Their impact would then be reflected in the estimated standard deviations derived for observations. The bias correction target is to reduce residual biases as a function of SZA and latitude relative to OMI-TOMS generally to within 1 %.

Both July-August ~~2014 (Fig, 5)~~ and January-February periods are considered for a comparison of bias estimates between seasons within a yearly cycle.~~2015 (Fig, 6) indicate global averaged biases in the range of -3.5 to 2 %.~~ The two sets of ~~results with SBUV/2 are for the~~ SBUV/2 total column ozone values obtained from the two wavelengths retrieval (SBUV/2-TC) and the sum of the retrieved partial column profiles (SBUV/2-NP) are included in the comparisons to OMI-TOMS.~~),~~ These have been added to extend the evaluation of the OMI-TOMS data conducted in Section~~section~~ 3.

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

4.~~,~~1 Colocation approach

Formatted: Font: +Headings (Times New Roman), English (United States)

This method estimates the bias as the mean differences of colocated observations with OMI-TOMS. Separate bias estimations are conducted for each distinct instrument-platform. Here, the criteria for observations to be considered to be colocated are for the points to be within 200 km and ±12h, and have solar zenith angle differences smaller than 5° for SZA under 70° and smaller than 2° for SZA between 70° and 90°.~~,~~ The latitude and solar zenith angle bins have a size of 5° each for total column ozone measurements, and 10° each for summed partial column ozone profiles, except for solar zenith angles above 70°, where bin sizes are reduced to 2°.~~averaging over all bins In any case, only data with SZA under 84° are used in assimilation considering the larger uncertainties at higher SZA.~~ The smaller bins at high SZA were chosen since stronger gradients in the differences between instruments arise for these values. The larger bin sizes for summed partial column ozone profiles are in consideration of the smaller density of profile measurements. The resultant bias corrections are assigned to the midpoint of each bin with a two dimensional piecewise linear interpolation applied to points at intermediate SZA and latitude values; data that would require corrections from extrapolation are instead discarded.

Formatted: Font: +Body (Times New Roman), English (United States)

Formatted: Font: +Body (Times New Roman)

Mean differences for each latitude/SZA bin are generated for individual six-hour intervals with, as a precaution, the removal of outliers beyond two standard deviations about the initial mean when there are at least 100 points per bin. Instead of monthly mean bias estimation, a moving window using the previous two weeks of data was applied to better capture variations in time. The six-hour mean differences over the two-week moving window were weighted in time with a Gaussian weighting function with a half width at half maximum of 4.7 days. The six-hour mean differences were generated starting two weeks prior to the start of assimilations to provide data over the full window at the start of the assimilation. Another two standard deviation outlier removal was applied, this time according to the variability of the six-hour mean differences over the two-week period. A minimum of 25 total contributing differences originating from at least four six-hour intervals is imposed for valid bias estimates for each bin.

The time mean differences with OMI-TOMS for July-August 2014 (and January-February 2015 are shown in Figs. 2 and 3, respectively. The figures indicate global averaged biases in the range of -3.5 to 2 % (Table 2). The maximum time mean biases per bin reach to sizes of ~5-9 % for some datasets. These mean differences are in general larger than the~~), with weighting by the number of colocations,per bin, gives mean biases of -1.8 % (-3.5 %) for GOME-2A, 0.05 % (-0.5 %) for GOME-2B, -1.3 % (0.1 %) for OMPS-NM, 1.1% (2.0 %) for OMPS-NP, 1.5 % (1.3 %) for SBUV/2-TC, and 1.2 % (0.6 %) for SBUV/2-NP. The~~ mean differences of OMI-TOMS with ground-based data. The mean differences typically vary by roughly 3 % over the ranges of bins for SZA values lower than 70°, while larger variations of up to ~7 % can be seen at higher SZA values. The mean differences from SBUV/2 typically vary less between bins as compared to the other instruments. GOME-2A and GOME-2B give the largest and smallest mean differences globally, respectively. The standard errors of the mean differences shown in Figs. 2 and 3 are below 0.1 % for most bins, except for some bins at high solar zenith angles (above 70 °), due to the smaller number colocations, where the maximum standard errors found over all datasets is 0.6 %.

The discontinuity appearing at 70° in SZA for both GOME-2 instruments, as seen in Figs. 2 and 3, may be associated with the switch in the wavelength for reflectivity retrieval between lower and higher SZA from 331.3 nm to 360.1 nm (Table 1.13 from Zhand and Kasheta, 2009). As such, when bias corrections were applied for GOME-2, no interpolation was applied over the SZA value of 70°. For the DOAS retrieval products, Hao et al. (2014) showed mean differences with Norther Hemisphere ground-based data that varied seasonally between roughly zero and 4 % over the period 2007 to Summer 2013. For the TOMS retrieval products used in this study, the seasonal variation can be seen from Figs. 2 and 3 and Tables 2 and 3, with larger differences for GOME-2A of up to about 3 %. Hao et al. (2014) also showed differences between GOME-2A and GOME-2B of less than 1 % covering December 2012 to November 2013, except in the south pole region and in the Southern Hemisphere for May to September where it reaches at least 2 %. This differs for the TOMS-based GOME-2 retrieval products used here that typically showed larger differences between the two instruments for the times studied.

The pattern ~~(of opposite sign to the mean differences of the satellite data with OMI-TOMS) over the various sites are usually smaller in size than the average mean differences over all bins between the various satellite-borne instruments and OMI-TOMS. The mirroring patterns~~ about the equator in Fig. 3 (Jan-Feb) appears inverted as compared to Fig. 2 (July-August)~~Figs. 5 and 6~~ for ~~at least~~ SBUV/2 and OMPS-NP (which can also be seen in Table 3) suggests~~suggest~~ the possibility of some seasonally dependent differences with OMI-TOMS for these instruments~~data~~. The results for the provisional OMPS-NM data are smaller than the roughly 2.5 % determined at the Tsukula station for the more recent product version (Bai et al., 2016). However, this is only for a single station. The overall variations in longitude of the mean differences with OMI-TOMS are notably weaker than that in latitude. As such, one would expect the remaining spatially varying residual biases to be small. The percentage of non-empty bins with time mean differences exceeding 2 % in magnitude for the six datasets range from 0 % for SBUV/2-NP

Formatted: English (United Kingdom)

Formatted: English (United Kingdom)

to 69 % for GOME-2A (Table 2), both for Jan-Feb 2015. For a 1% threshold, these percentages increase by factors of 1.2 to 4, depending on the instrument and season.~~2016; Section 2.2.3), but then this is a~~

Formatted: Font: +Body (Times New Roman), 10 pt, Not Bold, English (United States), Check spelling and grammar

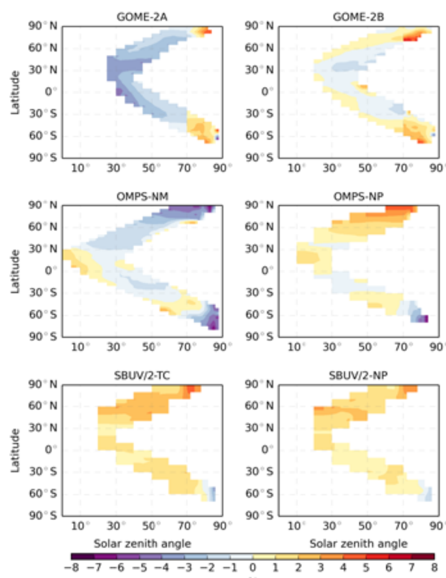
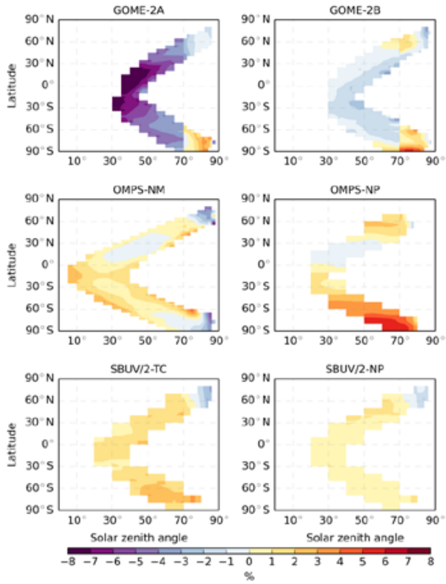


Figure 2~~5~~. Mean total column ozone differences (%) between GOME-2A/B, OMPS-NM/NP, SBUV/2-TC/NP and colocated OMI-TOMS data for the period of July-August 2014. The SBUV/2-TC total column ozone values stem from the two wavelength retrieval, while those for -SBUV/2-NP are the sums of the retrieved 21-layer partial columns. The colours blue to purple denote negative differences and the colours yellow to red refer to positive differences.

Formatted: Left

Formatted: Font: Calibri, 11 pt, Check spelling and grammar


GOME-2A
GOME-2B
OMPS-NM
OMPS-NP
SBUV/2-TC
SBUV/2-NP
Latitude
Solar zenith angle
%

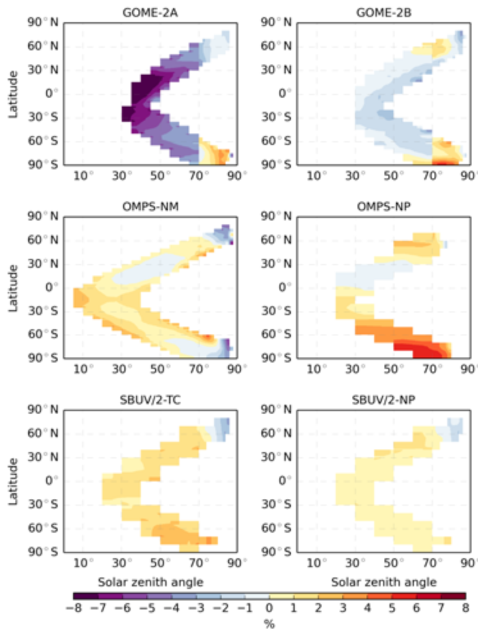


Figure 36. Same as Fig. 25 for January-February 2015.

Table 2. Global diagnosticssingle station. As aside, considering the gradients over the images of differences in total column ozone Figs. 5 and 6, the bin sizes for bias estimation could be increased by factors of at least two to four for most regions, this being adaptable to each instrument.

The standard deviations of the mean differences over the bins (with removal of 2-σ outliers) range between satellite instuments0.5 and 2.4 % with values per dataset of 1.4 % (2.4%) for GOME-2A, 0.8 % (1.3 %) for GOME-2B, 1.3 % (0.9 %) for OMPS-NM, 1.3 % (2.0 %) for OMPS-NP, 0.7 % (0.9 %) for SBUV/2-TC, and 0.7 % (0.5 %) for SBUV/2-NP. The latter two sets of values reflect a smaller variability of the differences between SBUV/2 and OMI-TOMS. As well, the time mean biases per bin reach sizes of ~5-9 % for some datasets.

The differences between both SBUV/2 sets and OMI-TOMS are generally consistent with, and more on the lower end of, the about 1-2 % mean differences found in other studies mentioned in the introduction, with better agreement obtained here for SBUV/2-NP. They also suggest better agreement of OMI-TOMS with mean tropical and mid-latitude northern hemisphere ground-based data for these periods, considering Table 1. The winter polar mean differences between SBUV/2 and OMI-

Formatted: Font: Times New Roman, 9 pt, Check spelling and grammar

Formatted: Justified

Formatted: Font: Calibri, 11 pt, English (Canada)

Formatted: Left, Line spacing: single

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 9 pt, English (Canada)

TOMS remain positive, reaching a maximum of ~4 % while the summer polar means for SZA<~75-80° are less than ~1 %. This seems to support the possibility of OMI-TOMS underestimation near or in the polar regions for SZA<~75-80°, but with less underestimation for the 60-90° S region than is suggested in Table. The disagreement of SBUV/2 and OMI-TOMS in the winter pole regions at even higher SZAs reaches only -1.9 % for SBUV/2-NP, while reaching -3,2 % for SBUV/2-TC, with SBUV/2-NP expected to be more accurate than SBUV/2-TC at higher SZAs. However, identifying which dataset is notably closer to the truth between particularly SBUV/2-NP and OMI-TOMS in the polar regions is somewhat obscured by the differences found in the mid-latitudes and the tropics between SBUV/2-NP, OMI-TOMS and the ground-based data. Considering results for the various datasets, during these periods only, and the results of Table 1, OMI-TOMS serving as the anchor is appropriate, with the caveat that the best reference for the polar regions and SZA>70-80° (and or possibly ozone effective temperatures below ~200 K) may be unclear.

GOME-2B and SBUV/2-NP provide smallest differences globally and both GOME-2 instruments show strong latitudinal gradients in bias estimates for SZA>84°, which are not used in assimilation. The solar zenith angle regions below and above 70° for GOME-2A and GOME-2B give rise to an approximately bimodal distributions of differences with OMI-TOMS for July-August 2014 (and January-February 2015. The diagnostics consists of global mean differences and) centered at about -2.3 % (-4.5 %) and 1.8 % (-0.5 %) for GOME-2A and -0.5 % (-1.0 %) and 2 % (2 %) for GOME-2B. The percentages of non-empty SZA/latitude bins with time mean differences exceeding 2 % in magnitude for the six data sets are 50 % (69 %), 14 % (13 %), 28 % (19 %), 23 % (47 %), 30 % (22 %), 16 % (0 %); for a 1% threshold, percentages increase by factors 1.2 to 4 depending on the instrument and season. Mean differences therefore exceed the observation random error levels for a notable fraction of the datasets.

Instrument	Mean difference (%)		Percentage of bins with |mean differnces| > 2 %.	
	July-Aug. 2014	Jan.-Feb. 2015	July-Aug. 2014	Jan.-Feb. 2015
GOME-2A	-1.8	-3.5	50	69
GOME-2B	0.1	-0.5	14	13
OMPS-NM	-1.3	0.1	28	19
OMPS-NP	1.1	2.0	23	47
SBUV/2-TC	1.5	1.3	30	22
SBUV/2-NP	1.2	0.6	16	0

Table 3. Mean differences of the total column ozone (%) between satellite instuments and OMI-TOMS for July-August 2014 and January-February 2015 for Northern and Southerm Hemispheres, for solar zenith angles below and above 70°.

Instrument	July-Aug 2014				Jan-Feb 2015			
	SZA < 70°		SZA > 70°		SZA < 70°		SZA > 70°	
	NH	SH	NH	SH	NH	SH	NH	SH
GOME-2A	-2.3	-1.8	0.3	1.7	-5.1	-4.5	-1.1	0.9

Formatted: Space After: 10 pt, Line spacing: Multiple 1.15 li

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 9 pt, English (Canada)

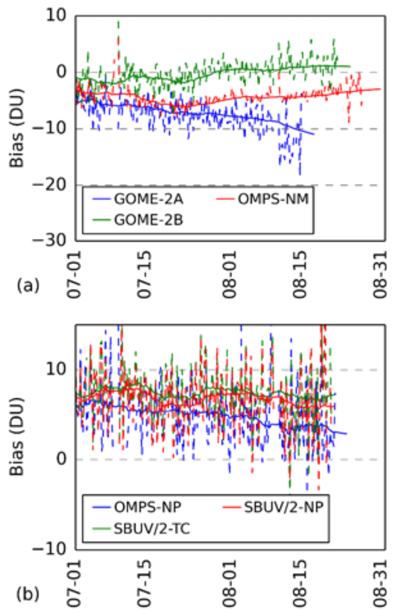
GOME-2B	-0.1	-0.3	1.3	1.6	-0.7	-1.1	0.4	1.7
OMPS-NM	-1.6	-0.6	-4.1	-1.1	-0.1	0.6	-0.6	-0.6
OMPS-NP	1.5	0.1	3.8	-1.1	0.3	3.1	1.6	4.5
SBUV/2-TC	1.8	1.2	4.1	0.3	1.4	1.6	-0.5	2.8
SBUV/2-NP	1.5	0.7	3.6	0.2	0.8	0.6	-0.5	0.7

The SBUV/2-NP dataset could have been an alternative candidate as the anchor considering the temporal stability in the quality of the data and its level of agreement with ground-based data indicated in earlier studies. The comparisons of the SBUV/2 products with OMI-TOMS in Figs. 2 and 3 and Tables 2 and 3 suggest that OMI-TOMS may be generally closer to the ground-based data for these two periods (Table 1). OMI-TOMS also appears to be in better agreement with SBUV/2 in the Antarctic region than with the ground-based data. The agreement between OMI-TOMS and SBUV/2-NP was usually found to be slightly better than the agreement between OMI-TOMS and SBUV-TC, with the agreement being more notably better in the Jan-Feb 2015 Antarctic region.

