The response to the Referees shall be structured in a clear and easy-to-follow sequence: (1) comments from Referees, (2) author's response, (3) author's changes in manuscript.

Anonymous Referee #1
Received and published: 17 March 2018

General comments:

The article is in general very comprehensive and detailed. The level of detail is very useful, but so dense it is easy for the reader to get lost. Several tables and figures would benefit greatly by additional labelling to orient the reader. Particularly figures with multiple panels should be labelled with instrument names, reference quality, etc. as appropriate so that at a glance the reader can identify what distinguishes one panel from another and one figure from the next for those that are very similar in appearance.

The ordering of two sections seems illogical. This is based on the concept that the satellite data should be fully discussed before discussing the FRM. Yet a sentence in the section on screening implies that the screening is not solely based on satellite data quality, but additionally on coincidence opportunities with FRM. If this is the case, the order presented makes sense, but how and why the coincidences with FRM factor into the screening is not motivated or explained.

The section describing the L3 data gridding process is not clear for the novice, and overkill for an expert. Choose your audience, and make adjustments.

Detailed comments:

(1) P2, line 16-17: Needs references for SBUV/2, GOME and OMPS.
(2) The authors agree that references are required here. References to (Heath et al., 1975), (Burrows et al., 1999), and (Flynn et al., 2006) have been added in the text and in the reference list.
(3) The second sentence of the introduction now reads as follows: "Atmospheric ozone concentration profiles have been retrieved from solar backscatter ultraviolet radiation measurements by nadir viewing satellite spectrometers since the 1960s, starting with the USSR Kosmos missions in 1964-1965 (Iozenas et al., 1969) and NASA's Orbiting Geophysical Observatory in 1967-1969 (Anderson et al., 1969) and Backscatter Ultraviolet (BUV) on Nimbus 4 in 1970-1975 (Heath et al., 1973), and continuing with the Solar SBUV(/2) series after 1978 (Heath et al., 1975), the Global Ozone Monitoring Experiment (GOME) family of sensors since 1995 (Burrows et al., 1999), and the Ozone Mapping Profiler Suite (OMPS-nadir) series started in 2011 (Flynn et al., 2006)."

(1) Section 2: An introduction to the orbital characteristics of the satellite vehicles will be useful for the reader to better understand the later discussions on gridding and colocation of ground data. The beginning of Section 2 might be a good place for such a discussion. Section 3.3: As previously noted in section 2, knowledge of the orbital characteristics of the satellite vehicles would help in the understanding of the points in this section.
(2) The authors agree that some knowledge on the orbital characteristics of the satellites might be of help to the user. This information has been added in Section 2.1. However, regarding the co-location criteria that are used, knowledge of the LST (as indicated in Table 5) is sufficient. Section 3.3 has therefore been slightly extended with reference to the orbital characteristics mentioned in Section 2.1.
(3) Section 2.1 has been extended with the following sentence: "All instruments listed in Table 1 are on satellite vehicles with a sun-synchronous low-earth-orbit, resulting in fixed local solar overpass times (also see Section 3.3)." and Section 3.3 has been slightly changed with reference to the first addition: "These time windows are chosen to generally have at least one satellite co-location with each FRM, given the satellite's fixed local solar time (LST, also see Section 2.1) and the fact that ozonesondes are typically launched around local noon, while lidar measurements are taken during the night."

(1) Table 1: Additional columns indicating physical characteristics (vertical units/resolution/range, horizontal grids) of the measurements would be useful. These are all discussed in the text, but Table 1 is an opportunity for easy reference.
(2) The authors agree that such overview would be helpful in understanding all CRDP products, and have added two columns to Table 1 and extended its caption.
(3) Two columns have been added to Table 1, and its caption has been extended as follows: "The products' vertical range (with number of levels or layers between brackets) and original units are added in the last two columns."

(1) P5 line 23: Change 'has to stay' to 'must stay'.
(2) Agree
(3) On page 5 line 23 "has to stay" has been replaced by "must stay"

(1) P6 line 14: The A priori for RAL and FORLI are both constructed from the same source as indicated. Are they also both global? It is not clear from this statement.
(2) The authors agree that this statement is not fully clear and have modified the text to make the similarities and distinctions between RAL and FORLI prior data clear.
(3) P6 line 13-15 has been updated as follows: "The a priori information used in the FORLI algorithm consists of a single global ozone prior profile. The prior variance-covariance matrix is built from the McPeters-Labow-Logan climatology (McPeters et al., 2007), as for RAL."

