The reviewer has raised a number of points that we have carefully considered in revising the manuscript.

One major disappointment of this paper is that the authors do not bother to show how the ozone normalization factor, and the relevant uncertainties in the measurements, changed from pre-homogenization to post-homogenization. They make the point both early, and in the conclusions, that there was real improvement, but do not bother to show one of the major results of this work. This could also be real motivation for other ozonesonde scientists to homogenize their own data sets. Following that line, one example of a station's long term trend for the pre- and post-homogenized data would be highly interesting. Did it make a difference? Was it worth doing, or is there minimal impact on the trends? Several times the authors hint that there is minimal impact on trends, to keep alive previous literature based on trends in the Canadian ozonesonde data, but they do not provide the example graphs to drive this point home, which is important.

We have tried to make the point brought up by the reviewer on the improvement of the data record through the homogenization effort using the example for Edmonton in Table 3. Edmonton was chosen because it has the most complete record of independently measured column ozone. For the Brewer-Mast record there is moderate improvement in both the variability of the normalization factor and the trend. For the ECC record there is a substantial improvement in the variability and the trend in the normalization factor is essentially eliminated. We preferred to use the table over a plot of the normalization factor to provide a more quantitative assessment.

We have removed the remarks about "...and so should not invalidate past analyses that have used Canadian network data", and have simply calculated the trends, using ECC data, before corrections are applied and presented these for comparison in Figure 15.

5217.1-5. This first sentence is better, but still doesn't work. I had to read it three times to understand the intent. Make it two sentences, breaking it when stratospheric ozone is discussed.

We agree; we made it three sentences.

5217-second paragraph. The literature cited is highly uneven. Why on some important topics is one reference "and references therein" provided, whereas for other topics up to 7 citations, prefaced with e.g., are included? The 7 citations beg the e.g. Use a balanced approach, and do the work if you don't know the references. Check out the paper where they are found. An even distribution of references would help the argument as to the importance of the network.

We chose to use "the references therein" format in the introductory material where we provide the motivation for an improved ozonesonde record. Since this is an instrument/data set focused paper a detailed discussion of the geophysical importance of ozone in both the troposphere and stratosphere did not seem to be warranted. We have dropped that terminology in reference to Arctic ozone loss and added

additional references. We also dropped the “e.g.” as suggested since the references given provide the examples.

5218.9-15. ECC is not a word. I believe there are some references associated with the development of this instrument, which could be useful (and it was not Tarasick). Note that Brewer-Mast sondes received such respect, and they are no longer used. Later in the paragraph we find there has been three decades of work and minor changes for the ECC, and not one reference!

The citation to Komhyr (1969) was missing from the text (although it was in the reference list). It has been added. We have defined “ECC”. The Tarasick et al. (2005) reference was included because it dealt specifically with the Canadian ECC record. We cite a number of papers that have investigated various aspects of ECC performance when discussing specific areas where corrections or adjustments were made. The Smit et al. (2012) reference that is cited here provides a summary of the changes,

5219.1-18. Is this the operating principle of all ozonesondes, or is this chemistry particular to one type? Where is the number, given to 4 significant figures, in equation 1 from? It results from well-known physical constants which should be used in spite of what Komhyr published. The equation could also benefit from using flow rate rather than “time to pump 100 ml”, which is flow rate upside down. The formulation of the equation as it is shown, and specifying the units of the quantities, is operator manual stuff, not scientific paper stuff. If the equation is presented scientifically, units don’t have to be presented. They will be universal and the detail only necessary when the equation is applied, which is not here. I’m not even sure what “kelvins” are. What are “bias changes”?

Yes, this is the operating principle of all modern ozonesondes. Equation (1) is typically quoted that way in the literature, but we have added the derivation, as the referee requests. The kelvin is the SI unit of temperature. We say “bias changes” in order to distinguish from simple random errors.