The variations in time of the bias corrections for a selected single bin in shown in Fig. 4 for the July-August 2014 period. This figure displays the bin with the latitude, solar zenith angle centered on (52.5°N, 37.5°) for instruments with 5° wide bins and the ~~bias correction per~~ bin centered on (55°N, 35°) for instruments with 10° wide bins. The time variations for many bins~~over the two months~~ are most often within ±1 % from the time mean, but some bins can vary by ~3 % in time. The~~with a few bins reaching variations of ±2-3 % (Figs. S10 to S13 and Figs. 7 and 10). These~~ variations in time for different instruments can differ not only in size but also in tendency within the short 1-2 month periods. While the resulting moving averages usually change gradually in time, the random variation~~, or scatter,~~ of the individual six-hour means about the moving averages can be small (~~at~~ within ~1 %)~~%~~ to more significant (reaching at least ~3 %) as can been seen in ~~% (~~Fig. 4. The~~7 and Fig. S14); the~~ number of colocations per bin for each six-hour interval ranges from a few to a few hundred, while the number of colocations~~points to a few hundred for the six-hour intervals, with larger scatter also present for large numbers per bin. A verified cause of the larger sized six-hour mean differences, as seen in Fig. 7, is the occasional temporal data gaps of OMI resulting in biasing 6-hr mean differences due to a persistent time difference between colocation pair elements given the assigned ±12h colocation search window. The influence of this added bias of individual 6-hr mean differences is circumvented through outlier removal in the weighted averaging over the two-week moving window. Unaccounted biases such as from the differing handling of cloud cover are also likely contributing to the varying sizes of 6-hr mean differences about the moving averages. The number of colocated total column points applied in bias estimation~~ for each ~~bin over the~~ two-week moving window typically exceeds a thousand but can be below one hundred for a~~some~~ few bins~~ (e.g., Fig.~~. As, the~~S14). The~~ number of colocations are significantly reduced for OMPS-NP and SBUV/2 measurements, as it would be for other profilers, ~~and so~~ the averaging might~~, while not done here, could~~ benefit from longer time windows, a wider Gaussian filter, ~~and~~ or larger bin sizes.

Formatted: English (Canada)

Formatted: English (United Kingdom)



~~While capturing much of the differences with OMI through latitude, time, and SZA dependent corrections, one would expect remaining spatially varying residual biases (and likely temporally varying as well). The overall variations in longitude of differences with OMI-TOMS are notably weaker than those in latitude, associated to the usually larger latitudinal gradients in ozone and the solar zenith angle variation in latitude along the orbit track, with occasional longitudinally isolated areas of larger differences (e.g., such as near (10° S, 0° E) for GOME-2A/B, about (60° S, 60° W) for OMPS-NM and over Greenland for OMPS-NP as shown in Fig. S15).~~

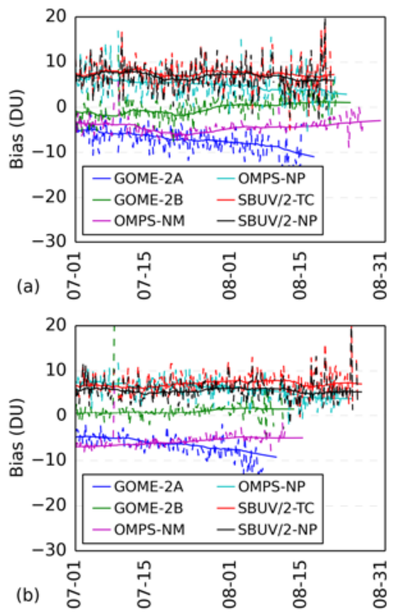
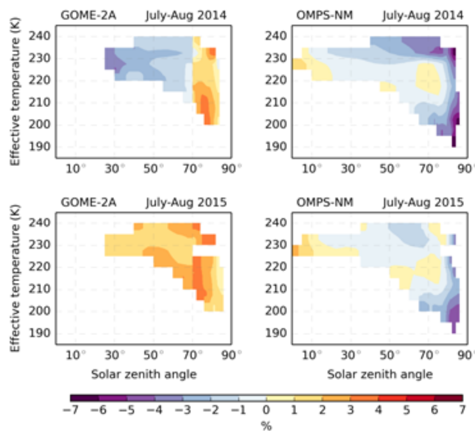


Figure 47. Time series of total column ozone bias corrections (DU) for July and August 2014 for GOME-2A/B, OMPS-NM/NP, and SBUV/2-TC/NP as derived from the colocation method described in Section 4.1.(see Fig. 5). Dashed vertical lines show individual six-hour mean differences with OMI-TOMS, while the solid curves of the same colour show the two--week moving average bias corrections. The particular (latitude, solar zenith angle) bins plotted are (a) 5° wide bins centred on (52.5°N, 37.5°) for GOME-2A/B and OMPS-NM and a 10° wide bin centred on (55°N, 35°) for OMPS-NP and SBUV/2-TC/(b) 5° wide bins centred on (62.5°N, 42.5°) for GOME-2A/B and OMPS-NM and a 10° wide bin centred on (65°N, 45°) for OMPS-NP. Time coverage for individual bins- do not necessarily cover complete months.

Formatted: Font: 10 pt, English (United States), Check spelling and grammar

Formatted: Space After: 0 pt, Line spacing: 1.5 lines



4.2 Bias estimation involving

~~**Figure 8. Mean total column ozone**~~ **differences** ~~**with forecasts**~~

~~An alternative bias estimation approach utilizes the differences of the original retrieved observation~~ between GOME-2A, OMPS-NM and colocated OMI-TOMS data with short-term model forecasts, with the same binning in latitude~~as a function of ozone effective temperature~~ and solar zenith angle over a two-week moving window. This would be applicable for near-real time or reanalysis data assimilations. These bias estimates can be constructed by considering observation (*O*) differences with forecasts (*F*). Bias estimates can be obtained by taking the differences~~the periods~~ of *O-F* (innovations in an assimilation context) between an instrument and a reference (*O* denotes retrieved observations prior to bias correction), which may be done with or without colocation requirements. We identify three different options for this case:

d) $\langle (O-F)-(O-F)_{\mathrm{ref}} \rangle$ with the same colocation requirements as Section 4.1,
~~a)~~ $\langle O-F \rangle - \langle O-F \rangle_{\mathrm{ref}}$ ~~without the above colocation requirements, or~~
e) $\langle O-F \rangle$

where the angular brackets denoting averages and the subscript 'ref' denoting differences for observations of the anchor set (OMI-TOMS for our case). Option (a) provides the potential benefit of accounting for spatial differences between paired colocation points, while options (b) and (c) bring the potential advantage of bias correction in the absence of sufficiently close colocation pairs. If previous observations of the reference or other bias corrected instruments were assimilated into the system that produce the short-term forecasts *F*, then option (c) provides a bias correction method for times or locations where the reference is not available. In this work, options (b) and (c) become successive fallback approaches to (a) in the absence of colocated anchor measurements for a bin, with option (b) automatically reducing to option (c) in the absence of the OMI or

Formatted: Font: +Headings (Times New Roman), 10 pt, English (United States), Check spelling and grammar

Formatted: Font: +Body (Times New Roman), 10 pt, English (United States), Check spelling and grammar

Formatted: Font: +Body (Times New Roman), 10 pt, English (United States), Check spelling and grammar

Formatted: Font: +Body (Times New Roman), 10 pt, English (United States), Check spelling and grammar

Formatted: Font: +Body (Times New Roman), 10 pt, English (United States), Check spelling and grammar anchor data. For option (c), innovations would be of more benefit when the forecasts more strongly reflect the influence of the anchor data from previous analyses than that of the model and initial condition errors. In addition, a cutoff criterion for the use of option (c) can be imposed by requiring reference data to have been assimilated within a certain past time period to ensure that these data sets have adequate influence over the forecasts. The same binning and time averaging as done in Section 4.1 are used in this section. As options (b) and (c) are able to use more data than option (a), the extension of (a) to use (b) and (c) as successive fallbacks can increase the number of usable bins in the bias estimation, which would be more evident at high SZAs.

All three of the above options for total column ozone bias estimation were performed and compared to the estimates from Section 4.1. Mean differences with forecasts would normally be determined and applied for bias estimation during the assimilation and forecasting cycle. For convenience, here we instead used the differences with six-hour forecasts from a separate assimilation and forecasting run (the 'OMI' assimilation run summarized in Table 5), which is described in more detail in Section 5. In practice, the forecasts used for this approach, if applied in a near-real time setting, would come from runs that assimilate the bias-corrected observations using the correction method considered in this section. In this section, all observational data sets used for bias estimation are thinned to 1°.

Bias estimates using the options (a) to (c) above for July-August 2014 are shown in Fig. 5, which also shows the colocations only method of Section 4.1 for comparison, and are summarized in Table 4. Differences between the biases resulting from options (a) to (c) and colocation alone are within 1 % over the two-month period except for a few bins, which are mostly at high SZA. The standard errors of the mean differences for all cases are mostly less than 0.1 %, but can as high as 1 % for the options (a) to (c) cases at very high SZA for bins with little data. The time evolution of these bias estimates from the two-week moving window for two different bins is shown in Fig. 6. All bias estimates (both those that do and do not use forecast differences) follow the same general evolution in time, varying within 1 % of one another. The top and bottom panels of Fig. 6 show examples of a bins that have a larger and smaller evolution in time, respectively, where for these bins the bias estimates change by ~10 DU and ~2-3 DU (~3 % and ~1 % for a total column of 300 DU), respectively.

The bias estimates that use differences with forecasts are largely consistent with estimates that use colocation alone. The estimates that utilize differences with forecast can provide additional benefits over using colocations alone if the forecasts well represent the spatial variation in total column ozone for options (a) and (b), or if the forecasts have been sufficiently de-biased for option (c).

Formatted: Font: +Body (Times New Roman)

Formatted: Font: 10 pt, English (United States), Check spelling and grammar

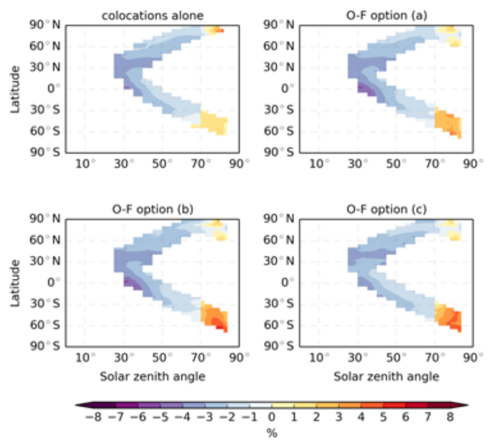


Figure 5. Time mean total column ozone biases (%) between GOME-2A and OMI-TOMS for ~~and~~ July-August 2014 from colocation alone and for the options (a), (b), and (c) of Section 4.2 that use observation-minus-forecast differences. For options (a), (b), and (c), the forecasts were taken from the 'OMI' assimilation run (see Table 5~~2015.~~). The bias in the 'colocations alone' panel was computed using the thinned observation data set to compare to the other cases that use thinned observations. The colours blue to purple denote negative differences and the colours yellow to red refer to positive differences.

Table 4. Mean differences in total column ozone (%) between satellite instuments and OMI-TOMS for July-August 2014 using the options (a), (b), and (c) from Section 4.2, for Northern and Southerm Hemispheres and solar zenith angle below and above 70°.

Instrument	Colocation alone				O-F option (a)				O-f option (b)				O-F opton (c)			
	SZA<70°		SZA>70°		SZA<70°		SZA>70°		SZA<70°		SZA>70°		SZA<70°		SZA > 70°	
	NH	SH	NH	SH	NH	SH	NH	SH	NH	SH	NH	SH	NH	SH	NH	SH
GOME-2A	-2.3	-1.8	0.4	1.7	-7.4	-1.8	-0.1	2.5	-2.6	-1.7	0.0	3.5	-2.3	-1.8	-0.2	3.3
GOME-2B	-0.1	-0.3	1.3	1.6	-0.2	-0.3	1.1	1.5	-0.3	-0.3	1.2	1.9	-0.1	-0.3	1.0	1.8
OMPS-NM	-1.6	-0.6	-4.9	-1.1	1.4	-0.5	-4.7	-1.2	-1.3	-0.4	-4.6	-1.7	-1.3	-0.5	-4.8	-1.9

Formatted: Space After: 0 pt, Line spacing: single

Formatted: Font: 9 pt, English (United Kingdom)

Formatted: Line spacing: single

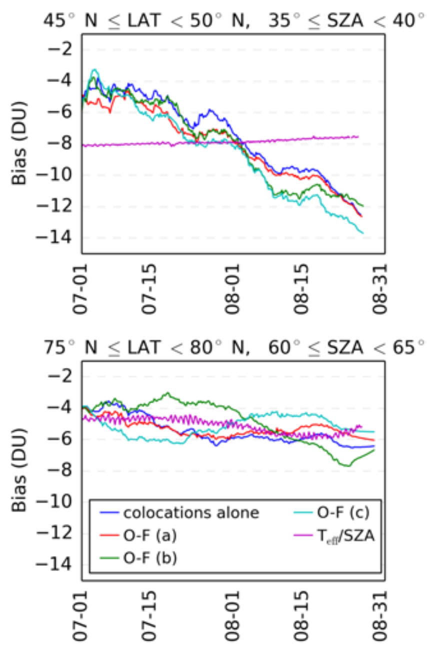


Figure 6. Time series of total column ozone bias corrections (DU) for two latitude/SZA bins covering July-August 2014 for GOME-2A using different bias correction methods. All cases that include colocation methods use thinned observation sets. The 'O-F' curves additionally use the differences of forecasts described in Section 4.2 following the assimilation of OMI-TOMS.~~3.2.2~~ The 'colocations alone' and 'O-F' curves were calculuated using the Gaussian two-week moving average with HWHM of 4.7 days. The 'T_{eff}/SZA' curves, described in Section 4.3, result from mapping each observation that falls within the latitude/SZA bin onto the ozone effective temperature/SZA bias estimate for July-August 2014 (shown in Fig. 7), followed by taking the average of these bias estimate values for each time.

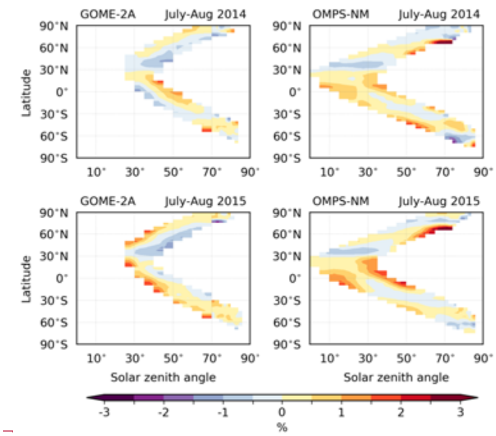
4.3 Variation with ozone effective temperature

An alternative parameterization for the bias estimation consists of using~~in the use of time averaged differences as a function of~~ ozone effective temperature and solar zenith angle, as done in van der A et al. (2010).~~.~~ A motivation for a dependency on ozone effective temperature is to~~ more directly~~ compensate for any unaccounted temperature sensitivity of the ozone absorption coefficients used in retrievals. In this case~~Here~~, bias estimation is~~ made~~ implicitly dependant on time through temporal changes of the ozone effective temperature (and solar zenith angle). This captures at least the seasonal variations of biases associated with~~to~~ changes in temperature in addition to constant offsets. In this section, we briefly consider such a parameterization.

OzoneThe ozone effective temperatures were calculated from ECCC'sthe GEM meteorological model, with LINOZ weather and ozone short-term ozone forecasts driven byrelying on the LINOZ model and launchedrespective weather analyses from ECCC and ozone analyses from the assimilation of total column ozone, all of which are described in more detail in Section 5. For these estimates, we return to the methods of Section 4.1, in which . The resulting time mean differences with OMI-TOMS are computed using only colocated observations (i.e. no use of forecasts).

Bias estimates for GOME-2A and OMPS-NM for July-August 2014 and 2015 using anfrom OMI as a function of ozone effective temperature parameterizationand solar zenith angle can be seen in Fig. 7. By comparing the bias estimates for the same months from notably differ for different years, we see that these bias estimates can differ notably for different time periods. With this parameterization, the bias estimate for GOME-2A differs by roughly 3-4 % , depending on the instrument and data processing setup, as shown for GOME-2A and OMPS-NM in Fig. 8. For GOME-2A (TOMS), typical differences between 2014 and 2015 of the order of roughly 3-4 % for SZAs less than 70°. -These differences are larger than the effect of gradual long term trends of about -2.2 DU, or roughly -0.6 to -0.8 %, per year for GOME-2A (DOAS) estimated by van der A et al. (2010) for GOME-2A (DOAS), although we note that all GOME-2 data used in this study were retrieved using the TOMS method.). Differences in retrievals methods and time periods might be factorsa factor in explaining these differences., if not also the differing time periods. For both time July-August 2014 and 2015 periods shown in Fig. 7, applying their respective corrections as a function of ozone effective temperature and solar zenith angles result in time averaged residual biases as a function of latitude and solar zenith angle typically within 1 %, with only a few bins over 2 %.% (Fig. 9). This supports use of ozone effective temperature dependence as alternative to latitude (and time) dependence, as in other studies, with the stipulation that one accounts for otherwise remaining and notable temporal changes in some fashion where necessary. Adding an explicit dependency on time may be compensating for some otherwise unaccounted time varying biases other than long-term trends. If the temporal changes may be spatially dependent, it is possible that an explicit spatial predictor (such as latitude) may be required. The variation of the bias corrections as a function of ozone effective temperature and SZA within these periods, of up to two months, is comparatively small relative to larger changes in time that can be seen from the two-week moving window bias estimations as a function of latitude and SZA as shown in Fig. 10 for GOME-2A and is also observed for GOME-2B and OMPS-NP as well. The identified differences in variation over these relatively short periods are typically within ±1 %.

