Scientific
Research
Publishing

Statistical Tests of the Validation of TCO Satellite Measurements, Recorded Simultaneously by TOMS-OMI (2005) and OMI-OMPS (2012-2018)

Molina-Almaraz Mario, Pinedo-Vega Jose Luis*, Ríos-Martínez Carlos, Mireles-García Fernando

Universidad Autónoma de Zacatecas, Zacatecas, México
Email: *jlpv85@gmail.com

Abstract

Two statistical validation methods were used to evaluate the confidence level of the Total Column Ozone (TCO) measurements recorded by satellite systems measuring simultaneously, one using the normal distribution and another using the Mann-Whitney test. First, the reliability of the TCO measurements was studied hemispherically. While similar coincidences and levels of significance > 0.05 were found with the two statistical tests, an enormous variability in the levels of significance throughout the year was also exposed. Then, using the same statistical comparison methods, a latitudinal study was carried out in order to elucidate the geographical distribution that gave rise to this variability. Our study reveals that between the TOMS and OMI measurements in 2005 there was only a coincidence in 50% of the latitudes, which explained the variability. This implies that for 2005, the TOMS measurements are not completely reliable, except between the −50° and −15° latitude band in the southern hemisphere and between +15° and +50° latitude band in the northern hemisphere. In the case of OMI-OMPS, we observe that between 2011 and 2016 the measurements of both satellite systems are reasonably similar with a confidence level higher than 95%. However, in 2017 a band with a width of 20° latitude centered on the equator appeared, in which the significance levels were much less than 0.05, indicating that one of the measurement systems had begun to fail. In 2018, the fault was not only located in the equator, but was also replicated in various bands in the Southern Hemisphere. We interpret this as evidence of irreversible failure in one of the measurement systems.

Keywords

TOMS-OMI and OMPS Data, Global Statistical Validation, Total Column

Ozone, Mann-Whitney Test

1. Introduction

The atmosphere, due to its composition and stratification, is the element of the planet that makes the Earth a unique planet, the only one that houses a plethora of living species. If the composition of the atmosphere changed radically, life on Earth would no longer be as we know it and a large number of living beings could become extinct. Although not in a radical way, the composition of the atmosphere is in fact changing day by day. This change is mainly due to the release of large amounts of gas emissions and fine particles into the atmosphere. Therefore, the study of the atmosphere is undoubtedly becoming more important and, in turn, more complex.

Ozone, despite being a trace gas that only represents 0.0000006% of the mass of the atmosphere, fulfills important and irreplaceable functions: it participates in the dynamics of the atmosphere and climate, and attenuates both UV-B radiation and a part of UV-A radiation ozone and plays a crucial role also in radiative processes controlling the energy balance on the Earth.

The release into the atmosphere of chlorofluorocarbons (CFCs), which began in the 1930s and reached its peak in the 1990s, came to disturb the balance, both of concentration and distribution, of ozone in the atmosphere, causing the destruction of important amounts of ozone [1] and the appearance of the ozone layer in Antarctica [2].

The 1992 Montreal Protocol concluded with the agreement to ban the production of CFCs in the world. However, banning the production of CFCs does not mean in any way that CFCs have disappeared from the atmosphere [3]. With data from the Earth System Research Laboratory/NOAA, it is observed that in the case of CFC-12, the current concentration is of the same order of magnitude as in 1990 (500 ppt); in the case of CFC-11, the current concentration is of the same order of magnitude as in 1985 (230 ppt); while the concentration of HCFC-22 went from 50 ppt in 1980 to 260 ppt in 2018 [4] [5]. In addition, there is already the presence of new polyatomic molecules in the atmosphere. Even though these molecules can be benevolent with ozone, they are not helpful in greenhouse terms. HFC-134a concentration increased from 0 ppt in 1995 to 105 ppt in 2018.