5220.14-23. How is the climatology of McPeters and Labow used to renormalize? The climatology is used to extrapolate the sounding from ceiling to top of atmosphere. Then the sonde total column is normalized to a total ozone measurement. There are three steps. First the total ozone record is corrected, second the integrals of the ozone soundings are corrected with the climatology, and then the two are compared to provide finally a normalization factor for the sounding. This paragraph could be clearer. Turn it around so the correction to the soundings is discussed first, then the integration, then the total ozone data, just as it is done in practice.

This section has been revised as suggeted. We note that the total ozone normalization is applied only after all other corrections have been applied to the non-normalized data; that is, any previous normalization is first removed.

2.2 Why no mention of the work done by the Swiss with their Brewer Mast record, Stubi e al.?

We are quite aware of René's excellent work with the Payerne record. Unfortunately it is not applicable to our case, nor are the Hohenpeissenberg or Uccle revisions, as all three improved on the Brewer-Mast preparation practices suggested by the manufacturer quite early on, while the Canadian program simply followed the manufacturer's directions. We understand that the Australian practice was similar to the Canadian practice, as noted.

5222.25. Even though the absolute concentration differences could be considered minor, it is a factor of 2 difference in concentration of KI, which is not exactly minor.

A fair point. The effects are minor, but the concentration change isn't. Text revised.

2.6 Were the corrections were made following Smit et al.? That is the corrections were spelled out in Smit et al.? If so state it, otherwise the correction details should be spelled out here.

Done.

2.7 This section contains a surprise not mentioned in 2.6. The "box" temperature, or other interior temperature, is unavailable before 1999. That is pretty recent. Why are we then talking about pump temperature measurements for sensors 3A and 4A which were in use before 1999, and which thus could also not be corrected?

We said that the pump temperature profile was recorded in the WOUDC file only after 1999. Unfortunately the original telemetry data (raw cell currents, box temperature, etc.) have not been preserved for older flights. One can work backwards from Equation (1) to raw cell currents, but only by making assumptions about the temporal variation of pump temperature. This only affects the background current correction, as the other variables in (1) are multiplicative. We can calculate the ozone adjustment that results from a relative change in box temperature, without knowing what the box temperature is.

This also raises another question. Is the Canadian network homogenization dependent on the WOUDC for its data records? I would have expected that they had more complete records at each station where a measurement was made. The WOUDC is merely a data archive for the finished product. The authors need to be clear about what data are used in this homogenization project. Is it all based on only what is available from the WOUDC, or are there other sources?

Most of the information is in the WOUDC. Individual station (paper) records were consulted with regard to changes to station practices, and to add pre-flight information that was not in the older WOUDC records. Part of the exercise was to transcribe this information into the revised WOUDC records.

5224.21-5225.4. I doubt that these day to night differences are the result of the pressure sensor which should not have a diurnal component, but rather to poor temperature measurements due to some deficiencies in handling radiation on the temperature sensor. This needs to be made clear whether the authors are ascribing this diurnal difference to the pressure sensors and if so why. I don't understand the rationale for the one sentence paragraph beginning, "Pressure Errors . . .", nor why the agreement of the local noon flights suggest that the pressure errors are more variable. These two paragraphs need to be combined and made more intelligible. What exactly is being claimed? How are these differences ascribed only to the pressure sensor and not to issues also in temperature? How is it possible to separate pressure and temperature if the comparison is between the calculated altitude from the sonde and an external altitude reference such as a radar or gps.

While the referee's suggestion about day-to-night differences would seem to make sense, unfortunately to ascribe the large height differences to temperature errors implies unreasonably large temperature errors --- so we are left with a puzzle about these differences. We have revised the text.

2.8. This sub section is longer than practically any other sub section in section 2, yet the end result is that nothing can be done, because of the difficulties inherent in the changes, and their unknown final impact on any individual sonde. Does the reader really need all the detail? Not this reader. I would prefer to see this section summarized much more succinctly, and a statement right at the beginning to indicate that none of what follows will be used in the correction of the data, other than as a source of uncertainty. Then spell out what that uncertainty will be and how it will be incorporated, e.g. as an uncertainty on sonde height based on year/day/night/ . . . , or . . .