An equivalent time evolution of a latitude/SZA bin can be made from the time-averaged effective temperature/SZA bias estimate shown in Fig. 7: First, the ozone effective temperature of each observation falling within a selected latitude/SZA bin is used to map that observation onto the ozone effective temperature/SZA bias



estimate (Fig. 7), then the bias estimate at at each observed ozone effective temperature/SZA point is averaged for each six-hour time period. The resulting curves are shown Fig. 6 for the latitude/SZA bins selected. The small temporal evolutions of these curves (typically well within 1 %) reflects the slight changes in the ozone effective temperature/latitude relationship in time. The greater the variation in time of the bias estimates based on the time varying latitude/SZA parameterization, the larger the differences with the estimates based on the time independent temperature and SZA parameterization (an example of which is illustrated by comparing the top and bottom panels of Fig. 6). Adding an explicit sub-seasonal to seasonal dependency on time to the ozone effective temperature/SZA bias estimate would compensate for these otherwise unaccounted for time variations. Overall, this supports the use of an ozone effective temperature parameterization as an alternative to latitude (and time) parameterization, with the stipulation that one accounts for any remaining notable temporal changes in some fashion when necessary.

Formatted: English (United Kingdom)

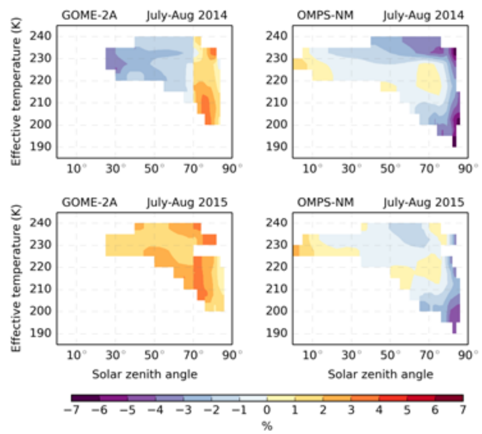


Figure 7. Mean~~**9.** Residual average~~ total column ozone differences (%) between GOME-2A, OMPS-NM and colocated OMI-TOMS data as a function of ozone effective temperature (degrees Kelvin)~~latitude~~ and solar zenith angle (degrees) for the periods of July-August 2014 and July-August 2015. The colours blue to purple denote negative differences and the colours yellow to red refer to positive differences.

5 Assimilation system and results

In this section, we examine the effects of ~~following~~ bias correction on global ozone assimilation and compare the six-hour forecasts launched from these analyses to ground-based observations and to OMI-TOMS. Corrections of observation biases were updated every six-hours using a two-week moving window from colocations with OMI-TOMS. Assimilation experiments were conducted for July-August 2014, with a start date of 28 June 2014, 18 UTC, with and without bias correction. All bias corrected observations used in assimilation used the colocation approach without use of forecast differences (Section 4.1) to obtain bias estimates~~as a function of ozone effective temperature and solar zenith angle~~.

The forecasting model used was the Global Environmental Multiscale (GEM) numerical weather prediction model (Côté et al., 1998a and 1998b; Charron et al., 2012; Zadra et al., 2014a,b; Girard et al., 2014) of Environment and Climate Change Canada coupled to a linearized ozone model (LINOZ) (McLinden et al., 2000; de Grandpré et al., 2016). The LINOZ model uses pre-computed coefficients generated as monthly mean climatologies for calculating the ozone production and sink contributions throughout the stratosphere and upper troposphere down to 400 hPa. A relaxation towards the climatology of Fortuin and Kelder (1998) was imposed between the surface and 400 hPa to constrain deviations away from the climatology, with a relaxation time scale of 2 days. The GEM model was executed with a 7.5 min time step with a uniform 1024×800 longitude-latitude grid and a Charney-Phillips vertically staggered grid (Charney and Phillips, 1953; Girard et al., 2014) with 80 thermodynamic levels extending from the surface to 0.1 hPa. The horizontal grid corresponds to a resolution of ~0.23° in

Formatted: Line spacing: 1.5 lines

Formatted: Font: 10 pt, English (United States), Check spelling and grammar

Formatted: Font: 10 pt, English (United States), Check spelling and grammar

Formatted: Justified, Space After: 0 pt, Line spacing: 1.5 lines latitude and ~0.37° in longitude, representing a 25 km resolution at latitude 49°. In assimilation, inconsistencies stemming from the differences in resolutions between the model forecasts and the observations would usually be reflected by some corresponding increase of applied observation error variances. This is not explicitly done here. The vertical resolution in the upper-troposphere/lower-stratosphere (UTLS) region is in the range of 0.3 to 0.6 km with the resolution gradually changing to ~1.6 km at 10 hPa and 3 km at 1 hPa.

Formatted: Font: 10 pt, Not Bold, English (United States), Check spelling and grammar

Assimilation was done using an incremental three-dimensional variational (3D-Var) approach with first guess at appropriate time (FGAT; Fisher and Andersson, 2001). This assimilation system uses components of the ECCC Ensemble-Variational data assimilation system (Buehner et al., 2013 and 2015) adapted by the authors and P. Du (ECCC) for constituent assimilation and being run without ensembles. Successive short-term three to nine hour forecasts were generated from analyses provided for 00, 06, 12, 18 UTC synoptic times.

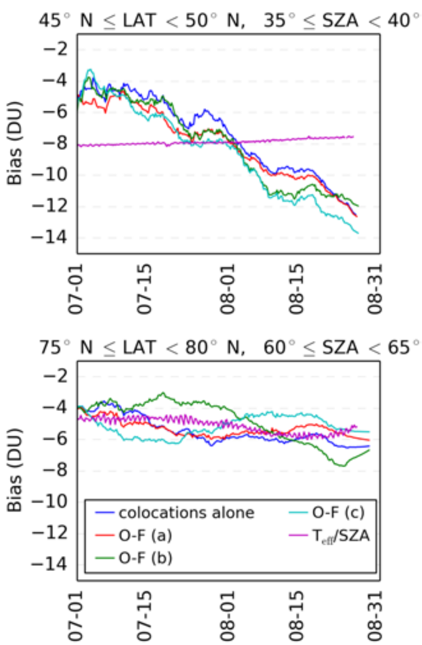


Figure 10. Time series of total column ozone bias corrections for two latitude/SZA bins covering July-August 2014 for GOME-2A using different bias correction methods. All cases that include colocation methods usethinned observation sets. The 'O-F' curves additionally use the differences of forecasts described in Section 2.4.2 following the assimilation of OMI-TOMS. The 'colocations alone' and 'O-F' curves were calculuated using the Gaussian two-week moving average with HWHM of 4.7 days. The 'T_{eff}/SZA' curve is the result of mapping each observation that falls within the latitude/SZA bin onto the ozone effective temperature/SZA bias for July-August 2014 (shown in Fig. 8) and taking the mean of these values.

Formatted: Justified

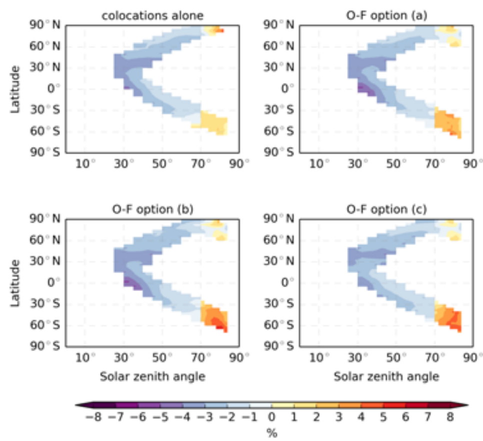


Figure 11. Time mean total column ozone biases between GOME-2A and OMI-TOMS for the period of July-August 2014 from colocation alone and approaches (a), (b), and (c) using observation-minus-forecast differences (see Section 2.4.2). For approaches (a), (b), and (c) the forecasts were taken from the 'OMI' assimilation run (see Table 2). The bias in the 'colocations alone' panel was computed The analyses are a composite of the already available ECCC operational meteorological analysis and the ozone analyses generated from this assimilations study. Assimilation runs were compared to runs without ozone assimilation but that used the same meteorological analyses as employed by the ozone assimilation runs. The initial ozone field used was an analysis from an earlier assimilation.

The background error covariances used have latitude varying error standard deviations with values at sample vertical levels of 1, 10, 50, and 300 hPa in the ranges of ~6-12 %, ~3-5 %, ~5-15 %, and ~15-24 %, respectively. The vertical correlations have half width at half maximum values between 0.5 and 1 km between the top of the boundary layer and 100 hPa, and are nearly equal the model vertical resolutions above 100 hPa with values ranging from ~0.5 km at 100 hPa to ~3.5 km at 1 hPa. The horizontal correlation half widths at half maximum are ~125 km near the surface and increase from ~165 km at 100 hPa to just under 750 km at 1 hPa. The applied observation error standard deviations assigned to all total column measurements of all sources for the conducted assimilations were set to a constant of 2 %.

As assimilating column ozone data from two or more sources ensures that data is continually available in the event of occasional to permanent interruption of data availability from specific instruments, both individual and combined observation datasets were assimilated. For near–real time assimilation, the interruption of the availability of the anchor dataset implies the need for contingency planning for transitions of bias correction references. One might opt to assimilate data from some sensors and monitor the data from others through comparisons with the assimilation analyses. While not necessarily negating the need for bias correction, one could always select to assimilate data from sensors with retrieval products having the smallest initial biases as compared to other products. The effects of bias correction on assimilation when assimilating both individual and multiple sensors will be examined.

using the thinned observation data set as the thinned data set was used in the assimilation.

3.2.3 Approaches based on differences with forecasts

Bias estimation and correction from differences with forecasts with the three options (a) to (c) of Section 2.4.2 was applied for GOME-2A for July-August 2014 using differences with six-hour forecasts. In this study, the forecasts used for these differences came from one of the previously run assimilation/forecasting experiments that are summarized in Table 2. However, the forecasts used for this approach could come from runs that assimilate the bias-corrected observations using the correction method considered in this section. Results from use of the forecasts from the OMI assimilation are provided in Figs. 10 and 11. Differences between the biases resulting from options (a) to (c) and colocation alone biases are within 1 % over the two month period except for a few bins, which are mostly at high SZA. For bins with larger bias estimates with the thinned data, there can be some improvement toward colocation alone results in the absence of thinning such as for (a) at the bin centered at the latitude of 52.5° S and SZA of 65° (Fig. 5 and Fig. S16). This may be due to spatial differences being better reflected in colocation binning through additional use of the forecasts. For this period and dataset, the extension of (a) to use (b) and (c) as successive fallbacks extends the number of usable bins, this being more evident at high SZAs in the Northern Hemisphere. Results from the assimilation of non-corrected data results in differences of bias estimates from colocation alone increasing nearly always by less than 1 % for (a) and (b), with the larger increases for (b), and by roughly up to 1-2 % for (c). The latter would reflect the resultant forecast bias from the assimilated non-corrected datasets. In the absence of ozone assimilation, the increases for (a) and (b) remain usually within 1 % for this period while increasing substantially for (c). [The implications of the other experiments are shown through the sample comparison of Fig. 10 with Figs. S17 and S18.] This demonstrates that use of differences with forecasts, and, by implication, innovations during assimilation runs, is a valid alternative to use of colocation by itself and can provide additional benefits conditional on the forecasts representing well the spatial variation in total column ozone for options (a) and (b) and being sufficiently de-biased for option (c).

3.3 Evaluation of ozone short-term forecasts

Successive short-term forecasts were generated from assimilations performed from 29 June to 31 August 2014 with different sets of total column ozone observations with and without bias correction, as well one experiment with assimilation of MLS ozone profiles. These are compared to each other and to the case without assimilation. All results with bias correction have bias estimates obtained from the colocation approach without use of forecast differences. All presented assimilation experiments except for one were conducted with the original background and observation error standard deviations. The short labels specified in the figure legends as of Fig. 13 identifying the different experiments are described in Table 2.

The applied evaluation metrics consist of mean differences, standard deviations, and anomaly correlation coefficients (ACC), i.e.,; e.g. WMO, 1992) over the July-August period. These were applied to forecasts of both total column ozone and the three-dimensional (3D) ozone field. The ACC, computed over different regions, provides a measure of the spatio-temporal anomalies between an ensemble (or time series) of forecasts and the verifying data (or analyses) as compared to their respective differences to a reference or climatological field, specified here as the 3D temporal mean of the no assimilation case.

Table

Mean differences, $m(O-F) = N^{-1}\sum_{i=1}^{N}(O_i - F_i)$ (1)

Standard deviations, $\sigma(O-F) = \sqrt{(N-1)^{-1}\sum_{i=1}^{N}[(O_i - F_i) - m(O-F)]^2}$ (2)

Anomaly correlation coefficients, $ACC = \frac{(N-1)^{-1}\sum_{i=1}^{N}[(O_i - C_i) - m(O-C)][(F_i - C_i) - m(F-C)]}{\sigma(O-C)\sigma(F-C)} = \frac{\mathrm{cov}(O-C, F-C)}{\sigma(O-C)\sigma(F-C)}$ (3)

where O_i, F_i, and C_i denote observations, forecasts, and climatological values at the observation locations, respectively. The ACC (e.g. WMO, 1992) provides a measure of the spatio-temporal correlation between the deviations of forecasts and a verifying dataset (observations or analyses) from a reference (often a climatological field). For this study, the mean forecast values for the no assimilation case over July-August 2014 were used the reference *C* instead of a climatology. It was verified that choosing the reference in the ACC as the 2D ozone climatology of Fortuin and Kelder (1998) instead does not significantly change the results. As anomaly correlation coefficients in assimilation typically compare forecasts with analyses instead of observations, OMI data in this case, it was also verified that both give similar results. In the tables and legends of the figures referred to in this section (Table 5 and Figs. 8 and 9), the short labels that denote the different assimilation runs are described in Table 5.

We first examine the global differences of Brewer and Dobson total column ozone measurements with six-hour forecasts following assimilation with and without bias correction. The mean and standard deviations of these differences are shown in Table 6. Note that assimilating GOME-2A observations alone without bias correction actually increases the absolute size of the global mean differences relative to the no assimilation case to over 2 %. The smaller value for the no assimilation case stems specifically from the cancellation of larger positive and negative mean differences in the tropical and extra-tropical regions, respectively (Fig. 8). Runs assimilating GOME-2A and OMPS-NM alone, as well as GOME-2A/B and OMPS-NM, have the global mean biases from both Brewers and Dobsons reduced from above to well below 1% when bias correction is introduced. Bias correction reduces the global mean differences to less than 0.3 % in size for all cases. For the south polar region, the inclusion of bias correction in the assimilation of OMPS-NM reduced mean difference from ~4-5 % to ~1-2 % (less reduction is seen for GOME-2 since it does not reach as far south). Introducing assimilation reduces the standard deviations from 3.4-3.8 % to ~2.3-2.9 %, while bias correction further reduces the standard deviations modestly to ~2.3-2.6 %. The standard deviations obtained from the assimilation of uncorrected observations incorporates the effect of the latitude and SZA variation of the biases of the different instruments. This contribution would be reduced when assimilating bias corrected observations. The small reductions in standard deviations from introducing bias correction indicate that the effect from the reduction of bias variability on the variances is small as compared to the sum of the other variance contributions. These contributions could include the variation of inter-station ground-based instrument calibration errors and/or representativeness errors associated to the model resolution, in addition to the forecast errors and the ground-based instruments random errors.

Table 5.2. List of assimilation experiments and their corresponding identifiers. In the second column, an asterisk (*) next to the instrument denotes that the bias-corrected observations (using the colocation method of Section 4.1) were assimilated.