Although the presence of CFCs in the atmosphere has not been eradicated, several publications have emerged announcing the recovery of ozone. Among the most recent, Keeble *et al.* used values of the total column ozone from a set of simulations obtained with the UM-UKCA model and simulated the progress of ozone recovery [6]. Shortly before, but in the same year, measurements analyzed by Ball *et al.* sadly showed no signs of ozone recovery [7].

The contrast in findings between these two opposite studies is a call for caution and a warning against premature rejoicing over the supposed recovery of

ozone. This disagreement between simulations and measurements alerts the scientific community to the importance of recognising that ozone recovery studies should remain under discussion.

Naturally, we all wish for the recovery of the stratospheric ozone. But one thing is the desire for this to happen and another is the scientific facts. The fundamental question we ask the defenders of the simulations is: Are the models infallible? Or, in other words, are models consistently validated? Obviously, validation from simulated measurements cannot be accepted. In fact, validation from a set of point measurements at ground stations cannot be accepted either. For models to be confirmed, these would have to be validated against actual global measurements. These questions can also be extended to the defenders of the measurements: Are measurements infallible? Are global measurements consistently validated? In this study, we attempt to provide answers to this last question.

The largest and most detailed total column ozone measurements (TCO) that exist worldwide are satellite measurements. The validation studies of satellite measurements are not new. Usually, they are carried out by comparing point measurements at ground stations with corresponding satellite measurements.

Among the validations:

Balis *et al.* presented results of validation of the OMI-TOMS and OMI-DOAS data through comparisons with ground measurements made by the Dobson and Brewer spectrophotometer instruments [8]. They found a global average similarity of more than 1% for OMI-TOMS data and better than 2% for OMI-DOAS data with observations from the ground.

Mc Peters *et al.* carried out two types of validation. The first validation was performed through comparison with an ensemble of 76 Northern Hemisphere ground station network of Dobson and Brewer ground stations [9]. They found that OMI-TOMS total column ozone averages 0.4% higher than the station average, with station-to-station standard deviation of ±0.6%. The second validation method was carried out through aircraft campaigns using the NASA DC-8 and WB-57 aircraft. Ozone above the aircraft was measured using an actinic flux instrument and it was compared with OMI ozone specifically to validate Aura. The comparison shows that the OMI-TOMS ozone was stable over the 2-year period with no evidence of drift relative to the ground network. The OMI-DOAS product is also stable but with a 1.1% offset and a seasonal variation of ±2%.

Lalongo *et al.* (2008), with data collected since 1992 using the Brewer spectrophotometer at the Rome station, found a satisfactory agreement between the OMI total ozone data and the Brewer measurements, both for the OMI-TOMS ozone algorithms and for OMI-DOAS (with biases of −1.8% and −0.7%, respectively) [10].

Anton *et al.* (2009) carried out the comparison of the Total Column Ozone data from the Ozone Monitoring Instrument (OMI) with ground-based measurement recorded by Brewer spectroradiometers located at five Spanish remote sensing ground stations between January 2005 and December 2007 [11]. They

found the largest relative differences between these OMI Total Column Ozone, of the order of 5% with a significant seasonal dependence. They say that Total Column Ozone from OMI-TOMS are on average a mere 2.0% lower than Brewer data. For OMI-DOAS data, the bias is a mere 1.4%.

In a priori approach, it can be interpreted from these studies that there is a reasonable approximation between measurements at ground stations and satellite measurements. But in fact, studies show that the relative differences are not of the same order. Most studies find higher averages in satellite measurements compared to terrestrial measurements but with unmatching orders of magnitude. In addition, all these studies are located in the northern hemisphere, therefore, in all rigor they cannot be extrapolated to all satellite measurements.

A likely explanation for the differences between estimates is that they do not coincide either in geographical location or in the analysis time periods. But above all, the cause of the differences might lie in the dynamics of ozone itself, which involves formation (through UV-C radiation photolysis), destruction (by UV-B or UV-A radiation photolysis), and transport. The ozone layer is not at all a uniform carpet of constant density and distribution.