This is an important source of error that is larger than is commonly recognized, so we feel it needs to be discussed. While it is somewhat unsatisfying to not be able to correct this source of uncertainty for each sounding, we feel it is necessary to provide a full discussion in order to point out that the final result of an ozonesonde profile encompasses more than the ozonesonde sensor itself.

Figures 1-4. These should be combined into one 4 panel plot, so they can be compared easily. The legends are not clear. What is the difference between renormalized and normalized, what do the labels, TARASICK et al., MATEER, DESCHLER (sic) refer to, why are some labels in all caps and other upper and lower case? Here it shows temperature corrections throughout the profile for ECC sondes in the 1980s, whereas earlier it was claimed the temperature profiles were not available for data prior to 1999. How can these

facts be reconciled? It seems the numbers on the abscissae are off by 10^5. People usually don't display ozone partial pressure in hPa.

Those typos have been corrected. Apologies to Terry D.

5226.18-23. Why isn't a temporal history of these normalization factors shown for the Canadian record? This would be quite interesting, and a good indication of the usefulness and impact of the corrections. In fact a before and after graph would be very illustrative of the usefulness of the corrections and the stability of the Canadian measurements.

This was one of the major criticisms raised by the reviewer in the overall remarks at the beginning of the review. We have repeated our rationale for the choice of a table instead of a plot below.

We have tried to make the point brought up by the reviewer on the improvement of the data record through the homogenization effort using the example for Edmonton in Table 3. Edmonton was chosen because it has the most complete record of independently measured column ozone. For the Brewer-Mast record there is moderate improvement in both the variability of the normalization factor and the trend. For the ECC record there is a substantial improvement in the variability and the trend in the normalization factor is essentially eliminated. We preferred to use the table over a plot of the normalization factor to provide a more quantitative assessment.

4.3 Pump calibration. It is not clear if this just means flow rate measurement, or pump calibration to determine efficiency as a function of pressure. The paragraph suggests the former, the title the latter. Pumps are not calibrated for a fixed flow rate, rather the flow rate is just measured.

Yes, that's right. We've taken the referee's suggestion and changed the text accordingly.

4.4 Confusing. What graduated cylinder? What is at 100% RH, the cylinder or the air being expelled? What is the uncertainty then on: the flow rate, the pump calibration, the ozone sensed? This paragraph may only be understood by those intimate with making ozone measurements, all others forget it.

We have revised this text.

4.5 Is the 0.5% just a guess, or is there some basis, and if a guess how so accurate?

This is approximately the standard error of the mean in the Deshler et al. measurements. Text revised to note this.

5228.19-20. What latter seven lines? Do the authors mean the latter seven items in the list, but there were only 5. Is the reader to count back seven lines? Do the authors mean the next 7 items? Is it too difficult for the authors to list the items with errors varying throughout the profile? If the errors vary is that indicated in Table 4, and then how can a constant error be stated? How is the variation with altitude handled?

We have now numbered the items in Table 4.

4.6 Now we're back to pumps. Why aren't 4.3 and 4.6 at least grouped together? What is meant by pump correction and its error? The uninformed reader will again not know what is being discussed, or that it has a pressure dependence, until maybe the last line. What larger values have been averaged? If I am correct the corrections to flow rate at low pressure can be significant, > 10% at 10 hPa. Table 4 indicates just about 3% at 10 hPa. The authors need to clarify this section.

We noted that the first five items in Table 4 were independent of pressure, while the latter seven were pressure-dependent. Table 4 lists the magnitudes of the pump correction errors, not the magnitudes of the pump correction itself (these are listed in the WOUDC file).

4.7 Can the 2.5 mL bias the measurement high? According to the explanation it only misses ozone, it doesn't add it. So why the plus/minus 4%? What is the pressure dependence?