Assimilation experimentexpt identifier	Instruments assimilated	Error std. dev. estimate
CTRL	None	-
OMI	OMI	Original
GOME2A	GOME-2A	Original
GOME2B	GOME-2B	Original
OMPSNM	OMPS-NM	Original
G2AB+NM	GOME-2A/B, OMPS-NM	Original
ALLTC	GOME-2A/B, OMPS-NM, OMI	Original
GOME2A bc	GOME-2A*	Original
GOME2B bc	GOME-2B*	Original
OMPSNM bc	OMPS-NM*	Original
G2AB+NM bc	GOME-2A*/B*, OMPS-NM*	Original
ALLTC bc	GOME-2A*/B*, OMPS-NM*, OMI	Original
MLS+OMI	MLS, OMI	Original
G2AB+NM bc us	GOME-2A*/B*, OMPS-NM*	Updated

*denotes bias-corrected observations

Formatted Table

Formatted: Font: 9 pt, Bold, English (United States)

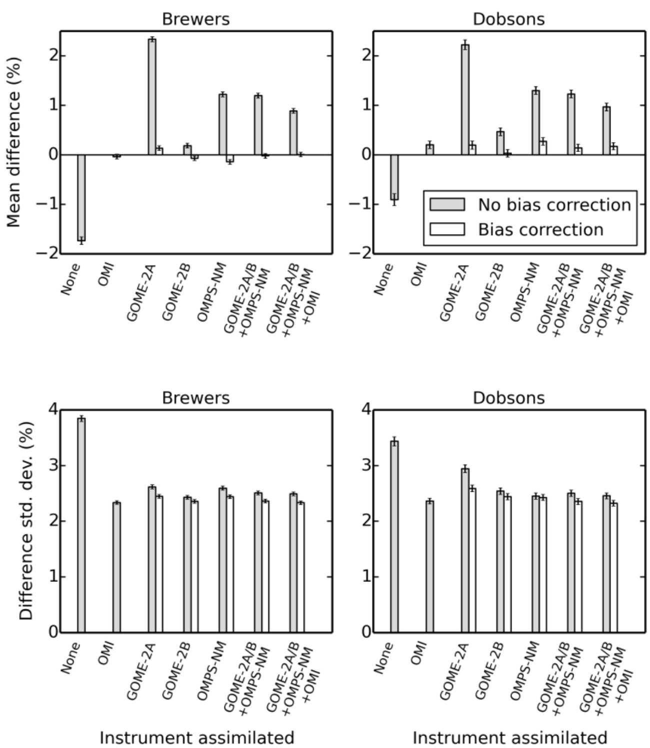


Table 6

~~Figure 12~~. Global mean differences (%) between Brewer and Dobson total column ozone measurements and short-term forecasts for July-August 2014. Bias-corrected~~The instuments assimilated for each run are shown on the horizontal axis and the unshaded (shaded) bars indicate that all~~ observations from~~were (were not) bias corrected. For assimilations that assimilated bias-corrected observations,~~ the colocated observation bias correction scheme (Section 4.1) were applied in the assimilations~~was used~~. The Dobson measurements used were adjusted as a function of the ozone effective temperature (see Section 2.~~2.~~4). The uncertainties~~error bars~~ denote the standard error ~~square root of the sample variance~~ of the mean differences~~difference~~ or ~~difference~~ standard deviations of the differences~~deviation~~. The data from the two Antarctic stations have been included here even though their mean differences with OMI are outliers relative to most mean differences~~differnces~~ (Tables S1 and S2).

	Assimilated instruments	Mean difference (%)		Difference std. dev. (%)	
		No bias correction	Bias correction	No bias correction	Bias correction
Brewers	None	-1.73 ± 0.08	-	3.85 ± 0.05	-
	OMI	-0.03 ± 0.05	-	2.34 ± 0.03	-
	GOME-2A	2.33 ± 0.05	0.13 ± 0.05	2.62 ± 0.04	2.45 ± 0.03
	GOME-2B	0.19 ± 0.05	-0.07 ± 0.05	2.43 ± 0.03	2.36 ± 0.03
	OMPS-NM	1.22 ± 0.05	-0.14 ± 0.05	2.59 ± 0.04	2.44 ± 0.03
	GOME-2A/B + OMPS-NM	1.20 ± 0.05	-0.02 ± 0.05	2.51 ± 0.03	2.36 ± 0.03

	GOME-2A/B + OMPS-NM + OMI	0.89 ± 0.05	0.01 ± 0.05	2.49 ± 0.03	2.33 ± 0.03
Dobsons	None	-0.91 ± 0.12	-	3.43 ± 0.08	-
	OMI	0.20 ± 0.08	-	2.36 ± 0.05	-
	GOME-2A	2.22 ± 0.10	0.20 ± 0.08	2.94 ± 0.07	2.59 ± 0.06
	GOME-2B	0.47 ± 0.08	0.03 ± 0.08	2.54 ± 0.06	2.44 ± 0.05
	OMPS-NM	1.30 ± 0.08	0.27 ± 0.08	2.45 ± 0.06	2.43 ± 0.05
	GOME-2A/B + OMPS-NM	1.23 ± 0.08	0.14 ± 0.07	2.51 ± 0.06	2.36 ± 0.05
	GOME-2A/B + OMPS-NM + OMI	0.97 ± 0.08	0.17 ± 0.07	2.46 ± 0.06	2.32 ± 0.05

3.3.1 Impact on total column ozone forecasts

Global differences, over the two month period, of Brewer and Dobson total column ozone with short-term forecasts following assimilation without bias correction span the range of near zero to about 2.3 %, nearly all being positive, as compared to about -1.8 % for Brewers and -0.8 % for Dobsons in the absence of assimilation (Fig. 12). These mean differences for the assimilation cases without bias correction are consistent with the global mean differences from the colocation comparison of the satellite and ground-based observations for July-August 2014 provided in the first paragraph of Section 3.1. The mean differences with forecasts from the assimilation of the uncorrected provisional OMPS-NM are only slightly larger than direct differences with OMPS-NM indicated in section 2.2.3 for more recent product versions. Bias correction reduces the global differences to less than 0.3 % in size for the various cases, be it from assimilation of individual or multiple data sources. The assimilating GOME-2A observations alone without bias correction actually increases the absolute global mean differences relative to the no assimilation case. Introducing assimilation reduces the standard deviations from 3.4-3.8 % to ~2.3-2.8 %. The reduction of standard deviations from bias correction is less than 0.35 %. These smaller reductions in standard deviations are limited by the error level of the observations and forecasts and the spatio-temporal variability of the biases about the global mean differences as shown in the following sentence. Brewer and Dobson error standard deviations of 1.5 %, a forecast error standard deviation of 0.5 % (likely an underestimate) and, considering Table 1, a standard deviation over the stations time mean differences of 1.5 % give a resultant standard deviation of 2.2 % from the square root of the sum of squares, quite close to the numbers from Fig. 12.

Comparisons of OMI-TOMS measurements with forecasts for the various experiments with and without bias correction and without any assimilation are shown in Figs. 8 and 913 to 16 for the July-August 2014 period. For assimilation of only GOME2-A, in most of Figure 13 (see also Figs. S19 and S20) displays the topics and northern extra-tropics, the reduction of the time mean differences from assimilationwhich seem more prominent in the tropical region for the control case, i.e. without assimilation, reaching above 10 %. The separate assimilations of all column ozone sources with bias correction as compared to the no assimilation case is roughly the same order of magnitude as the reductions resulting from introducing bias correction as compared to the no (labelled 'ALLTC bc') and of OMI yield mean differences with OMI usually within ±1 %. The absence of bias correction case. However, assimilation of the other instrumentsresults in an overall time-mean reduction of column ozone by 1-3 % due to the significance of the GOME-2A and OMPS-NM biases. The results of assimilations of various individual and combined sources in terms of mean differences, standard deviations, and anomaly correlation, are provided for different latitude bands in Figs. 14 and 15. These show that the first order improvements stem from assimilation in general, while bias corrections result in and updating of error standard deviations (case 'G2AB+NM bc us' of Fig. 15) comparatively imply second order changes. GOME-2A and OMPS-NM show the largest reductions in mean differences from bias correction as would be expected from Fig. 3. Figure 15 shows that the updated error variances have resulted in a reduction in impact in some areas due to the larger reduction of the background error variances relative to the change in the observation error variances. Adding the assimilation of MLS ozone profiles to that of OMI-TOMS has small varying effects within 1 % on total column ozone mean differences as a function of latitude except in Antarctica at just over 1.5 % (Fig. S21). The latter seems to support an underestimation of OMI-TOMS in Antarctica by at least that amount for July. The added MLS assimilation also suggests that any OMI-TOMS underestimation in the Arctic would be much less for that period.

For ACC, forecasts from the assimilation of GOME-2B in the tropics appear better than from the assimilation of OMI-TOMS when compared to the OMI-TOMS observations. This is likely due to the larger volume of GOME-2B data, in addition to its low bias. As well, the ACC is the diagnostic that demonstrates more marked improvements in multiple sensor assimilation as compared to OMI-TOMS assimilation. The advantage of multiple sensor assimilation is therefore more notable in increasing the quality of the pattern and variation of the forecast fields. As anomaly correlation coefficients in assimilation typically compare forecasts with analyses instead of observations, OMI data in this case, it was verified, for completeness, that both give similar results (Fig. 14 compared to Fig. S22). It was also verified that the choice of the reference field (climatology) did not change the overall qualitative results; switching from the time average of the no assimilation case (CTRL) to the 2D ozone climatology of Fortuin and Kelder (1998) results in changes to the ACC for the case without assimilation of within ~0.l and of much less for the column ozone assimilation cases (Fig. S23).

In the absence of assimilation, the mean differences in the extra-tropics have opposite sign to those in the tropics. This stems, at least partly if not entirely, from the differences between the observations and tendency of the forecast in the absence of assimilation to move toward the ozone model equilibrium state. This is also seen by the temporal changes of the mean differences of Fig. 16 over the two month period having opposite tendencies for the extra-tropics and the tropics. The temporal changes are believed to result from a long spinup period in moving from the initial ozone field based on an earlier assimilation, toward the ozone model equilibrium state. Beginning with an initial ozone field at the model equilibrium state would likely not have improved the ACC of the control case, as implied by Fig. 16, and would have increased its mean observation minus forecast differences. The changes in time of Fig. 16 also reflect the high predictability of ozone medium range forecasts, with an increase in total column ozone forecast regional mean error of less than 5 % in ten days, this in addition to the relevance of assimilation. Both the temporally averaged and time varying mean differences of forecasts with OMI-TOMS wereare reduced to within 1 % over the latitude ranges where satellite data are assimilated for the allvarious cases with bias correction, with the results for GOME-2A only assimilation being the exception, slightly exceeding 1 % in some places. The GOME-2A and OMPS-NM datasets show the largest reductions in mean differences from bias correction, as would be expected from Fig. 2, where these biases are reduced from levels of ~1-3 % when no bias correction is performed to well within 1 % for bias correction cases (excluding latitudes below 60°S). The assimilation of bias corrected observations from multiple sensors (labelled as 'ALLTC bc') does not notably reduce the mean differences as compared to the assimilation of individual bias corrected sensors. Considering the earlier comparisons of forecasts with ground-based data and these results, the reduction of biases to the 1 % target appears to be achieved for the short-term forecasts in most%. Figures 14 and 16 both show the strongest improvement from bias correction of multiple sensor assimilation ('ALLTC' and 'ALLTC bc') in the northern extra-tropics. The regions with assimilated data.

Assimilation of total column observations improves the standard deviations of differences between the six-hour forecasts and OMI-TOMS across all latitudes, as seen in Fig. 8, although relatively little impact is seen for the GOME-2A/B assimilations in the southern extra-tropics where relatively few observations are available. The impact of bias correction on standard deviations of forecast is not very significant.of most significant temporally averaged impact from assimilation are near the Antarctic polar vortex and in the tropics. The large mean differences and standard deviations for GOME-2A/B assimilations below 60°S stem from these datasets not reaching much further south during this period. This reflects the importance of observations nearin the winter poles in the absence of heterogeneous chemistry in LINOZ.

In the absence of assimilation, the mean differences between the forecasts and OMI-TOMS observations in the extra-tropics have opposite sign to those in the tropics, as seen in the top panels of Fig. 8. Also, notice that in Fig. 9 the mean differences in the extra-tropics diverge in the opposite direction as compared to the tropics. The drift of the mean biases in time in the absence of assimilation are due to the tendency of the forecast to move toward the ozone model equilibrium state. For the GEM-LINOZ model, this results in a long spin-up period in which ozone field moves from the initial ozone field, based on an earlier assimilation, toward the ozone model equilibrium state. Beginning with an initial ozone field at the model equilibrium state would have increased its mean observation minus forecast differences and would likely not have improved the ACC of the control case, as implied by Fig. 9. Also from Fig. 9, we can see that the error of the total column ozone forecast increases by less than 5 % over the course of fifteen days, reflecting the high predictability of ozone medium range forecasts. This limited deterioration would not deter, for example, in properly forecasting the movement of low column ozone regions during these periods and the corresponding changes in clear-sky UV Index.

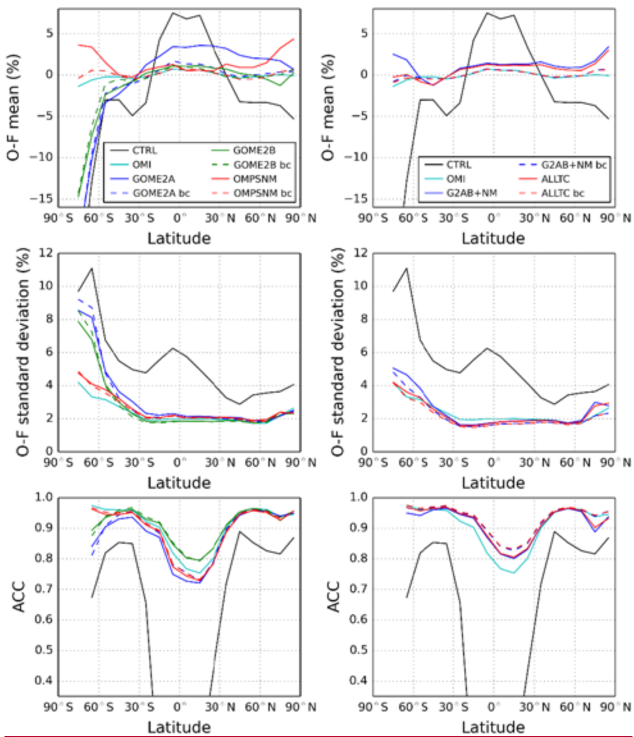
For the ACC, forecasts from the assimilation of GOME-2B in the tropics appear better than from the assimilation of OMI-TOMS when compared to the OMI-TOMS observations. This is likely due to the larger volume of GOME-2B data (when comparing the thinned dataset from spatial sampling at 1° resolution), in addition to its low bias. Furthermore, the ACC demonstrates a more marked improvement in multiple sensor assimilation in the tropical region as compared to OMI-TOMS assimilation alone, which is not seen in the mean differences. The advantage of multiple sensor assimilation is, therefore, more notable in increasing the quality of the pattern and variation of the forecast fields.

The deterioration of the ACC with time in Fig. 9, as well as the low time mean ACC in the tropics in Fig. 8 The deterioration of the ACC with time in Fig. 16, and the low time mean ACC in the tropics in Figs. 14 and 15, in the absence of assimilation reflects an increase in the spatio-temporal variations of the observation-minus-forecast differences as compared to the cases with assimilation. To examine this further, we can rewrite the expression for the anomaly correlation coefficient in observation space as

$$ACC = \frac{\mathrm{cov}(O-C, F-C)}{\sigma(O-C)\sigma(F-C)} = \frac{1}{2}\frac{\left[\sigma^2(O-C)+\sigma^2(F-C)-\sigma^2(O-F)\right]}{\sigma(O-C)\sigma(F-C)} \quad (4$$

(1)

wherewhere *O* are observations and *C* a climatology used to evaluate forecasts *F* and σ are the standard deviations of the quantity in its brackets. As shown in Fig. 10,In the case when assimilation is not performed, during the time period when the ACC deteriorates rapidly in the tropics, (as seen in Fig. 16), $\sigma(O\text{-}C)$ and $\sigma(F\text{-}C)$ do not change substantially (roughly at 14 DU and 9 DU, respectively), while $\sigma(O\text{-}F)$ increases from about 10 to 20 DU, illustrating the temporal deterioration in the tropics fromform the model. Similar increases in $\sigma(O\text{-}F)$ are also seen in the other regions as well for the no assimilation casethough. Introducing assimilation rapidly and substantially reduces the values of $\sigma(O\text{-}F)$ to around 5-7 DU while pushing the values of $\sigma(F\text{-}C)$ up closer to that of $\sigma(O\text{-}C)$, so that the first two terms in Eqn. 41 are of roughly the same size and much larger than the third term in regions where measurements are assimilated. This results, resulting in ACC values notably closer to unity., this being observed in all latitudinal regions (Fig. S24).



While the $\sigma(O\text{-}F)$ values at mid-latitudes and in the polar regions can be smaller or larger than in the tropics when no assimilation is performed (Fig. 14), their sizes relative to both $\sigma(O\text{-}C)$ and $\sigma(F\text{-}C)$ are smaller in the extra-tropics (Fig. S24). The larger $\sigma(O\text{-}C)$ and $\sigma(F\text{-}C)$ in the extra-tropics result in larger ACCs as compared to the tropics. The smaller relative size of $\sigma(O\text{-}F)$ in the extra-tropics stems from both the observations and forecasts being correlated to the spatio-temporal changes in weather patterns.