(1) P7 line 7-9: Are these rejected data included before or after the 'screening' discussed later in the paper?
(2) This question is not fully clear to the authors. All data screening is discussed in Section 2, summarised in Table 3, and studied in Section 4.1. The relative screening numbers in Figure 3 refer to all screening as discussed in Section 2 relative to the total number of retrieved profiles.
(3) No further action has been taken.

(1) Section2.4 L3 monthly gridded data: This section is not needed for experts in gridding data, and not helpful to the novice, so it is not clear who the authors are writing to. Figure 1 and this section would benefit for a discussion of the orbital characteristics of the satellite vehicles (either here or at the beginning of Sect 2 as suggested.). Also relate A, B,. . . and 1,2,3. . . to the physical items they represent. Refer to the profiles of the L2 data, and the grid points of L3. If A, B, etc. are the grid points, and 1, 2, 3 are the L2 profiles (and it is not clear that this is the case), is there an advantage to this approach of 4 grid points defining a rectangle, and subdividing the enclosed area, or is it the same as creating a rectangle around a grid point and assigning all profiles within that rectangle to the grid point? The latter seems so much simpler conceptually at least to a novice. What is the subtle missing difference?
P7, line 21: Is there a reference for the GOME/GOME2 convention?
Caption to Fig 1: Why is TM5 assimilation grid referenced here? This figure is used to illustrate the creation of L3 data, not the assimilated L4 data.
Section 2.5: There is a detailed, though difficult, description of how to create the L3 gridded data, but no discussion of how to move to the 2x3 degree L4 grid. This is confusing since Fig 1 refers to the transport model. This needs a little clean up.
(2) In the context of the KNMI L3 product, a pixel refers to a satellite measurement, while a lat-lon grid cell refers to the regular 1x1 degree latitude-longitude grid for which the mean and standard deviation are calculated. Each pixel is divided into 25 subpixels, which are assigned to the grid cell containing the subpixel. The mean and standard deviation for the grid cell are calculated according to the equations given in the text. The authors agree with both reviewers that the text on which subpixels are assigned to which grid cell is unclear and the text of section 2.4 and the caption of figure 1, which is preferably maintained, have been updated accordingly.
(3) The paragraph before Eq. (1) has been replaced by the following: "Monthly-averaged L3 profile products are produced from the filtered RAL v2.14 GOME, GOME-2A, SCIAMACHY, and OMI data by the Royal Meteorological Institute of the Netherlands (KNMI). Version 0004 of the KNMI L3 products has been used in this work (see Table 1). The KNMI level-3 data consist of monthly ozone

profile averages, also on a one-by-one degree latitude-longitude grid, containing 19 layers between 20 fixed pressure levels at each grid-point. The algorithm that calculates the monthly-averaged ozone fields assumes that the L2 satellite ground pixel vertices (labelled ABCD) are ordered as indicated in Figure 1. Each pixel's across-track direction is defined by the lines AD and BC, while the along-track direction is defined by the lines AB and DC. The satellite pixel is divided into 25 subpixels, five in the along-track direction and five in the cross-track direction, and each subpixel is assigned to the L3 grid cell (the boundaries are indicated with the dashed lines in Figure 1) containing the subpixel. The subpixel values x_i are weighted by the square inverse of their uncertainties (σ_i^{-2}), so the weighted mean grid cell value x_c and the corresponding standard deviation σ_c are given by"
The caption of Figure 1 now reads as follows: "Figure 1: A L2 satellite pixel ABCD is divided into subpixels (diamonds 1 to 7). Each subpixel is assigned to a L3 grid cell (indicated with the dashed boundaries) and the average and standard deviation are calculated (see text). In this example, subpixels 1-3 would be assigned to the lower-right grid cell and subpixels 4-7 would be assigned to the lower-left grid cell. The satellite pixel ABCD may have any orientation with respect to the L3 grid."

(1) P8, line 9: 44 ozone layers in what altitude range?
(2) The authors agree that this was not clear.
(3) "surface to 1 hPa" has been added as a clarification between brackets.