Again, we're talking about the error in the correction, which can be positive or negative. The error is proportional to pressure, as noted in Table 4.

4.8 Why is the difference randomly variable? One is a constant, the other pressure dependent, thus it shouldn't be random, but decay smoothly as pressure decreases. To indicate why it is always positive mention that the pressure dependent part is proportional to the ratio of pressure with surface pressure. This is a poor way to introduce Figures 5-8 which are showing more than the effect of the background current, and finally the graphs do not show an average magnitude, rather an uncertainty that has a profile which is not easily averaged.

The magnitude of the error (or of the difference between the two corrections) is randomly variable. The effect of the difference in relative terms is largest just below the tropopause, where absolute amounts of ozone tend to be lowest. We have removed the reference to Figures 5-8, which was indeed out of order.

Figures 5-8. These should be combined into one 4 panel plot so the reader can see the changes easily. How many of these uncertainties are two sided? What do the JOSIE ECC/BM, Vanscoy ECC labels refer to? These figures need to be introduced more properly,

rather than off hand in one item in section 4. They can then be used when some of the items already discussed are described. Their introduction doesn't come until much later. Why are the ordinates not shown, only the tick marks? If their introduction is to be delayed then don't mention them in section 4.8.

All the uncertainties are two sided. These labels are defined at the bottom of Page 5230,where the first reference to the figures now appears. Not sure about the axis lines – appears to be a bug in the PDF generation software.

4.9 I doubt that many besides the authors, and I am not sure all of them, understand what is done for this correction. This paper's readers will certainly not understand it.

We have changed the text to note that the correction is largest (i.e. 7%) at 1000 hPa and declines quadratically with log(pressure).

4.10 What is the empirical estimate based on? At least here the pressure dependence is specified. Why is solution evaporation more important in Brewer Mast than in ECC sondes?

The Brewer-Mast sensor has a somewhat more open construction that may allow more solution evaporation. Brewer-Mast sondes also use a much weaker (0.1%) KI solution, which may allow significant iodine evaporation (Brewer and Milford, 1960; Tarasick et al. 2002). We have added this to the text.

4.11 It is hard to reconcile this section with table 4, where the ascent rate variation is +/- 12% * exp(del t/tau sub del O3), whatever the quantities in the exponent mean. They are not explained. Yet here in the text it is a "modest (< 1%)". Why are these errors modest while previous errors of 0.5% are specified? Are they modest as well?

Text revised. Yes, they are modest.

4.12 What are the values of the manufacturer's 1 sigma uncertainty? What is multiplied by the ozone gradient? Why is the ozone gradient used? If the manufacturer's 1 sigma uncertainties are used what values from Richner and Phillips and Steinbrecht et al. are used?

These are quoted in Table 4; the values are given by Richner and Phillips and Steinbrecht et al.. The uncertainty in ozone is the pressure uncertainty times the gradient of ozone with respect to pressure.

5230.17-5231.9, Figures 5-8. So now the figures are introduced, but there are questions reconciling them with table 4.

1) The only corrections explicitly shown are pump correction error and background current, and if these two were added together they should come close the “same balloon” and their difference with the “same balloon” should indicate the impact of all the other errors, but they don’t. In Figures 6-8 the sum of the pump correction error and the background current exceeds the same balloon. In fig. 6 at the top of the profile there is only 1% difference between background current and same balloon, while the pump correction factor is 4%.

The uncertainties don’t add linearly, but rather as the root of the sum of their squares. Taking this into account gives the result we have shown

2) If the errors in any column in Table 4 are added except for the pump correction and background current they would be in the range of 6%, e.g. 5A/6A, with the major contribution from the 2.5ml solution error, while the pump correction is 1-3% (so stated) and the background current error range not easily displayed, but from the graphs it is on the order of 6% and less. So why then isn’t there a bigger separation between same balloon and the sum of pump correction and background current. The figures cannot be resolved with the table.

As in the previous case the uncertainties don’t add linearly so that when we account for this we get the result shown.