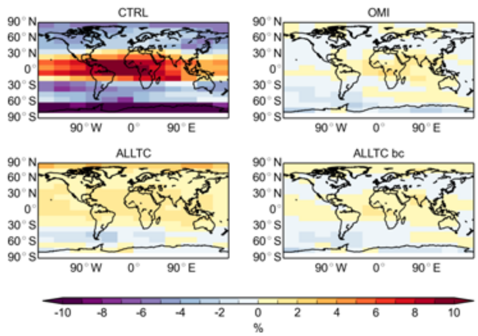


Figure 813. Mean total column ozone differences between OMI-TOMS measurements and short-term forecasts as a function of spatial location for July-August 2014. The plot titles indicate the assimilation run (see Table 2 for description). In the case of no assimilation, many of the values south of 60º S exceed the lower limit of the colour bar, in some instances with values as low as -30%.

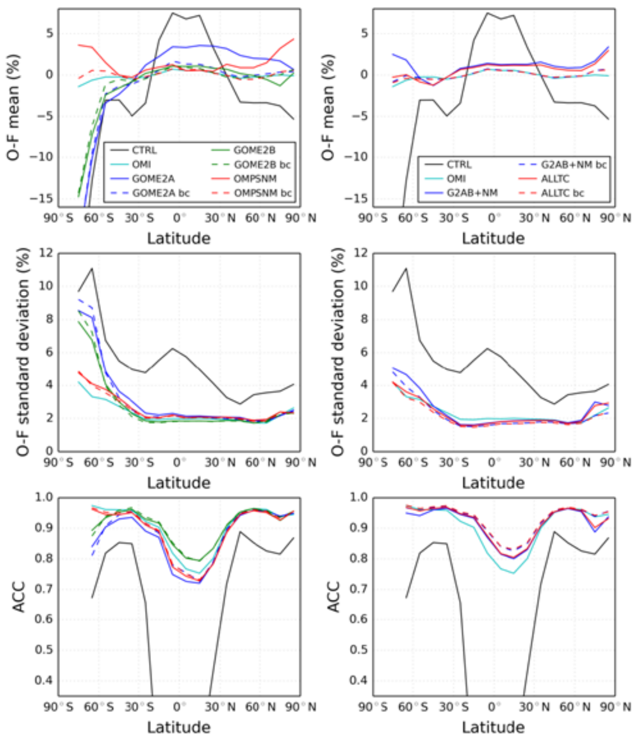
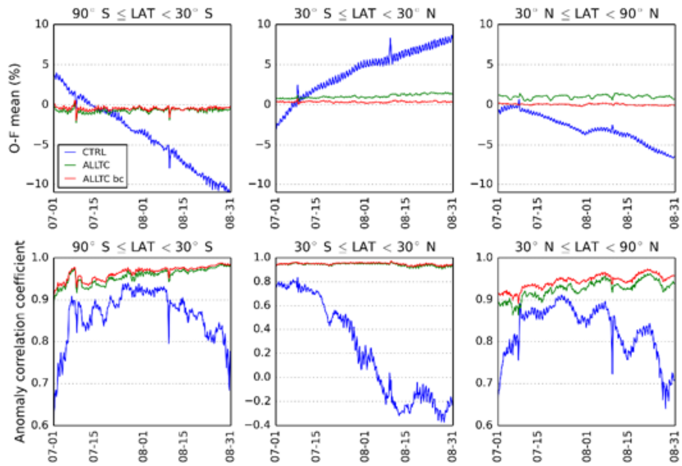


Figure 14. Zonal mean total column ozone statistics of mean differences (%), standard deviations (%), and anomaly correlation coefficients (ACC; unitless) as a function of latitude (degrees) for the comparison between OMI-TOMS measurements and short-term forecasts for July-August 2014. The legends in the top plots indicate the assimilation runs (see Table 6 for description) and apply to all plots in the same column.

Formatted: English (United Kingdom), Check spelling and grammar



90° S ≤ LAT < 30° S
30° S ≤ LAT < 30° N
30° N ≤ LAT < 90° N
O-F mean (%)
Anomaly correlation coefficient
CTRL
ALLTC
ALLTC bc
07-01
07-15
08-01
08-15
08-31

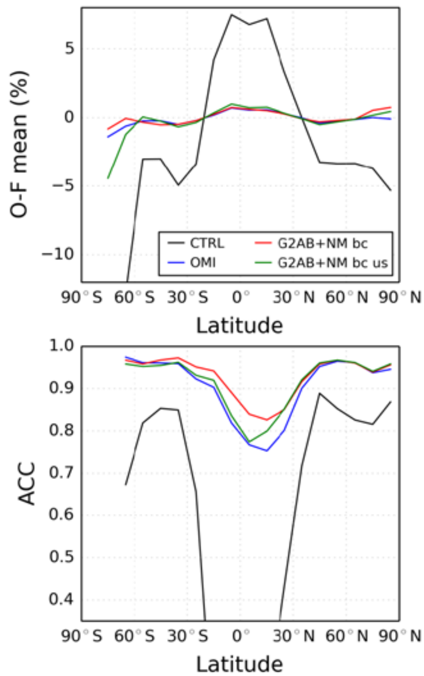


Figure 915. Zonal mean total column ozone differences (%) and anomaly correlaton coefficients (ACC) between OMI-TOMS observations and short-term forecasts for July-August 2014 showing the effect of using the updated sets of observation and background error variances in the assimilation. The legend indicates the assimilation run (see Table 2 for description).

Formatted: Font: Not Bold, Check spelling and grammar

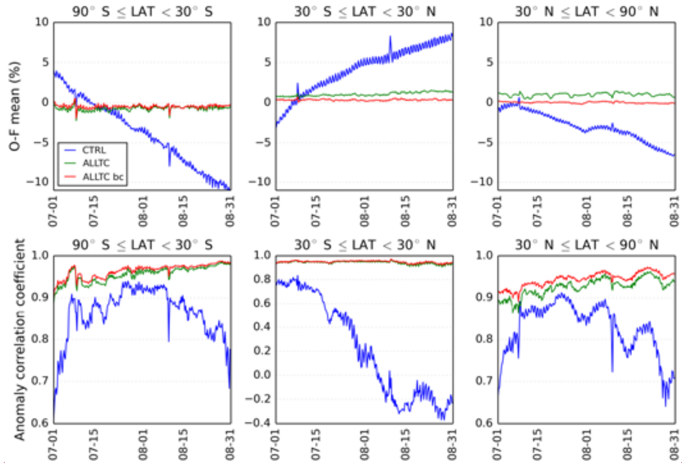
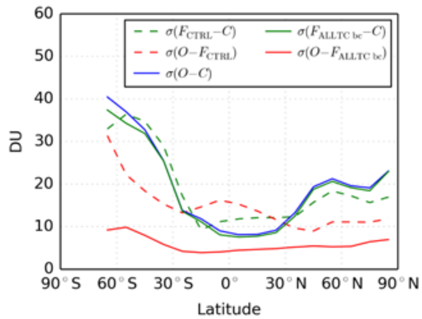


Figure 16. Zonal mean differences and anomaly correlation coefficients (unitless) for total coumn ozone between OMI-TOMS observations and short-term forecasts -as a function of time. Results are shown for the case without assimilation as well as with the assimilation of OMI, GOME-2A/B, and OMPS-NM (both with and without bias correction). The legend indicates the assimilation run (see Table (date). Results are shown for the case without assimilation as well as with the assimilation of OMI, GOME-2A/B, and OMPS-NM (both with and without bias correction). The legend indicates the assimilation run (see Table 6 for description).2 for description). Each value plotted was calculated using a 24 hour time window.



3.3.2 Impact on the vertical structure of the ozone field

The short-term ozone field forecasts from a few assimilation experiments are compared to MLS and ozonesonde profiles using the same metrics and for the same period as above. Significant impact of total column ozone assimilation on forecasts is seen in Figs. 17 to 20, usually below 10-20 hPa, with the largest influences in the lower half of the stratosphere. A reduction in impact above 2-5 hPa stems at least partly from the short photochemical time scales at these levels reflected in the ozone forecast model. Deteriorations are also observed in the lower stratosphere and upper troposphere regions in some panels, e.g. the mean observation minus forecast differences in the northern hemisphere below about 70 hPa for Fig. 17. The presence of improvements and deteriorations as a function of vertical level for the mean differences of Fig. 17 (upper panels) and 18 seems to reflect, to some degree, a vertically weighted profile shift needed for the non-assimilated case to better agree with the total column ozone observations. This is consistent with not having negative background vertical error correlations. The most significant improvements from total column ozone assimilation alone are found in the tropics for the anomaly correlation coefficients throughout the lower stratosphere, as seen in Fig. 17, followed by the time mean differences in the same region. This partly stems from the relative change in vertical distribution of mixing ratio increments for each total column measurement being proportional to the product of the background error covariance matrix and the model pressure layer thicknesses vector. The former includes not only the influence of the background error variances and vertical correlations, but also the influence of the background horizontal error correlations increasing away from the surface, which would impact local profile increments considering the proximity of neighbouring total column measurements. The relaxation time scale of two days toward climatology in the lower troposphere would also have some impact on the forecasts moving away from the influence of the observations. Figures 17 and 19 also show, in support of Fig. 15, that the updated error variances have resulted in a slight reduction in impact from assimilation in the upper troposphere and lower stratosphere.

While some improvements in the vertical structure can result from assimilation of total column ozone, unsurprisingly, the greatest improvements are seen, in Figs. 18 and 20, when adding the assimilation of ozone profile measurements as done here with Aura-MLS (see also Struthers et al., 2002, as another example), which occurs even though the horizontal density of profiles is much smaller than from total column mapping instruments (e.g., Fig. S1).

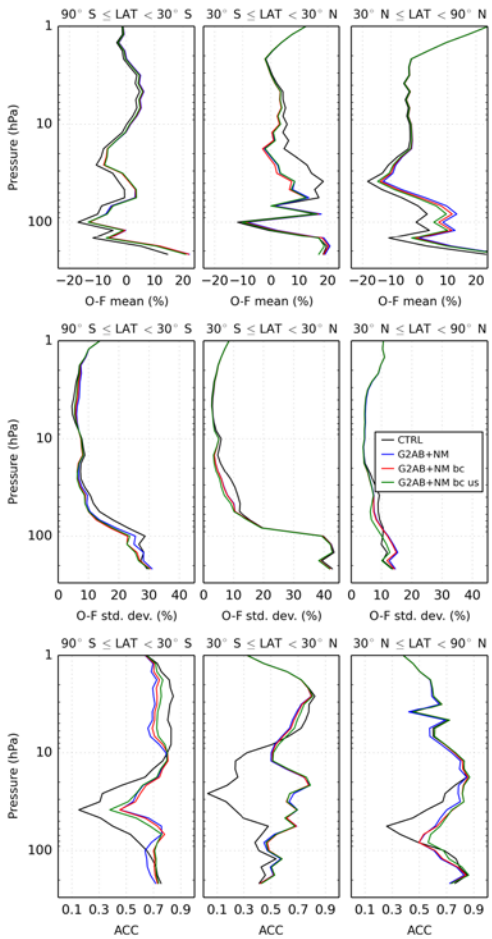


Figure 10. Zonal mean ~~**17.** Mean differences,~~ standard derivations (DU)~~deviations~~ of the differences~~, and anomaly correlation coeffients (ACC)~~ between ~~MLS observations and~~ short-term forecasts *F*, OMI-TOMS observations *O*, and climatological values *C* over July-August 2014. For the short-term forecasts,~~The legend indicates~~ the assimilation run from which the forecast was launched is indicated in the subscript, with the labels described in ~~(see~~ Table 5.~~2 for description)~~.

6

Formatted: Justified

Formatted: Font: Italic, Check spelling and grammar

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 10 pt, Bold, English (United States), Check spelling and grammar

Formatted: Justified, Line spacing: 1.5 lines

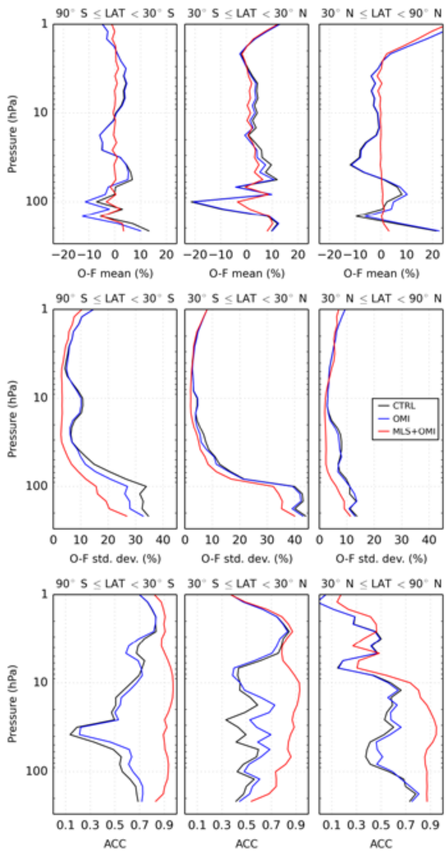


Figure 18. Similar to Fig. 17 but only covering 1-23 July 2014. This shorter period in this figure is imposed by the 'MLS+OMI' assimilation having been conducted only for this duration prior to local computing system changes.

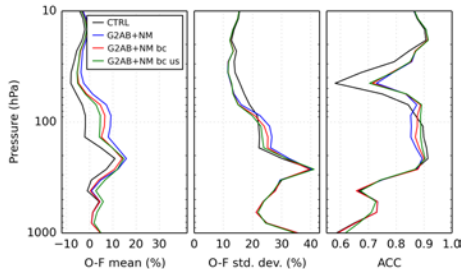


Figure 19. Global mean and standard deviation differences, and anomaly correlation coeffients between ozonesonde observations and short-term forecasts over July-August 2014. The legend indicates the assimilation run (see Table 2 for description).

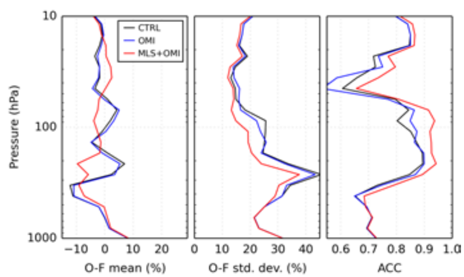


Figure 20. Similar to Fig. 19 but over 1-23 July 2014.

4 Conclusions

Bias correction of total column ozone data from satellite instruments was performed using threea few different approaches. Two of the methods parameterized the relying on bias estimation as a function of latitude, and solar zenith angle binning, and time, while the remaining method used the ozone effective temperature instead in place of latitude and time. These approaches consisted of using observation with a two-week moving time window. The approach consisting of applying colocation between satellite-borne instruments and a reference, referred to in this paper as the anchor. Different variants of the bias estimation scheme were explored, including examining the effect of including by itself was compared to variants that used short-term forecasts within . The differences in corrected biases between these approaches were nearly always under 1 %. These were also compared to the bias estimation. Differences between bias estimations from different methods that used the latitude/ as function of ozone effective temperature and solar zenith angle parameterization were generally within 1 %. without the two-week moving time window. While the two month time-averaged bias estimates from the ozone effective temperature

Formatted: Left

parameterizationcorrections from this scheme were similar to those from the other approaches, the lackrelevance of an explicitexplicitly including time dependence caused departureswas shown to potentially matter, this likely depending on shorter time scales between these estimates and those from methods that include an explicit time dependence, where these estimates were different by ~2-3 % for some instrumentsthe instrument and/or retrieval process.

The anchor used in thedata source from which bias estimation schemesestimates were generated was chosen asthe OMI-TOMS data product. The quality of the OMI-TOMS data product, dueis supported by other publications in addition to its wide coverage in both time and space, and its good agreement with ground-based instruments. In this study, for the timeevaluations on two month periods examined, OMI-TOMS was foundin this paper through comparisons to have global and regional mean differences with ground-based Brewer and Dobson spectrophotometers, and filter ozonometers . Comparison of OMI-TOMS with the ground-based data yielded regional and global mean differences within 1 %, except in the polar regions. Similar to larger mean differences of OMI-TOMS with SBUV/2 data were found, with OMI-TOMS generally being in better agreement to the ground-based data for the examined periods.

For the July-August 2014 and January-February 2015 periods, the observations based on TOMS retrievals for the GOME-2A instrument were found to have the largest mean slightly larger differences with OMI-TOMS, which could be as high as 8 % in some regions of the parameter space for solar zenith angles below 70°. The GOME-2B instrument showed much better agreement with OMI-TOMS, with mean differences generally confinedwere found with SBUV/2 data. While no adjustments were made in this paper to ~1-2 %, excluding at very high solar zenith angles. The provisional OMPS ozone column products, both the total column and summed partial column profile, typically had mean differences somewhere between the two GOME-2 instruments, with mean differences generally confined to ~3-4 % (again excluding high solar zenith angle regions). As the quality of the different versions of OMPS retrieved data may differ, one might expect a reduction in bias of the more recent version of the OMPS products based on the SBUV V8.6 retrieval algorithmsthe OMI-TOMS data, the choice of the anchor does not exclude the possibility of applying pre-determined corrections to its data when deemed necessary.