(1) P8, line 27: 'data harmonization' means different things to different people. Many think of it as bias correcting as a step preliminary to combining data. Perhaps use 'harmonization of data reporting units' to clarify.
(2) Thanks for pointing out this ambiguity. The authors have changed the text to clarify.
(3) "data harmonisation" has been replaced by "harmonisation of data representation in terms of vertical sampling and units"

(1) P 9 line 17: It would be beneficial to add a line or two about the additional screening criteria used in this study and Hubert et al. 2016 for the ozonesonde data.
(2) The authors agree. A sentence has been added after the reference to Hubert et al.
(3) Added sentence: "Entire FRM profiles are discarded when more than half of the levels are tagged bad or when less than 30 levels are tagged good."

(1) P 9 line 26: State measurement variables and resolution for the lidar as a parallel to the ozonesonde description in the previous paragraphs.
(2) The authors believe that the information requested by the reviewer was already available at the end of the paragraph under consideration (thus not above).
(3) No changes have been made.

(1) Figure 2: When ozonesonde is removed as an FRM for the level 4 data, there is little left in the tropics to validate L4.
(2) The authors are somewhat confused by this statement. Nowhere it is stated that ozonesondes are not used for L4 validation. On the contrary, it is stated in the text that "For the six-hourly assimilated L4 data, the unique temporally closest ground-based reference measurement is always less than 3 hours away." Meaning that there is a co-location for each FRM.
(3) No action has been taken.

(1) P10, line 13-14: Do you mean within one month (+/- one month) or within relevant month?
(2) Thanks for pointing out this unclarity. The text has been updated to make elucidate this statement.
(3) "All FRM within this grid cell within one month are included in the analyses for the L3 comparisons." has been replaced by "All FRM within this grid cell and within the relevant month are included in the analyses for the L3 comparisons."

(1) Table 4: The column name SPI needs more explanation. How to the numbers in this relate to Figure 1?

(2) The authors admit that the meaning of the SPI values had erroneously not been mentioned in the text. Therefore the text has been extended with reference to Table 4. This however does not immediately relate to Figure 1, as should now be clear from the updated text.
(3) After the first reference to Table 4 in Section 3.3, the following sentence has been added: "The possible satellite pixel index (SPI) values within each cross-track scan and the resulting number of pixels per scan are provided for each instrument in Table 4 (taking into account pixel co-adding, see Section 2)." The notation of the possible SPI in Table 4 has been changed from X:X:X (start, step, end) to X,X,X,…,X (start, start+1, start+2,…,end).

(1) Section 3 leads with a description of the layout of the next several sections. This is very helpful given the complexity of the paper. But it is unclear why the choice is made to shift at this point to a description of the FRM data before completing the discussion on information content (screening) of the satellite data. Are these not separate concepts? Why not continue with the evaluation of satellite, and complete it before moving onto the description of the FRM? (See also related comment in section 4.1 specifically P12, line 14.)
(2) The authors agree that this approach might be somewhat misleading as the pre-processing of the data might not have been fully clear from the text: The data and information content studies are performed on ground station overpass data, i.e. satellite pixels must be within a 300 km radius from a FRM station. Section 3.1 has been rewritten to make this clear and motivate the subsequent ordering of sections.
(3) The end of Section 3.1 has been replaced by the following: "The satellite data collection and post-processing (mainly L2 profile screening) is described by the previous section. The L2 datasets have however been reduced to 300 km ground station overpass datasets for the quality assessment in this work, in order to reduce the total amount of data processing (i.e. satellite pixels must be within a 300 km radius from a FRM station). The FRM data selection, co-located datasets study, and data harmonisation are therefore included as the successive subsections within this section. The satellite data content studies and information content studies are discussed in the next Section 4. These include statistics on the L2 station overpass data screening and spatiotemporal coverage, and averaging kernel-based information content measures, respectively. The comparative analysis with both spatially and temporally co-located FRM data follows later in Section 5."