Therefore, the validation of satellite measurements is not a trivial problem because it involves working with populations with a considerable amount of data.

This work addresses one aspect of the validation of global TCO measurements problem. Our purpose was precisely to study the similarity between TCO satellite measurements, performed simultaneously by two satellites: TOMS (Total Ozone Measuring System) and OMI (Ozone Monitoring Instrument) in 2005 and OMI and OMPS (Ozone Mapping and Profiler Suite) in 2012 and 2018.

2. Materials and Methods

Since the discovery of the ozone layer hole, National Aeronautics and Space Administration (NASA) has monitored the atmosphere to quantify the Total Column Ozone by using satellite measurements. NASA has used 5 satellites equipped with spectrometric systems: TOMS (Total Ozone Mapping Spectrometers) from November 1, 1978 until December 14, 2005; OMI (Ozone Monitoring Instrument) since July 15th, 2004; and OMPS (Ozone Mapping and Profiler Suite) since January 26, 2012. In parallel, the ADEOS (Advanced Earth Observation Satellite) program has monitored ozone in Europe, but its measurements are not openly available unlike those of NASA, therefore they were not considered in this work.

Between October 1, 2004 and December 14, 2005, TOMS and OMI released TCO measurements simultaneously. And OMI and OMPS have measured simultaneously since the beginning of OMPS on January 26, 2012. In this work, the comparison between satellite measurements of TCO covers these periods.

To this aim, the application of two statistical tests was carried out:

The first one can be considered a modification of the method used to compare two means of independent samples whose data number is or may be different.

We have abbreviated it simply as proof of the *Normal Distribution*. Thus, if n_1 is the amount of data measured by a first satellite and n_2 the amount of data measured by the second satellite, the probability of finding a certain statistical difference can then be found by calculating the standard value (*U*) or standard score of the difference

$$U = \frac{\hat{X}_2 - \hat{X}_1}{\sigma_{\hat{X}_2 - \hat{X}_1}} \tag{1}$$

where, $\hat{X}_1$ and $\hat{X}_2$ are the mean values for a certain latitude, (or hemisphere according to the case) for measurements made by satellite 1 and 2, respectively; and the standard deviation of the difference is given by

$$\sigma_{\hat{X}_2 - \hat{X}_1} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{2}$$

The second test used was the *Mann Withney U test* [12]. The Mann-Whitney *U* test (also called the Mann-Whitney-Wilcoxon (MWW), Wilcoxon rank-sum test, or Wilcoxon-Mann-Whitney test) is a nonparametric <u>test</u> of the <u>null hypothesis</u> in which it is equally likely that a randomly selected value from one population to be less than or greater than a randomly selected value from a second population. This test can be used to investigate whether two *independent* samples were selected from populations having the same distribution. In which case, the standard score is given by

$$U = \frac{W - \frac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}} \tag{3}$$

The subscripts 1 and 2 correspond to the two data populations to be compared, n_1 and n_2 are the amount of data from both populations. *W*, is the sum of the ranges of the samples and is determined by

$$W = \sum_{i=1}^{N_r} |x_{2i} - x_{1i}| R_i \tag{4}$$

where R_i is the range for the sample or data group and N_r is the size of the sample.

Both tests were applied by hemispheres and by latitudes (degree by degree).

The null hypothesis was H_0: There is no statistical difference between the TCO obtained between both satellites. The alternative hypothesis was H_1: There is a statistical difference between the measurements of both satellites.