3) How many errors are double sided? It would be good to replot the data allowing all two sided errors to go both ways with zero on the abscissa in the middle. Then for example the 2.5 ml error would only show up on the right, as would, if I understand the text correctly, the background current. How about pump correction? Table 4 indicates all errors are two sided, but I don’t believe that is the case.

Based on the methodology we have described all of the uncertainties are double sided.

4) Why, since the background current error is always positive ib*(1-P/Po) is there such a large bump near 10 km? Is that because the ozone is lowest just below the tropopause? It would be useful to indicate this with some climatological temperature profile shown, so the reader knows how the errors are distributed between troposphere and stratosphere. But this illustrates the point that the percent error is dependent on the ozone profile itself. Still it is curious why the other factors, comprising a 6% error don’t appear more strongly.

Yes, that’s correct. We have changed the text to note that it is largest in relative terms just below the tropopause, where absolute amounts of ozone tend to be lowest.

5232.15-20. If I understand correctly these are natural halogens, bromine, inducing the ozone loss, so why should the frequency change?

Good question. Probably a complicated mixture of sea ice, snow cover and meteorology changes, but we don't know yet. We have added a reference to Oltmans, S.J., B.J. Johnson, and J.M. Harris (2012), (Springtime boundary layer ozone depletion at Barrow, Alaska: Meteorological influence, year to year variation, and long-term change, *J. Geophys. Res.*, 117, D00R18,doi:10.1029/2011JD016889) which also notes these influences.

5232.26-29. This statement on stratospheric ozone recovery is incorrect. No one expects to see ozone recovery in the 2000-2013 time frame. Stratospheric chlorine levels, even though declining, are still well above the point where recovery could be seen. Note that this statement is belied by the next two sentences which highlight ozone loss in the cold vortex years.

We are using the term "recovery" in the sense used in WMO/UNEP *Scientific Assessment of Ozone Depletion: 2010* (see Prologue Box 2). By that definition we are in the first stage of recovery, probably the second --- although WMO 2014 is careful about that As the referee points out, it is clear from the context that we do not mean "full recovery". We have altered the wording to make this clearer.

5233.22-25. These negative trends in the lower stratosphere do not appear at all stations and this should be mentioned.

They do at all sites except Eureka, whose record began in 1993. We have mentioned this (parenthetically) as requested.

5234.8-13. Again, these statements about ozone recovery are incorrect, should be removed, and are belied by the very next statements in the manuscript. The authors should investigate what the effective equivalent stratospheric chlorine (EESC) is right now. Paul Newman will send you his estimates of that. With this information the authors will see that the EESC is now equivalent to mid 1990 values, or about 4 ppbv. Ample to keep ozone depletion going for many years to come.

We have changed the wording to make it clearer that we are referring to an earlier stage in recovery, not "full" recovery.

5234.20-25. It is sad that these statements have to be taken on credit. None of this information was presented in graphical form. Yet this type of result should be the primary motivation for, and result of, all the work that was done. And as the authors correctly point

out this is one of the major conclusions. Minor issues: Abstract: Does the reader know what ECC, SD, mean?

We have calculated the trends for ECC data before corrections are applied and presented these for comparison in Figure 15. "SD" was introduced by AMT in the typesetting process. We wrote "standard deviation".

5217.7-9. Switch "they" and "ozone soundings" so the reader doesn't have to guess what "they" are when it appears first.

O.K.

5220.10-13. This punctuation is wrong. There is no need of a colon. Make it two sentences.

O.K.

5224.12. . . subsequently by Environment Canada? Who/what is Environment Canada? Has this been introduced?

Environment Canada is the government agency that runs the Canadian network, more or less equivalent to the US NOAA. But we changed this to "in Canada".

5224.17. Where does 1% at 10 hPa come from, this is a shift of 300 m, not 70 m.

We meant a shift in ozone, not altitude. Text revised.

5232.13. . . sites appears may be. . . ?

Corrected.