(1) Section 4.1 Data Content: It is not clear how a measure of percent of data screened is a measure of data content. It is apparent that the desire is knowledge as to the distribution of the satellite data in latitude and time. It is noted in the description of Figure 3 that for IASI-A, there is little data removed by the screening process leaving a featureless contour implying an even distribution of data. But it is also stated that this is due to pre-screening of data before release by the data providers. This technique does not show where the pre-screening removed data. Instead a more relevant measure of content and distribution would be the absolute number of measurements left after screening and its latitudinal and temporal distribution.
(2) In line with the previous comment and corresponding answer, the authors believe that the presentation of percentages is now better motivated: As station overpass data are studied, absolute numbers would be misleading and even more stress the spatial selection of the data. Figure 3 mainly wants to show were L2 data can be found and what the impact is of the screening suggested by the data providers. This has been made more clear in the first paragraph of Section 4.1.
(3) The beginning of Section 4.1 has been updated as follows: "The nadir ozone profile CRDP L2 data content study focuses on the spatiotemporal distribution and the effect of screening of the retrieved satellite profiles in the first place, next to the regular file structure, file content, and value checks for the quantities of highest relevance (also see Table 3). Figure 3 displays the latitude-time distribution per 10° latitude band and per month of the percentage of screened profiles for all nadir profile L2 station overpass (300 km) datasets (except for IASI on Metop-B)."

(1) P12, line 15: How can the latitudinal striping in the UV-Vis instruments be partially 'due to station overpass' if the screening is solely based on criteria in Table 3? Is screening based solely on data quality, or also on co-location? Additionally what data is in the CCI data release? Only the screened data? Only the screened co-located data?

(2) The authors agree that this was unclear from the original text, but believe that this is now clarified by the previous two answers on the use of 300 km station overpass data. It should be clear that the full L2 datasets are available in the CCI data release, without screening and without any co-location.
(3) No additional changes have been made to the text.

(1) Figure 3, first panel: What causes the gap in the GOME dataset after 2003 in the tropics?
(2) As ground stations are located near the South-Atlantic Anomaly (SAA) and a quite severe GOME data screening has to be applied, no (near) SAA data are left. This has now been made clear in the text.
(3) "The lack of GOME data in the southern mid-latitudes from 2003 onwards is due to severe screening of L2 overpass data for ground stations that are all located near the South-Atlantic anomaly (SAA)." has been added.

(1) Figure 3, caption: What is meant by 'The decreased GOME-2B data from 2015 onwards justifies additional screening' mean? Are you trying to say that it indicates additional screening?
(2) The authors agree that this statement is misleading. The caption has therefore been brought in line with the main text.
(3) "justifies additional screening" has been replaced by "points at a retrieval issue"

(1) P14, line 5: change to 'understanding of how the system'
(2) The reviewer's proposal for improving the readability has been followed, yet somewhat differently, in agreement with the suggestion by the second reviewer.
(3) "understanding how the system" has been replaced by "understand how the system"

(1) Figure 4, first panel: Why is the area in the tropics of missing data in the GOME panel larger than that in Figure 3?
(2) If all data are screened (100 % values in Fig. 3) than the DFS and other information content values are empty.
(3) No changes made.

(1) P17, line 20: From here after there is inconsistent use of BG and of Backus-Gilbert. BG is used extensively in the Figure labels and captions, and occasionally in the text. Introduce the acronym here, then use BG only after.
(2) Thanks for pointing this out. The acronym as been added.
(3) P17, line 20 "Backus-Gilbert spread" is replaced by "Backus-Gilbert (BG) spread"

(1) Figure 5: The offset in the second and third rows are labelled identically, but the graphs are different. The caption only states that 'different measures are used'. Are the measures direct and centroid? Differences in the measures for width are clearly indicated. Offset could also be simply added by label and in the caption.
(2) The authors agree that offset and spread indications in Figure 5 can be improved. The caption of Figure 5 has been updated accordingly.
(3) The caption of Figure 5 has been updated as follows: "Global GOME-2A (left) and IASI-A (right) information content in terms of vertical sensitivity, retrieval offset (in km), and averaging kernel width (in km) and their dependence on DFS, SZA, or thermal contrast (TC). Black dashed lines represent median values, while out-of-range profiles are plotted in magenta. Different measures are used for the offset and kernel width in the second and third rows, which include the centroid offset and Backus-Gilbert spread, and the direct offset and FWHM, respectively. Plot titles provide the absolute and relative amounts of profiles after screening, and the number of ground-based overpass stations."