3. Hemispheric Comparison between TCO Satellite Measurements

Figure 1 presents the statistical comparisons, day by day, of the TCO measurements from TOMS and OMI in terms of the significance level for both hemispheres for the year 2005, using the Normal Distribution and the Mann-Whitney test.
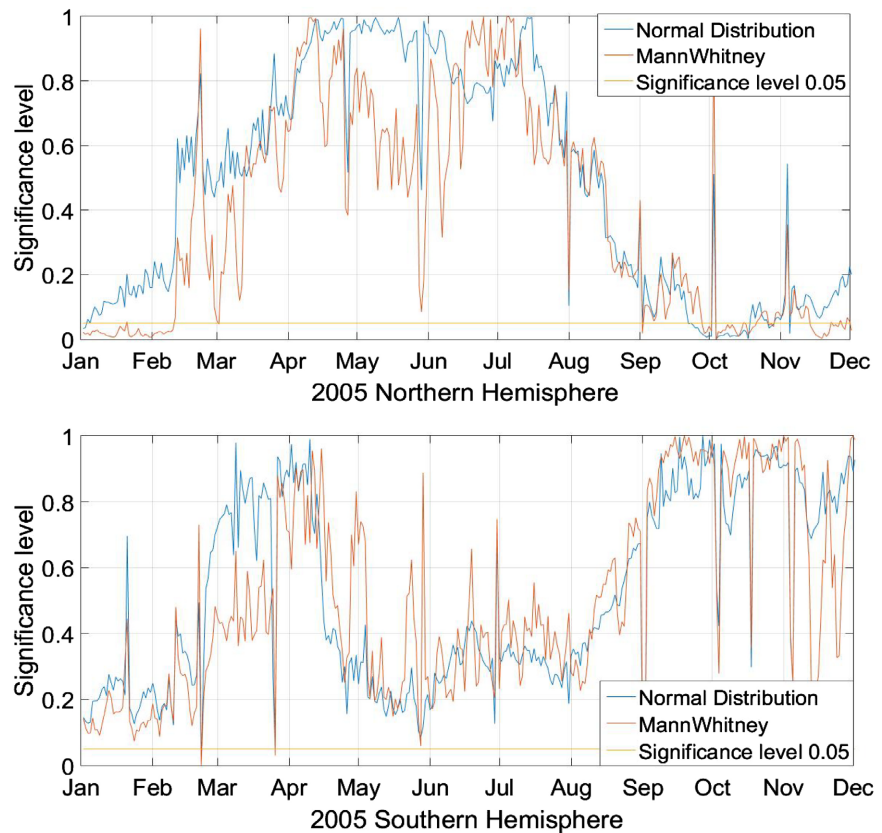
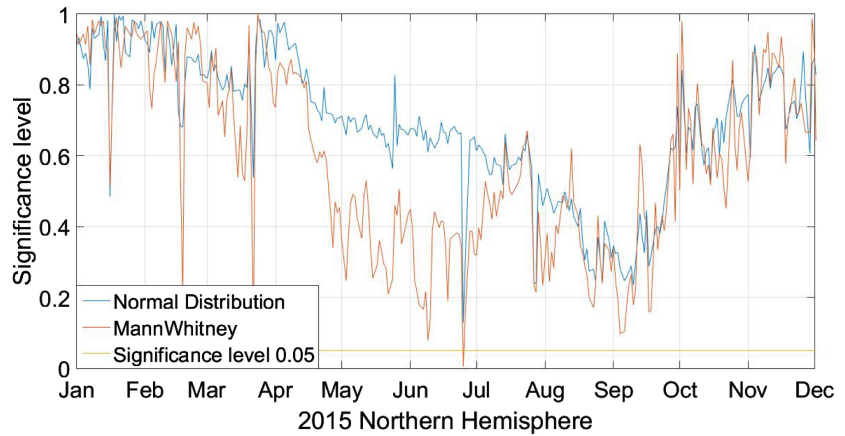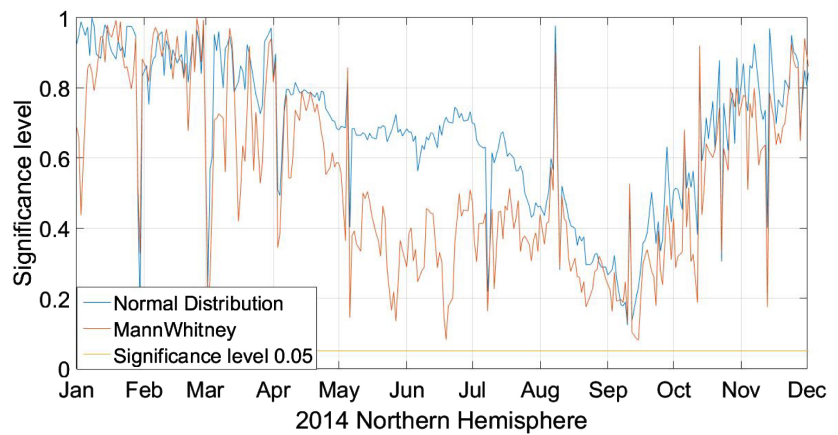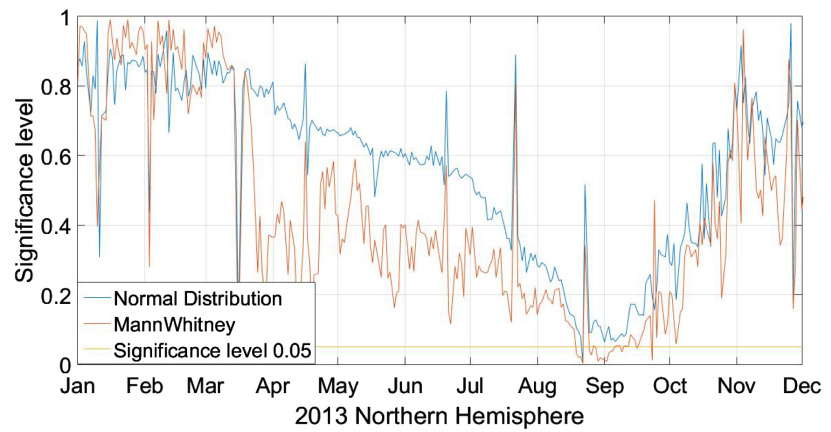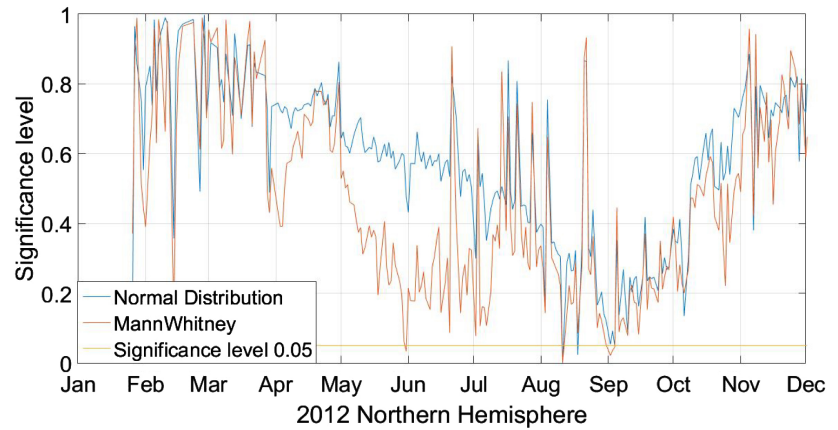**Figure 1.** Hemispheric statistical comparison between TOMS and OMI.