Formatted: English (United Kingdom)

It was demonstrated that Prior to bias correction for the assimilation of assimilated data, the GOME-2B (TOMS) products were found to be closest to OMI-TOMS, and thus to the ground-based data, with GOME-2A showing the largest differences. The differences for the provisional OMPS-NM products were typically in between these two cases.

Three-dimensional variational assimilations of total column ozone observations that include data measured from the satellite instruments have demonstrated the capability of reducing short-term forecast errors of total column ozone to within 1 % of the OMI-TOMS data in the latitude regions covered by the assimilated data when including bias corrections as derived in this study can improve the agreement between short-term forecasts and ground-based measurements. Using a three-dimensional variational applied in the paper. This applies to assimilation of single and multiple data sources, except for assimilation system, the assimilation of GOME-A without bias correction gives global and time mean differences between ground-based observations and short-term ozone forecasts of ~2.3 %. The assimilation of uncorrected OMPS-NM measurements reduced these mean differences of GOME-2A alone with residuals slightly to ~1.3 %. Assimilating instead the bias corrected observations brought these mean differences to well within 1 %. As minimal exceeding 1 % at some latitudes. For the assimilation of all instruments, while possible cancelation of different errors from different instruments is a factor in reducing forecast biases, harmonizing the differing datasets through bias correction better ensures achieving the target reduction in residual bias was found for GOME-2B, the assimilation of both corrected and uncorrected GOME-2B observations yielded mean differences within 1 %.. The benefit of including total column satellite data, even without bias correction, was most notable in the tropics, in addition to the polar vortex region. ; this might vary depending on the ozone forecast model. This was evident not only in improvements of mean differences, but also of anomaly correction coefficients. The most notable improvement from total column ozone assimilation without the profile data was obtained for the anomaly correlation coefficients in the tropical lower stratosphere. Based on the shorter assimilation with Aura-MLS, the addition of this profile data did not significantly alter total column ozone short-term forecasts in the regions where the total column data was assimilated; it can be surmised that it would improve the total column ozone forecasts in the winter pole regions when there are no total column ozone data available. As one would or might expect and as has been illustrated, the added assimilation of good quality profile data is important in assuring a better quality of the vertical structure of the ozone forecast field even if the spatial density of profile measurements over six hour periods is much less than that of the total column data.

The aforementioned results indicate that the reduction of biases to the 1 % target was achieved for most regions and cases, a likely exception being for conditions with high solar zenith angles. For the assimilation of two of more satellite sensors, while it is possible that the cancellation of errors from different instruments could reduce forecast biases, harmonizing the different datasets through bias correction better ensures target reductions in residual biases are achieved.

Code and data availability. The assimilation and forecasting system is integrated into the unique operational computing environments of ECCC. As well, the computing hardware used for these assimilation cycles has since been replaced at ECCC with accompanying changes to the cycling package. References of the system components are provided in this paper. The assimilation of bias corrected observations from multiple sensors does not notably reduce the mean differences as compared to the assimilation of individual bias corrected sensors. However, a notable improvement in multiple sensor assimilation was seen in the tropical region as compared to OMI-TOMS assimilation alone with the anomaly correlation coefficients metric. This improvement implies an increase the quality of the pattern and variation of the forecast fields.

The Fortran-based*Code and data availability.* The bias estimation and correction software with related shell scripts can be provided with the understanding that users will need to adapt the code to their preferred input/output data file formats. The observations can be obtained from the different centres identified in the text and the acknowledgments Section below. The assimilation and forecasting system relies on ECCC computing environment tools and file conventions–. As well, the computing hardware used for these assimilation cycles has since been replaced at ECCC with accompanying changes to the cycling package. References of the system components are provided in this paper. The large sets of model analyses and forecasts, and the observation minus forecast datasets, are saved with an in-house binary file format. Subsets could potentially be made available from the authors upon request. In addition to also containing a few complementary figures, the Supplement provides tables of station by station mean differences of OMI-TOMS with ground-based data related to Table 1 and Fig. 14.

The Supplement related to this article is available online at https://doi.org/XXXX-supplement.

Formatted: Font: +Headings (Times New Roman), Bold

Competing interests. The authors declare that they have no conflict of interest.

Acknowledments. The authors wish to thank Lawrence E. Flynn (NOAA) and Vitali Fioletov (ECCC) for information and advice regarding use of the different satellite datasets and the Brewer ground-based data, respectively, Jean de Grandpré and Irena Ivanova (ECCC) for the availability and assistance in use of the version of GEM with LINOZ, Ping Du and Mark Buehner (ECCC) for contributions in extending the variational assimilation code for use with constituents, Jose Garcia (ECC) and Vaishali Kapoor (NOAA), among others associated to NOAA for contributions, in facilitating data acquisition and the various instrument teams having generated the different observation sets. We also gratefully acknowledge the following centres for access to the observations used in this paper: the National Environmental Satellite, Data, and Information Service of the National Oceanic and Atmospheric Administration (NESDIS/NOAA), the Earth Observing System Data and Information System of the National Aeronautics and Space Administration (EOSDIS/NASA), the World Ozone and Ultraviolet Radiation Data Center (WOUDC), and the Global Monitoring Division of the NOAA Earth System Research Laboratory.

References

Adams, C., Bourassa, A. E., Sofieva, V., Froidevaux, L., McLinden, C. A., Hubert, D., Lambert, J. C., Sioris, C. E., and Degenstein, D. A.: Assessment of Odin-OSIRIS ozone measurements from 2001 to the present using MLS, GOMOS, and ozone sondes., Atmos. Meas. Tech. 7, 49-64, doi:10.5194/amt-7-49-2014, 2014.

Andersson, E., and Järvinen, H.: Variational quality control. Q. J. R. Meteorol. Soc., 125, 697–722. doi: 10.1002/qj.49712555416, 1999.

Antón, M., López, M., Vilaplana, J. M., Kroon, M., McPeters, R, Bañón, M., and Serrano, A.: Validation of OMI-TOMS and OMI-DOAS total ozone column using five Brewer spectroradiometers at the Iberian peninsula, J. Geophys., Res., 117, D14307, doi:10.1029/2009JD012003, 10 pp., 2009a.

Antón, M., Loyola, D., López, M., Vilaplana, J. M., Bañón, M., Zimmer, W. and Serrano, A.: Comparison of GOME-2/MetOp total ozone data with Brewer spectroradiometer data over the Iberian Peninsula, Ann. Geophys., 27, 1377-1386, 2009b.

Antón, M., Loyola, D., Clerbaux, C., López, M., Vilaplana, J.M., Bañón, M., Hadji-Lazaro, J., Valks, P., Hao, N., Zimmer, W., Coheur, P. F., hurtmans, D., and Alados-Arboledas, L.: Validation of the Metop-A total ozone data from GOME-2 and IASI using reference ground-based measurements at the Iberian Peninsula, Remote Sens. Environ., 115, 1380-1386, 2011.

Bai, K., Liu, C., Shi. R. and Gao, W.: Comparison of Suomi-NPP OMPS total column ozone with Brewer and Dobson spectrophotometers measurements, Front. Earth Sci., 9(3), 369–380, doi:10.1007/s11707-014-0480-5, 2015

Formatted: English (United Kingdom)

Bai, K., Liu, C., Shi, R., and Gao, W.: Validation of the Suomi NPP Ozone Mapping and Profiler Suite total column ozone using Brewer and Dobson spectrophotometers, Atmos. Meas. Tech. Discuss., 6, 4577-4605, 2013.

Bai, K., Chang, N. B., Yu, H. and Gao, W.: Statistical bias correction for creating coherent total ozone record from OMI and OMPS observations, Remote Sens. Environ., 182, 150-168, 2016.

Bak, J., Liu. X., Kim, J. H., Chance, K., and D. P. Haffner, Validation of OMI total ozone retrievals from the SAO ozone profile algorithm and three operational algorithms with Brewer measurements, Atmos. Chem. Phys., 15, 667–683, doi:10.5194/acp-15-667-2015, 2015

Balis, D., Kroon, M., Koukouli, M. E., Brinksma, E. J., Labow, G., Veefkind, J. P., and McPeters, R. D.: Validation of Ozone Monitoring Instrument total ozone column measurements using Brewer and Dobson spectrophotometer ground-based observations, J. Geophys. Res., 112(D24), 2007a.

Balis, D., Lambert, J.-C., Van Roozendael, M., Spurr, R., Loyola, D., Livschitz, Y., Valks, P., Amiridis, V., Gerard, P., Granville, J., and Zehner, C.: Ten years of GOME/ERS2 total ozone data—The new GOME data processor (GDP) version 4: 2. Ground-based validation and comparisons with TOMS V7/V8, J. Geophys. Res., 112(D7), 2007b2007a.

Balis, D., Kroon, M., Koukouli, M. E., Brinksma, E. J., Labow, G., Veefkind, J. P., and McPeters, R. D.: Validation of Ozone Monitoring Instrument total ozone column measurements using Brewer and Dobson spectrophotometer ground-based observations, J. Geophys. Res., 112(D24), 2007b.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, Q. J. R. Meteorol. Soc. 134, 1951–1970 , doi: 10.1002/qj.339, 2008.

Bhartia, P. K., McPeters, R. D., Flynn, L. E., Taylor, S., Kramarova, N. A., Frith, S., Fisher, B. and DeLand, M.: Solar Backscatter UV (SBUV) total ozone and profile algorithm. Atmos. Meas. Tech., 6(10), 2533–2548. doi:10.5194/amt-6-2533-2013, 2013

Bhartia, P. K. and Wellemeyer, C. W.: TOMS-V8 total O3 Algorithm, Chapter 2 in OMI Algorithm Theoretical Basis Document, Volume2, OMI Ozone Products, edited by P. K. Bhartia, 15-31, 2002. [Available from http://projects.knmi.nl/omi/documents/data/OMI_ATBD_Volume_2_V2.pdf]

Bhartia, P. K., McPeters, D., Mateer, C. L,. Flynn, L. E., and Wellemeyer, C.: Algorithm for the estimation of vertical ozone profiles from the backscattered ultraviolet technique, J. Geophys. Res., 101, 18,793-18,806, 1996.

Bernhard, G., Evans, R. D., Labow, G. J. and Oltmans, S. J.: Bias in Dobson total ozone measurements at high latitudes due to approximations in calculations of ozone absorption coefficients and air mass, J. Geophys. Res., 110, D10305, doi:10.1029/2004JD005559, 2005.

Buehner, M., Morneau, J. and Charette, C.: Four-dimensional ensemble–variational data assimilation for global deterministic weather prediction. Nonlinear Processes Geophys., 20, 669–682, doi:10.5194/npg-20-669-2013, 2013.

Buehner, M., McTaggart-Cowan, R., Beaulne, A., Charette, C., Garand, L., Heilliette, S., Lapalme, E., Laroche, S., Macpherson, S .R., Morneau, J., and Zadra, A.: Implementation of Deterministic Weather Forecasting Systems based on

Ensemble-Variational Data Assimilation at Environment Canada. Part I: The Global System Mon. Wea. Rev., 143, 2532-2559. doi: 10.1175/MWR-D-14-00354.1, 2015.

Callies, J., Corpaccioli, E., Eisinger, M., Hahne, A., and Lefebvre, A.: GOME-2 – Metop's second generation sensor for operational ozone monitoring, ESA Bull.-Eur. Space, 102, 28–36, 2000.

Cariolle, D. and Déqué, M.: Southern hemisphere medium-scale waves and total ozone disturbances in a spectral general circulation model. J. Geophys. Res., 91D, 10825–10846, 1986.

Cariolle, D. and Teyssèdre, H.: A revised linear ozone photochemistry parameterization for use in transport and general circulation models: multi-annual simulations. Atmos. Chem. and Phys. Disc., 7, 1655–1697, 2007.

Charney, J. C. and Phillips, N. A.: Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. J. Meteor., 10, 17–29, 1953.

Charron, M., Polavarapu, S., Buehner, M., Vaillancourt, P. A., Charette, C., Roch, M., Morneau, J., Garand, L., Aparacio, J. M., MacPherson, S., Pellerin, S., St-James, J., and Heilliette, S.: The stratospheric extension of the Canadian Global Deterministic Medium-Range Weather Forecasting System and its impact on tropospheric forecasts, Mon. Wea. Rev., 140, 1924–1944, doi:10.1175/MWR-D-11-00097.1, 2012.

Coldewey-Egbers, M., Loyola, D. G., Koukouli, M., Balis, D., Lambert, J.-C., Verhoelst, T., Granville, J., Van Roozendael, M., Lerot, C., Spurr, R., Frith, S., and Zehner, C.: The GOME-type Total Ozone Essential Climate Variable (GTO-ECV) data record from the ESA Climate Change Initiative, Atmos. Meas. Tech. Discuss., 8, 3923-3940, doi:10.5194/amt-8-3923-3940-2015, 2015.

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. Mon. Wea. Rev., 126, 1373–1395, doi:10.1175/1520-0493(1998)126,1373:TOCMGE.2.0.CO;2, 1998a.

Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A.: The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part II: Results. Mon. Wea. Rev., 126, 1397–1418, doi:10.1175/1520-0493(1998)126,1397:TOCMGE.2.0.CO;2, 1998b.

de Grandpré, J., Ménard, R., Rochon, Y. J., Charette, C., Chabrillat, S., and Robichaud, A. : Radiative impact of ozone on temperature predictability in a coupled chemistry–dynamics data assimilation system, Mon. Wea. Rev., 137, 679-692, doi:10.1175/2008MWR2572.1, 2009.

de Grandpré, J., Tanguay, M., Qaddouri, A., Zerroukat, M. and McLinden, C. A.: Semi-Lagrangian Advection of Stratospheric Ozone on a Yin-Yang Grid System, Monthly Weather Review, 144, 1035--1050, doi:10.1175/MWR-D-15-0142.1, 2016.

Dee, D. P.: Bias and data assimilation, Q. J. R. Meteorol. Soc., 131(613), 3323-3343, 2005.

Dee D. P.: Importance of satellites for stratospheric data assimilation, In Proceedings of the ECMWF Seminar on Recent developments in the use of satellite observations in numerical weather prediction, 3–7 September 2007. ECMWF: Reading, UK, 2008.

Dee, D. P. and Uppala, S.: Variational bias correction of satellite radiance data in the ERA-Interim reanalysis, Q. J. R. Meteorol. Soc., 135(644), 1830-1841, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S.B., Hersbach, H., Hòlm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart. F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. doi:10.1002/qj.828, 2011.

~~Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background, and analysis-error statistics in observation space. Q. J. R. Meteorol. Soc., 131, 3385-3396, 2005.~~

Dittman, M. G., Ramberg, E., Chrisp, M., Rodriguez, J. V., Sparks, A. L., Zaun, N. H., Hendershot, P., Dixon, T., Philbrick, R. H., and Wasinger, D.: Nadir ultraviolet imaging spectrometer for the NPOESS Ozone Mapping and Profiler Suite (OMPS), Proc. SPIE 4814, Earth Observing Systems VII, 25 September 2002, Seattle, WA, edited by Barnes, W. L., 111–119, 2002a.

Dittman, M. G., Leitch, J. Chrisp, M., Rodriguez, J. V., Sparks, A., McComas, B., Zaun, N., Frazier, D., Dixon, T., Philbrick, R., and Wasinger D.: Limb broad-band imaging spectrometer for the NPOESS Ozone Mapping and Profiler Suite (OMPS), Proc. SPIE 4814, Earth Observing Systems VII; 25 September 2002, Seattle, WA, edited by: Barnes, W. L., 120–130, 2002b.

Dragani, R. and. Dee D. P.: Progress in ozone monitoring and assimilation, ECMWF Newsletter 116, 35–42. 2008

Dupuy E., and Coauthors: Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), Atmos. Chem. Phys., 9, 287–343, 2009.

Durbin, P., Tilmes, C., Duggan, B., and Das B.: OMI near real time data processing, 2010 IEEE Intern. Geoscience and Remote Sensing Symp., Honolulu, Hawaii, USA, 586-588, July 25-30, 2010.

Evans, R., McConville, G., Oltmans, S., Petropavlovskikh, I., and Quincy, D.: Measurement of internal stray light within 30 Dobson ozone spectrophotometers, Int. J. Remote Sens., 30(15), 4247–4258, doi:10.1080/01431160902825057, 2009.

Fioletov, V. E., Kerr, J. B., Hare, E. W., Labow, G.J., and McPeters, R.D.: An assessment of the world ground-based total ozone network performance from the comparison with satellite data, J. Geophys. Res., 104, 1737–1747, 1999

Fioletov, V. ~~E., Kerr, J. B., McElroy, C. T., Wardle, D. I., Savastiouk, V., and Grajnar, T. S.: The Brewer reference triad, Geophys. Res. Lett., 32(20), 2005.~~

~~Fioletov, V.~~ E, Labow, G., Evans, R., Hare, E. W., Köhler, U., McElroy, C. T., Miyagawa, K., Redondas, A., Savastiouk, V., Shalamyansky, A. M., Staehelin, J., Vanicek, K., and Weber, M.: Performance of the ground-based total ozone network assessed using satellite data, J. Geophys, Res., 113, D14313, 19 pp, doi:10.1029/2008JD009809, 2008.