(1) P20, line 22: change 'fiver' to 'five'.
(2) Thank you for spotting this typo; the text has been corrected.
(3) 'fiver' has been changed to 'five'

(1) P21, line 18: Here 68% interpercentile spread is used for the first time, but the acronym IP68 is not introduced. Later in the text and graphs there is inconsistent use of the acronym and the full term. Introduce both here, and consistently use the acronym or the full name in later text.
(2) Actually the 68% interpercentile spread and its acronym are first introduced on page 20, line 8, as Q84-Q16.
(3) The authors have added "68 % interpercentile" explicitly to page 20, line 8 to avoid the impression of inconsistent use of terms and acronyms.

(1) P21, line 26-27: Should 'vertical averaging smoothing' be 'vertical smoothing'?
(2) The authors intended either "vertical smoothing" or "vertical averaging kernel smoothing". The latter has been chosen here. Yet in agreement with a comment by the second reviewer, the phrasing has been changed.
(3) "vertical averaging smoothing" is replaced by "vertical smoothing of ground-based reference data with averaging kernels"

(1) P39, line 10: Add the word ozonesonde: '64 ozonesonde stations'.
(2) The authors have followed the suggestion by the reviewer.
(3) "64 stations" has been changed into "64 ozonesonde stations"

(1) Figure 13 caption refers to top, middle and bottom instead of left, middle and right. Label each panel.
(2) The caption of Figure 13 has been written with the final mark-up of the paper in mind, i.e. with the three plots combined into a column (not a row).
(3) No action has been taken.

(1) P40, line 4-5: The Southern mid lats do not look smaller but similar to the tropics in the UTLS.
(2) It was the intention of the authors to state that the bias indeed looks similar, but has only positive values. This observation has been made more explicit in the update of the text.
(3) "but smaller" has been replaced by "but only positive"

(1) P40, line 19: Replace 'As for' with 'Similarly to' and remove the word 'now'.
(2) The authors have adopted the suggestions by the reviewer to increase readability.
(3) The sentence referred to now reads as follows "Similarly to the L2 RAL v2.14 UV-VIS retrievals, Figure 14 and Figure 15 contain…"

(1) Figure 3: This figure (and many after) need additional labelling. Label each panel with the satellite name so it is obvious at a glance.
Figure 4: Label each panel with the satellite name.
Figure 5: This figure is difficult to interpret and needs more explanation and labelling. Label the columns with the instrument name.
Figures 6-10: These are very hard to distinguish when trying to compare the results. Label each figure with the instrument and years (GOME 1996-2010 for example). Also label each panel with the influence quantity. These are stated in the caption, but are more easily interpreted if the panels are directly labelled.
Figure 11: Label the columns with Latitude and Quarter, and the rows with the instrument name for easy recognition.
Figure 12: Label the columns with L2, L3 and the rows with the instrument name for easy reference.
Figure 14, 15: Label each panel with the influence quantity displayed, and 'drift' in the final panel of Fig. 14.
(2) The authors agree that readability and interpretation of graphs can be improved upon insertion of satellite instrument and influence quantity labels on the relevant plots.
(3) All plots have been updated with the requested labels.

(1) Figure 16: Why is the time series shown for IASI L3, but not for others? It might also be enlightening to show the profile of the L3 drift.
(2) For the FORLI IASI product only tropospheric column L3 data are available, so only columnar values can be shown. Such values however allow for a more easy time series representation. Vertical drift profiles are not possible, and have been replaced by a trend line.
(3) No further action has been taken.

(1) P 51, Acknowledgements: Some of the NDACC PIs listed are retired. It might be of use to additionally include the current persons in these positions as is done for TMF.
(2) The authors acknowledge that some PI references require updating. The names of R. Querel and R. C. Schnell have been added.
(3) The lidar PI acknowledgement now reads as follows: "CNRS and CNES (Dumont d'Urville station and Observatoire Haute Provence, PI is S. Godin-Beekmann), DWD (Höhenpeißenberg station, PI is H. Claude), RIVM and NIWA (Lauder station, PIs is are D. P. J. Swart and R. Querel), NASA/JPL (Mauna Loa Observatory and Table Mountain Facility, PIs are I. S. McDermidR. C. Schnell and T. Leblanc), and NIES (Tsukuba station, PI is H. Nakane)"