Fisher, M., and Andersson, E.: Developments in 4-D-Var and Kalman filtering, in: Technical Memorandum Research Department, 347, ECMWF, Reading, UK, 2001.

Flynn, L., Seftor, C., Larsen, J., and Xu, P.: Introduction to the Ozone Mapping and Profiler Suite (OMPS), in Earth Science Satellite Remote Sensing Vol.1: Science and Instruments, edited by J. J. Qu et al., Springer, Berlin, ISBN: 978-3-540-35606-6, 2006. [Also http://npp.gsfc.nasa.gov/omps.html.]

Flynn, L. (editor): Solar Backscatter Ultraviolet Instrument (SBUV/2) Version 8 Ozone Retrieval Algorithm Theoretical Basis Document (V8 ATBD), 45 pp., 2 February 2007 [Available from: https://ozoneaq.gsfc.nasa.gov/media/docs/SBUV2_V8_ATBD_020207.pdf]

Flynn, L. E., McNamara, D., Beck, C.T., Petropavlovskikh, I., Beach, E., Pachepsky, Y., Li, Y. P., Deland , M., Huang, L.-K., Long ,C. S., Tiruchirapalli R., and Taylor, S.: Measurements and products from the Solar Backscatter Ultraviolet (SBUV/2) and Ozone Mapping and Profiler Suite (OMPS) instruments, International Journal of Remote Sensing, 30,4259-4272, 2009.

Flynn, L., Long, C., Wu, X., Evans, R., Beck, C. T., Petropavlovskikh, I., McConville, G., Yu, W., Zhang, Z., Niu, J., Beach, E., Hao, Y., Pan, C., Sen, B., Novicki, M., Zhou, S., and Sefton, C.: Performance of the ozone mapping and profiler suite (OMPS) products, J. Geophys. Res., 119(10), 6181-6195, doi: 10.1002/2013JD020467, 2014.

Fortuin, J. P., and Kelder, H.: An ozone climatology based on ozonsonde and satellite measurements, J. Geophys. Res., 103, 31 709-31 734, 1998.

Garane, K., Lerot, C., Coldewey-Egbers, M., Verhoelst, T., Elissavet Koukouli, M., Zyrichidou, I., Balis, D., S., Danckaert, T., Goutail, F., Granville, J., Hubert, D., Keppens, A., Lambert, J.-C., Loyola, D., Pommereau, J.-P., Van Roozendael, M., and Claus Zehner, C.: Quality assessment of the Ozone_cci Climate Research Data Package (release 2017) – Part 1: Ground-based validation of total ozone column data products, Atmos. Froidevaux, L., and Coauthors: Validation of Aura Microwave Limb Sounder stratospheric ozone measurements, J. Geophys., Res., 113, D15S20, doi:10.1029/2007JD008771, 24 pp., 2008.

Gaspari, G., and Cohn, S.: Construction of correlation functions in two and three dimensions, Q. J. R. Meteorol. Soc., 125, 723-757, 1995.

Meas. Tech., 11, 1385-1402, doi:10.5194/amt-11-1385-2018, 2018.

Gauthier, P., Chouinard, C., and Brasnett, B.: Quality control: Methodology and applications. In Data Assimilation for the Earth System, R. Swinbank et al. (eds.), NATO Science Series IV: Earth and Environmental Science Vol. 26, Kluwer Academic Publishers, 177-187, 2003.

Girard, C., Plante, A., Desgagné, M., McTaggart-Cowan, R., Cote, J., Charron, M., Gravel, S., Lee, V., Patoine, A., Qaddouri, A., Roch, M., Spacek, L., Tanguay, M., Vaillancourt, P., Zadra, A.: Staggered vertical discretization of the Canadian Environmental Multiscale (GEM), Model using a coordinate of the log-hydrostatic-pressure type, Mon. Wea. Rev. 142, 1183-1196, doi: 10.1175/MWR-D-13-00255.1, 2014

GOME User Manual: Product User Manual for GOME Total Columns of Ozone, NO2, tropospheric NO2, BrO, SO2, H2O, HCHO, OClO, and Cloud Properties (O3M-SAF OTO and NTO), Doc. No. DLR/GOME/PUM/01, Issue 2/E, Deutsches

Zentrum für Luft und Raumfahrt e.V. – DLR, Oberpfaffenhofen, Germany, 47 pp., 8 August 2012. [Available from: https://earth.esa.int/c/document_library/get_file?uuid=1bca1fc9-0525-4e67-b211-b23141ce83b7&groupId=10174.]

GOME-2 ATBD: Algorithm Theoretical Basis Document for GOME-2 Total Column Products of Ozone, NO2, BrO, HCHO, SO2, H2O and Cloud Properties (GDP 4.8 for O3M-SAF OTO and NTO), Doc. No. DLR/GOME-2/ATBD/01, Issue 3/A, Deutsches Zentrum für Luft und Raumfahrt e.V. – DLR, Oberpfaffenhofen, Germany, 58 pp., 30 March 2015. [Available from: https://wdc.dlr.de/sensors/gome2/DLR_GOME-2_ATBD_2A.pdf.]

Hao, N., Koukouli, M. E., Inness, A., Valks, P., Loyola, D. G., Zimmer, W., Balis, D. S., Zyrichidou, I., Van Roozendael, M., Lerot, C., and Spurr, R. J. D.: GOME-2 total ozone columns from MetOp-A/MetOp-B and assimilation in the MACC system, Atmos. Meas. Tech., 7(9), 2937-2951, 2014.

Hassinen, S., Balis, D., Bauer, H., Begoin, M., Delcloo, A., Eleftheratos, K., Gimeno Garcia, S., Granville, J., Grossi, M., Hao, N., Hedelt, P., Hendrick, F., Hess, M., Heue, K.-P., Hovila, Jønch-Sørensen, J., H., Kalakoski, N., Kauppi, A., S. Kiemle, S., Kins, L., Koukouli, M. E., Kujanpää., J., Lambert6, J.-C., Lang. R., Lerot, C., Loyola, D., Pedergnana, M., Pinardi, G., Romahn, F., van Roozendael, M., Lutz, R., De Smedt, I., Stammes, P., Steinbrecht, W., Tamminen, J., Theys, N., Tilstra, L. G., Tuinder, O. N. E., Valks, P., Zerefos, C., Zimmer, W., and Zyrichidou I.: Overview of the O3M SAF GOME-2 operational atmospheric composition and UV radiation data products and data availability, Atmos. Meas. Tech., 9, 383-407, doi:10.5194/amt-9-383-2016, 2016.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N.,Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team: The MACC reanalysis: an 8-yr data set of atmospheric composition, Atmos. Chem. Phys., 13, 407304109, doi:10.5194/acp-13-4073-2013, 2013.

IFS Documentation CY41R1: Part I: Observation, book chapter, European Centre for Medium-Range Weather Forecasts,72 pp., 2015. [Available from: http://www.ecmwf.int/sites/default/files/elibrary/2015/9208-part-i-observation-processing.pdf.]]

Jackson, D.R., Keil, M., and Devenish, B.J.: Use of Canadian Quick covariances in the Met Office data assimilation system, Q. J. R. Meteorol. Soc., DOI:10.1002/qj, 2008.

Kerr, J. B.: New methodology for deriving total ozone and other atmospheric variables from Brewer spectrophotometer direct sun spectra, J. Geophys., Res., 107(D23), 4731-4747, doi:10.1029/2001JD001227, 2002.

Komhyr, W. D., Mateer, C. L., Hudson, R. D.: Effective Bass-Paur 1985 ozone absorption coefficients for use with Dobson ozone spectrophotometers, J. Geophys. Res., 98(D11), 20451-20465, 1993.

Koukouli, M., Balis, D., Loyola, D., Valks, P., Zimmer, W., Hao, N., Lambert, J. -C., Van Roozendael, M., Lerot, C., and Spurr, R. J. D.: Geophysical validation and long-term consistency between GOME-2/MetOp-A total ozone column and measurements from the sensors GOME/ERS-2, SCIAMACHY/ENVISAT and OMI/Aura. Atmos. Meas. Tech., 5(5), 2169-2181, 2012.

Formatted: English (United Kingdom)

Koukouli, M. E., Lerot, C., Granville, J., Goutail, F., Lambert, J.-C., Pommereau, J.-P., Balis, D., Zyrichidou, I., Van Roozendael, M., Coldewey-Egbers, M., Loyola, D., Labow, G., Frith, S., Spurr, R., and Zehner, C.: Evaluating a new homogeneous total ozone climate data record from GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, J. Geophys. Res. Atmos., 120, 12,296-12,312, doi:10.1002/ 2015JD023699, 2015.

Koukouli, M., Zara, M., Lerot, C., Fragkos, K., Balis, D., van Roozendael, M., Allart, M., and Rvan, A. R.D: The impact of the ozone effective temperature on satellite validation using the Dobson spectrophotometer network, Atmos. Meas. Tech. 9, 5, 2055-2065, doi:10.5194/amt-9-2055-2016, 2016.

Kroon, M., Veefkind, J. P., Sneep, M., McPeters, R. D., Bhartia, P. K., and Levelt, P. F.: Comparing OMI-TOMS and OMI-DOAS total ozone column data, J. Geophys. Res., 113(D16S28), doi:10.1029/2007JD008798, 2008.

Labow, G. J., McPeters, R. D.,, Bhartia, P. K., and Kramarova, N.: A comparison of 40 years of SBUV measurements of column ozone with data from the Dobson/Brewer network, J. Geophys., Res. , 118, 7370–7378, doi:10.1002/jgrd.50503, 2013

Lahoz, W., and Errera, Q: Constituent Assimilation, in Data Assimilation: Making Sense of Observations with eds W, Lahoz, B. Khattatov and R. Ménard, Springer-Verlag Berlin Heidelberg, 449-490, doi:10.1007/978-3-540-74703-1_18, 2010.

Levelt, P. F., van den Oord, G. H. ., Dobber, M R., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lerot, C., Van Roozendael, M., Spurr, R., Loyola, D, Coldewey-Egbers, M., Kochenova, S., van Gent, J., Koukouli, M., Balis, D., Lambert, J.-C., Granville, J., and Zehner, C.: Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, J. Geophys. Res. Atmos., 119, 1639–1662, doi:10.1002/2013JD020831, 2014.

Lundell, J.O. V., and Saari, H.: The Ozone Monitoring Instrument, IEEE Trans. Geosci. Remote Sens., 44, 1093–1101, 2006.

Livesey, N. J., Van Snyder, W., Read, W. J., and Wagner, P. A.: Retrieval algorithms for the EOS Microwave Limb Sounder (MLS), IEEE Trans. Geosci. Remote Sens., 44, 1144–1155, doi:10.1109/TGRS.2006.872327, 2006.

Livesey, N. J., and Coauthors: Earth Observing System (EOS) Aura Microwave Limb Sounder (MLS): Version 3.3 and 3.4 Level 2 data quality and description document, JPL D-33509, Version 3.3x/3.4x-1.1, Jet Propulsion Laboratory, 158 pp., May 14, 2013. [Available from: http://mls.jpl.nasa.gov/data/v3_data_quality_document.pdf.]

Levelt, P. F., Joiner, J., Tamminen, J., Veefkin, J. P., Bhartia, P. K., Stein Zweers, D. C., Duncan, B., N., Streets, D. G. Eskes, H., van der A, R, McLinden, C., Fioletov, V., Carn, S., de Latt, J. DeLand, M., Marchenko, S., McPeters, R., Ziemke, J., Fu, D., Liu, X., Pickering, K., Apituley, A., González Abad, G., Arola, A., Boersma, S., Chan Miller, C., Chance, K., de Graaf, M., Hakkarainen, J., Hassinen, S., Ialongo, I., Kelipool, Q., Krotkov, N., Li, C., Lamsal, L., Newman, P., Nowlan, C., Suleiman, R., Gijsbert Tilstra, L., Torres, O., Wang., H., andf Wargan, K.: The Ozone Monitoring Instrument: overview of 14 years in space, Atmos. Meas. Phys., 18, 5699-5745, doi:10.5194/acp-18-5699-2018, 2018.

Loyola, D. G., Koukouli, M. E., Valks, P., Balis, D. S., Hao, N., Van Roozendael, M., Spurr, R. J. D., Zimmer, W., Kiemle, S., Lerot, C., and Lambert, J. -C.: The GOME-2 total column ozone product: Retrieval algorithm and ground-based validation, J. Geophys. Res., 116(D7), 2011.

Formatted: English (United Kingdom)

MACC-II Final Report: MACC-II Monitoring Atmospheric Composition and Climate - Interim Implementation, project funded from the European Union's Framework Programme, 137 pp., October, 2014. [Available from: http://atmosphere.copernicus.eu/sites/default/files/repository/MACCII_FinalReport_0.pdf.]]

Massart, S., Clerbaux, C., Cariolle, D., Piacentini, A., Turquety, S., and Hadji-Lazaro, J.: First steps towards the assimilation of IASI ozone data into the MOCAGE-PALM system, Atmos. Chem. Phys., 9, 5073-5091, 2009.

McLinden, C. A., Olson, S. C., Hannegan, B., Wild, O., Prather, M. J., and Sundet, J.: Stratospheric ozone in 3-D models: A simple chemistry and the cross-tropopause flux, J. Geophys. Res., 105, 14653–14665, 2000.

McPeters, R, D,, Frith, S., and Labow, G. J.: OMI total column ozone: extending the long-term data record, Atmos. Meas. Tech., 8, 4845–4850, doi:10.5194/amt-8-4845-2015, 2015.

McPeters, R. D., Bhartia, P. K., Haffner, D., Labow, G. J., and Flynn, L.: The version 8.6 SBUV ozone data record: An overview, J. Geophys. Res., 118., 1–8, doi:10.1002/jgrd.50597, 2013.

McPeters, R., Kroon, M., Labow, G., Brinksma, E., Balis, D., Petropavlovskikh, I., Veefkind, J. P., Bhartia, P. K., and Levelt, P. F.: Validation of the AURA Ozone Monitoring Instrument total column ozone product, J. Geophys. Res., 113(D15), 2008.

Moeini, A., Vaziri, Z., McElroy, C. T., Tarasick, D. W., Evans, R. D., Petropavlovskikh, I., and Feng, K.-H.: The effect of instrumental stray light on Brewer and Dobson total ozone measurements, Atmos. Meas. Tech. Discuss., 29 pp., doi:10.5194/amt-2018-2, 2018.

Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski, M., Holdak, A., Kokhanovsky, A., Livschitz, J., and Eisinger, M.: The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing – an overview, Atmos. Meas. Tech., 9, 1279-1301, doi: 10.5194/amt-9-1279-2016, 2016Eisinger, M., Anderson, C., Callies, J., Corpaccioli, E., Lang, R., Lefebvre, A., Livschitz, Y., and Albinana, A. P.: GOME-2 on MetOp, Proc. of The 2006 EUMETSAT Meteorological Satellite Conference, Helsinki, Finland, 2006.

Ozone Monitoring Instrument (OMI) Near Real Time Data User's Guide, 2010. [Available from: https://ozoneaqmacuv.gsfc.nasa.gov/media/docs/OMI-NRT-DUG.pdf.]]

Ozone Monitoring Instrument (OMI) Data User's Guide, produced by the OMI Team, 6266 pp.,January 5, 2012. [Available from: https://acdisc.gesdisc.eosdishttp://disc.sci.gsfc.nasa.gov/data/s4pa/Aura_OMI_Level2G/OMTO3G.003/doc/additional/documentation/README.OMI_DUG.pdf.]

Polavarapu, S., Ren, S., Rochon, Y., Sankey, D., Ek, N., Koshyk, J., and Tarasick, D.: Data assimilation with the Canadian Middle Atmosphere Model, Atmos.-Ocean, 43. 77-100. 2005.

Redondas, A., Evans, R., Stuebi, R., Köhler, U., and Weber, M.: Evaluation of the use of five laboratory-determined ozone absorption cross sections in Brewer and Dobson retrieval algorithms, Atmos. Chem. Phys., 14, 1635-1648, doi:10.5194/acp-14-1635-2014, 2014.

Formatted: Font: (Default) Times New Roman

Smit, M. A. H., and Sträter, W.: Jülich ozone sonde intercomparisons 2000 (JOSIE), Report No. 1225, World Meteorol. Org., Global Weather Watch, Geneva, 2004. [Available online at http://library.wmo.int/pmb_ged/wmo-td_1225.pdf]

SPARC: Stratospheric Processes And their Role in Climate (SPARC), Assessment of trends in the vertical distribution of ozone, SPARC report No. 1, in: WMO Global Ozone Res. and Monit. Project Rep. 43, edited by: Harris, N., Hudson, R., and Phillips, C., World Meteorol. Org., Geneva, 1998.

Staehelin, J., Kerr, J., Evans R., and Vanicek, K.: Comparison of total measurements of Dobson and Brewer spectrometers and recommended transfer functions, Report No. 149, Meteorol. Org., Global Atmosphere Watch, Geneva, 2003.

Steinbrecht, W., Hegglin, M. I., Harris, N., and Weber, M.: Is global ozone recovering?, C. R. Geoscience, 350, 368-375, doi:10.1016/j.crte.2018.07.012, 2018.

Struthers, H., Brugge, R., Lahoz, W. A., O'Neil, A., and Swinbank, R.: Assimilation of ozone profiles and total column measurements into a global general circulation model, J. Geophys., Res., 10(D20), 4438-4451, doi:10.1029/2001JD000957, 2002

Tereszchuk, K. A., Rochon, Y. J., McLinden, C. A., and Vaillancourt, P. A.: Optimizing UV Index determination from broadband irradiances, Geosci. Model Dev., 11, 1093–1113, doi:10.5194/gmd-11-1093-2018, 2018.

van der A, R. J., Allaart, M. A. F., and Eskes, H. J.: Extended and refined multi sensor reanalysis of total ozone for the period 1970–2012, Atmos. Meas. Tech., 8(7), 3021-3035, 2015.

van der A, R. J., Allaart, M. A. F., and Eskes, H. J.: Multi sensor reanalysis of total ozone, Atmos. Chem. Phys., 10(22), 11277-11294, 2010.

van Roozendael, M., Peeters, P., Roscoe, H. K., De Backer, H., Jones, A. E., Bartlett, L., Vaughan, G., Goutail, F., Pommereau, J.-P., Kyro, E., Wahlstrom, C., Braathen, G., and Simon, P. C.: Validation of ground-based visible measurements of total ozone by comparison with Dobson and Brewer spectrophotometers, J. Atmos. Chem., 29, 55-83, 1998.

Van Weele, M., Dameris, M., Braesicke, P., and van der A, R.: User Requirements Document Ozone_cci, Issue 3, Ozone_cci Phase-2 (Project Deliverable D1.1), ESA-ESRIN, Reference: Ozone_cci_URD_3, Ozone-CCI, 2016, {Available from: http://www.esa-ozone-cci.org/?q=webfm_send/175, 2017.]

Veefkind, J. P. and de Haan, J. F.: DOAS Total O3 Algorithm, Chapter 3 in OMI Algorithm Theoretical Basis Document, Volume, 2, OMI Ozone Products, edited by P.K. Bhartia, Version 2.0, 33-50, 2002.

Veefkind, J. P., de Haan, J. F., Brinksma, E. J., Kroon, M., and Levelt, P. F.: Total ozone from the Ozone Monitoring Instrument (OMI) using the DOAS technique, IEEE Trans. Geosci. Remote Sens., 44(5), 1239-1244, 2006.

Viatte, C., Schneider, M., Redondas, A., Hase, F., Eremenko, M., Chelin, P., Flaud, J.M., Blumenstock, T., and Orphal, J.: Comparison of ground-based FTIR and Brewer O 3 total column with data from two different IASI algorithms and from OMI and GOME-2 satellite instruments, Atmos. Meas. Tech., 4(3), 535-546, 2011.

Waters, J. W., and Coauthors: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite, IEEE Trans. Geosci. Remote Sens., 44, 1075–1092, doi:10.1109/TGRS.2006.873771, 2006.

WMO (World Meteorological Organization): Manual on the global data processing system, Volume I (Annex IV to the WMO Technical Regulations), Global Aspects, Report WMO-No. 485, Secretariat of the World Meteorological Organization, Geneva, Switzerland, 1992.

Zadra, A., McTaggart-Cowan, R., Vaillancourt, P. A., Roch, M.Bélair, S., and Leduc,A.-M.: Evaluation of tropical cyclones in the Canadian Global Modeling System: Sensitivity to moist process parameterization. Mon. Wea. Rev., 142, 1197–1220, doi:10.1175/MWR-D-13-00124.1, 2014a.

Zadra, A., Antonopoulos, S., Archambault, B., Beaulne, A., Bois, N., Buehner, M., Giguère, A., Marcoux, J., Petrucci, F., Poulin, L., Reszka, M., Robinson, T., St-James, J., and Rahill, A.: Improvements to the Global Deterministic Prediction system (GDPS) (from version 2.2.2 to 3.0.0), and related changes to the Regional Deterministic Prediction System (RDPS) (from version 3.0.0 to 3.1.0). Canadian Meteorological Centre Tech. Note, 88 pp., 2014b [Available from:online at http://collaboration.cmc.ec.gc.ca/cmc/CMOI/product_guide/docs/lib/op_systems/doc_opchanges/technote_gdps300_20130213_e.pdf.].]

Zhand, Z., and Kasheta, Z.: Global Ozone Monitoring Experiment-2 (GOME-2) Operational Ozone Product System Version 8: Interface Control Documnent (Documentation Version 1.0), June 2007, prepared for the National Environmental Satellite, Data, and Information Service (NESDIS), revision of July 2009.

Zhou, L., Divakarla, M., and Liu, X.: An overview of the Joint Polar Satellite System (JPSS) science data product calibration and validation, Remote Sens., 8, 139, 13 pp, doi:10.3390/rs8020139, 2016.

Table captions

Table 1. Regional and global relative mean differences (%) of total column ozone between OMI-TOMS and the specified ground-based instument types over July-August 2014/2015 and January-February 2015. The averaging excludes stations having outlier station mean diffferences for each period (see Supplement tables S1 to S3 and the text of Section 3.1) except for the two rows for the latitude region 60-90° S as described in the text. The standard deviations (S.D.) are for the inter-station variation of the station mean differences about the regional or global mean differences. Unavailable S.D. values for available mean differences imply the presence of only one station. The Dobson total column ozone measurements for the two July-August periods were adjusted as a function of the ozone effective temperature (see Section 2.2.4); those for the January-February period were not adjusted in the absence of the ozone effective temperature for the period. The impacts of the Dobson July-August period corrections on the global mean differences were reductions between 0.0 and 0.4 %.

Table 2. Global diagnostics of differences in total column ozone between satellite instuments and OMI-TOMS for July-August 2014 and January-February 2015. The diagnostics consists of global mean differences and percentages of non-empty SZA/latitude bins with mean differences exceeding 2 % in magnitude.

Table 3. Mean differences of the total column ozone (%) between satellite instuments and OMI-TOMS for July-August 2014 and January-February 2015 for Northern and Southerm Hemispheres, for solar zenith angles below and above 70°.

Table 4. Mean differences in total column ozone (%) between satellite instuments and OMI-TOMS for July-August 2014 using the options (a), (b), and (c) from Section 4.2, for Northern and Southerm Hemispheres and solar zenith angle below and above 70°.

Table 5. List of assimilation experiments and their corresponding identifiers. In the second column, an asterisk (*) next to the instrument denotes that the bias-corrected observations (using the colocation method of Section 4.1) were assimilated.

Figure captions

Figure 1. Updated set of applied observation errror standard deviation estimates for GOME-2A, GOME-2B, OMPS-NM and OMI total column measurements as a function of solar zenith angle band centers, for the combined period of July and August 2014. A 2% constant served in specifying the first set of applied error standard deviation estimates.

Figure 2. Original and updated (final) set of total column background error standard deviations obtained from projecting the applied latitude and vertically dependent background ozone error variances and the globally homogeneous background vertical error correlations to the total ozone observation space as a function of latitude for July and August.

Figure 3. Mean total column ozone differences between GOME-2A/B, OMPS-NM, and colocated OMI-TOMS data as a function of solar zenith angle for the Northern and Sourthern Hemispheres for August 2014. Differences were computed from observation colocations. For the GOME-2 instruments, separate continuous difference regions are specified about 70° (see text in Section 2.4.1 for additional information). Constant difference extrapolation is applied from bin midpoints at the edges of distinct regions.

Figure 4. Mean total column ozone differences between OMI-TOMS and Brewer, Dobson, and filter ozonometer measurements over July-August 2014.

Figure 5. Mean total column ozone differences between GOME-2A/B, OMPS-NM/NP and colocated OMI-TOMS data for the period of July-August 2014.

Figure 6. Same as Fig. 6 for January-February 2015.

Figure 7. Time series of total column ozone bias corrections for July and August 2014 for GOME-2A/B, and OMPS-NM/NP. Dashed lines show individual six-hour mean differences with OMI-TOMS, while the solid curves of the same colour show the two -week moving average bias corrections. The particular (latitude, solar zenith angle) bins plotted are (a) 5° wide bins centred on (52.5°N, 37.5°) for GOME-2A/B and OMPS-NM and a 10° wide bin centred on (55°N, 35°) for OMPS-NP and (b) 5° wide bins centred on (62.5°N, 42.5°) for GOME-2A/B

and OMPS-NM and a 10° wide bin centred on (65°N, 45°) for OMPS-NP. Time coverage for individual bins do not necessarily cover complete months.

Figure 8. Mean total column ozone differences between GOME-2A, OMPS-NM and colocated OMI-TOMS data as a function of ozone effective temperature and solar zenith angle for the periods of July-August 2014 and July-August 2015.

Figure 9. Residual average total column ozone differences between GOME-2A, OMPS-NM and colocated OMI-TOMS data as a function of latitude and solar zenith angle for July-August 2014 and July-August 2015 following bias correction as a function of ozone effective temperature and solar zenith angle.

Figure 10. Time series of total column ozone bias corrections for two latitude/SZA bins covering July-August 2014 for GOME-2A using different bias correction methods. All cases that include colocation methods usethinned observation sets. The 'O-F' curves additionally use the differences of forecasts described in Section 2.4.2 following the assimilation of OMI-TOMS. The 'colocations alone' and 'O-F' curves were calculuated using the Gaussian two-week moving average with HWHM of 4.7 days. The 'T_{eff}/SZA' curve is the result of mapping each observation that falls within the latitude/SZA bin onto the ozone effective temperature/SZA bias for July-August 2014 (shown in Fig. 8) and taking the mean of these values.

Figure 11. Time mean total column ozone biases between GOME-2A and OMI-TOMS for the period of July-August 2014 from colocation alone and approaches (a), (b), and (c) using observation-minus-forecast differences (see Section 2.4.2). For approaches (a), (b), and (c) the forecasts were taken from the 'OMI' assimilation run (see **Table** 2). The bias in the 'colocations alone' panel was computed **6** using the thinned observation data set as the thinned data set was used in the assimilation.

Figure 12. Global mean differences (%) between Brewer and Dobson total column ozone measurements and short-term forecasts for July-August 2014. Bias-correctedThe instuments assimilated for each run are shown on the horizontal axis and the unshaded (shaded) bars indicate that all observations from were (were not) bias corrected. For assimilations that assimilated bias-corrected observations, the colocated observation bias correction scheme (Section 4.1) were applied in the assimilations.was used. The Dobson measurements used were adjusted as a function of the ozone effective temperature (see Section 2.2.4). The uncertaintieserror bars denote the standard errorsquare root of the sample variance of the mean differencesdifference or difference standard deviations of the differencesdeviation. The data from the two Antarctic stations have been included here even though their mean differences with OMI are outliers relative to most mean differencesdiffernces (Tables S1 and S2).

Formatted: Font: Bold, Check spelling and grammar

Formatted: Left, Space After: 0 pt, Line spacing: single

Formatted: Font: Not Bold, Check spelling and grammar

Figure captions

Figure 113. Mean total column ozone differences (%) between OMI-TOMS and Brewer, Dobson, and filter ozonometer measurements overand short-term forecasts as a function of spatial location for July-August 2014. The colours blue to purple denote negative differences andplot titles indicate the colours yellow to red refer to positive differences.

Figure 2. Mean total column ozone differences (%) between GOME-2A/B, OMPS-NM/NP, SBUV/2-TC/NP and colocated OMI-TOMS data for the period of July-August 2014. The SBUV/2-TC total column ozone values stem from the two wavelength retrieval, while those for SBUV/2-NP are the sums of the retrieved 21-layer partial columns. The colours blue to purple denote negative differences and the colours yellow to red refer to positive differences.
Figure 3. Same as Fig. 2 for January-February 2015.

Formatted: Justified

Figure 4. Time series of total column ozone bias corrections (DU) for July and August 2014 for GOME-2A/B, OMPS-NM/NP, and SBUV/2-TC/NP as derived from the colocation method described in Section 4.1. Dashed vertical lines show individual six-hour mean differences with OMI-TOMS, while the solid curves of the same colour show the two-week moving average bias corrections. The particular (latitude, solar zenith angle) bins plotted are 5° wide bins centred on (52.5°N, 37.5°) for GOME-2A/B and OMPS-NM and a 10° wide bin centred on (55°N, 35°) for OMPS-NP and SBUV/2-TC/NP. Time coverage for individual bins do not necessarily cover complete months.

Figure 5. Time mean total column ozone biases (%) between GOME-2A and OMI-TOMS for July-August 2014 from colocation alone and for the options (a), (b), and (c) of Section 4.2 that use observation-minus-forecast differences. For options (a), (b), and (c), the forecasts were taken from the 'OMI' assimilation run (see Table 52 for description). In the case of no). The bias in the 'colocations alone' panel was computed using the thinned observation data set to compare to the other cases that use thinned observations. The colours blue to purple denote negative differences and the colours yellow to red refer to positive differences.

Figure 6. Time series of total column ozone bias corrections (DU) for two latitude/SZA bins covering July-August 2014 for GOME-2A using different bias correction methods. All cases that include colocation methods use thinned observation sets. The 'O-F' curves additionally use the differences of forecasts described in Section 4.2 following the assimilation of OMI-TOMS. The 'colocations alone' and 'O-F' curves were calculuated using the Gaussian two-week moving average with HWHM of 4.7 days. The 'T_{eff}/SZA' curves, described in Section 4.3, result from mapping each observation that falls within the latitude/SZA bin onto the ozone effective temperature/SZA bias estimate for July-August 2014 (shown in Fig. 7), followed by taking the average of these bias estimate , many of the values south of 60º S exceed the lower limit of the colour bar, in some instances with values for each time.

Figure 7. Mean total column ozone differences (%) between GOME-2A, OMPS-NM and colocated OMI-TOMS dataas low as a function of ozone effective temperature (degrees Kelvin) and solar zenith angle (degrees) for the periods of July-August 2014 and July-August 2015. The colours blue to purple denote negative differences and the colours yellow to red refer to positive differences. -30%.

Figure 814. Zonal mean total column ozone statistics of mean differences (%),, standard deviations (%),, and anomaly correlation coefficients (ACC; unitless) as a function of latitude (degrees) for the comparison between OMI-TOMS measurements and short-term forecasts for July-August 2014. The legends in the top plots indicate the assimilation runs (see Table 52 for description) and apply to all plots in the same column.

Figure 915. Zonal mean total column ozone differences (%) and anomaly correlaton coefficients (ACC) between OMI-TOMS observations and short-term forecasts for July-August 2014 showing the effect of using the updated sets of observation and background error variances in the assimilation. The legend indicates the assimilation run (see Table 2 for description).

Figure 16. Zonal mean differences and anomaly correlation coefficients (unitless) for total coumn ozone between OMI-TOMS observations and short-term forecasts -as a function of time (date). Results are shown for the case without assimilation as well as with the assimilation of OMI, GOME-2A/B, and OMPS-NM (both with and without bias correction). The legend indicates the assimilation run (see Table 3 for description).. Results are shown for the case without assimilation as well as with the assimilation of OMI, GOME-2A/B, and OMPS-NM (both with and without bias correction). The legend indicates the assimilation run (see Table 2 for description). Each value plotted was calculated using a 24 hour time window.

Figure 10. Zonal mean**17.** Mean differences, standard derivations (DU)diviations of the differences, and anomaly correlation coefficients (ACC) between MLS observations and short-term forecasts F, OMI-TOMS observations O, and climatological values C over July-August 2014. For the short-term forecasts,The legend indicates the assimilation run from which the forecast was launched is indicated in the subscript, with the labels described in(see Table 5.2 for description).

Figure 18. Similar to Fig. 17 but only covering 1-23 July 2014. This shorter period stems from the 'MLS+OMI' assimilation having been conducted only for this duration prior to computing system changes.

Figure 19. Global mean and standard deviation differences, and anomaly correlation coeffients between ozonesonde observations and short-term forecasts over July-August 2014. The legend indicates the assimilation run (see Table 2 for description).

Figure 20. Similar to Fig. 19 but over 1-23 July 2014. For this specific figure, the comparison to ozonesondes for the 'MLS+OMI' assimilation case was done using six-hour forecasts instead of nearest forecasts in the three to nine hour forecast periods at intervals of 45 minutes as the latter were not available for that comparison. This by itself would have deteriorated the results shown for 'MLS+OMI' which are still better that the other cases.

Formatted: Space After: 0 pt, Line spacing: single

Formatted: Font: Not Bold, Check spelling and grammar

Formatted: Justified

Formatted: Font: 9 pt, English (Canada)

Formatted: Font: 10 pt, Bold, English (United States), Check spelling and grammar