A great variability is observed in both hemispheres using both statistical tests. It is also observed that although the significance levels are not of the same order in both statistical tests, they present multiple coincidences; which are more evident when significance levels abruptly decrease or increase showing peaks. The fact that those peaks coincide on very precise days means that both statistical tests record the same events. However, the peaks do not coincide in date for both hemispheres, which means that there are events that give rise to these disturbances that do not occur simultaneously in both hemispheres.

In the northern hemisphere, there is a coincidence between the TOMS and OMI measurements with a significance level greater than 0.05 most of the year. The comparison of the measurements using the Normal distribution gives value at the significance level of less than 0.05 on some days between mid-September and mid-October, while the Mann-Whitney test shows values less than 0.05 between the months of January and February and some days between mid-September and December.

For the Southern Hemisphere, the significance level is higher than 0.05 for both statistical tests during practically the entire year.

However, the variability throughout the year cannot be explained, nor the causes for which in the Northern Hemisphere there are significance level values < 0.05.

Figure 2 shows the statistical comparison between the OMI and OMPS satellite

2012 Northern Hemisphere

2013 Northern Hemisphere

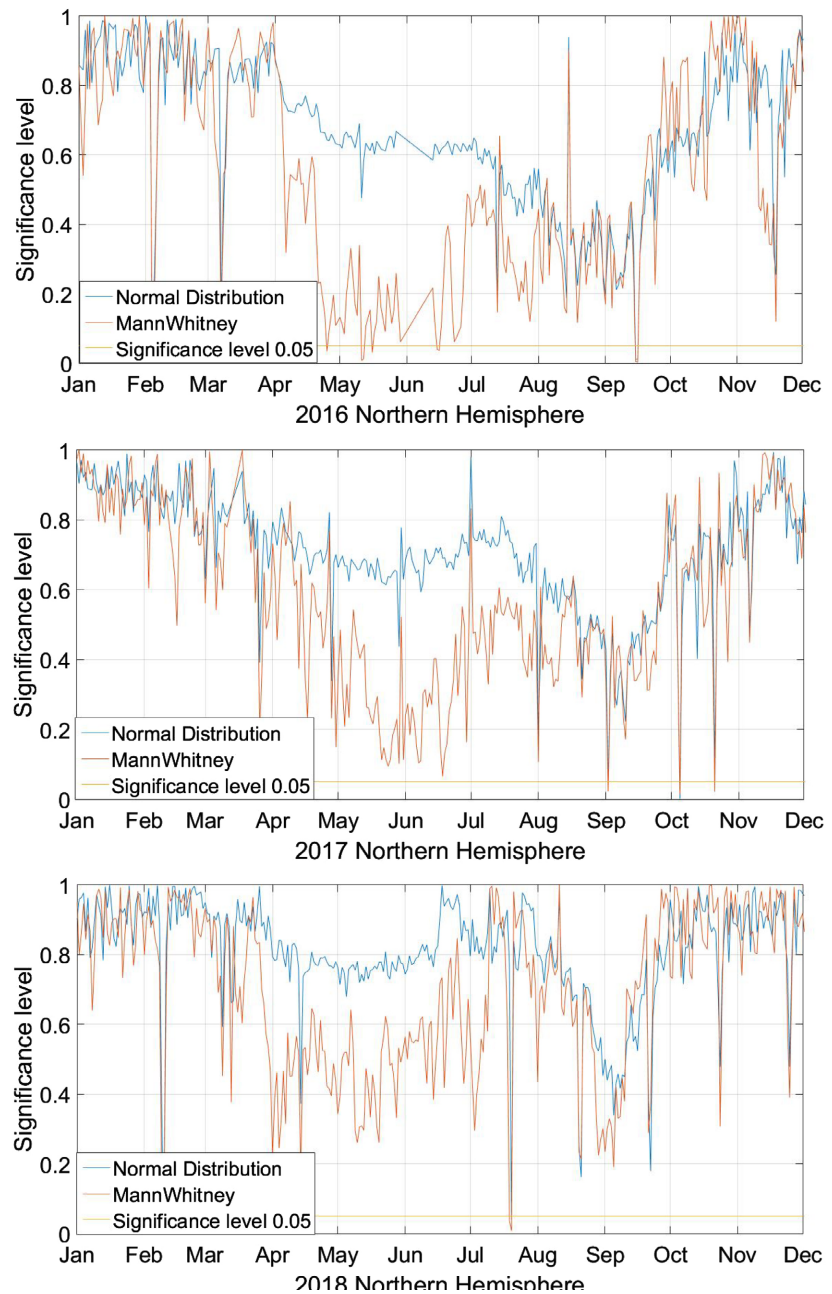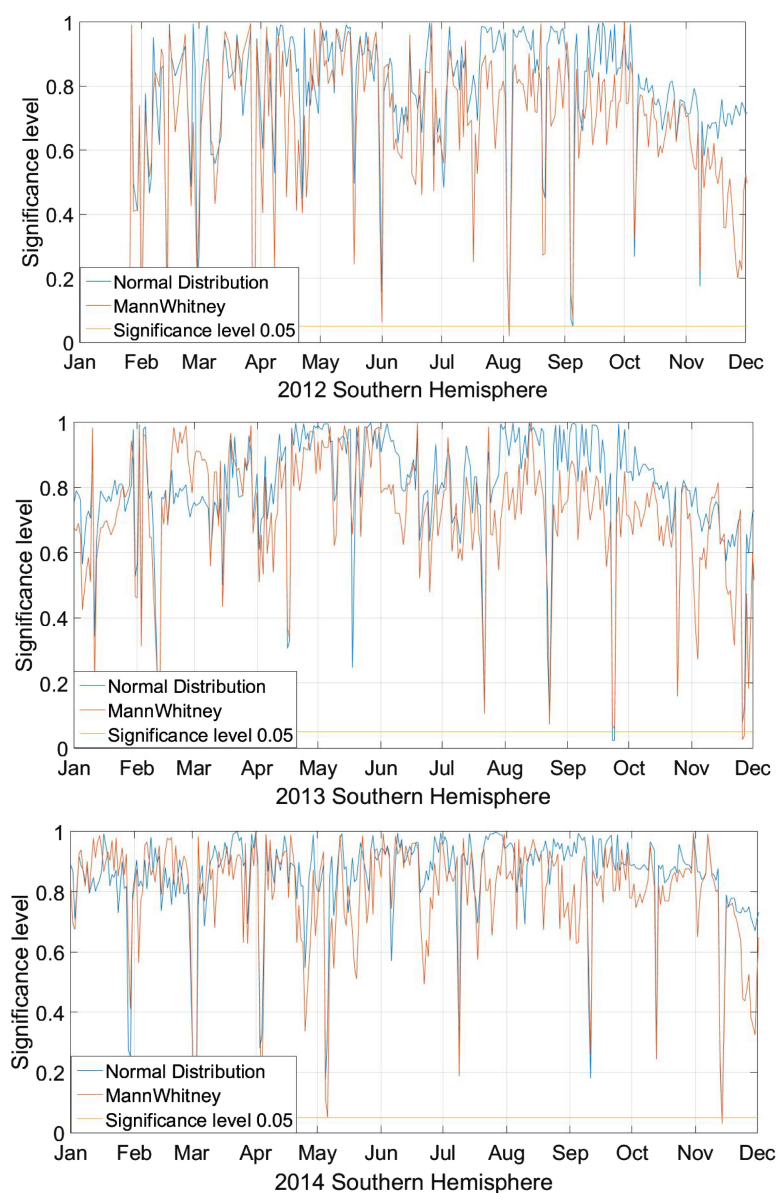2014 Northern Hemisphere

2015 Northern Hemisphere

Figure 2. Statistical comparison of measurements between OMI and OMPS for the Northern Hemisphere.

measurements in the Northern Hemisphere, from 2012 to 2018, year by year. It can be seen that the significance levels are clearly higher than 0.05, which means that the measurements from both satellites are similar with a confidence level higher than 95%. It can also be seen that the comparison between the OMI and OMPS measurements using the normal distribution gives better results than that using the Mann-Whitney test. For both statistical tests in the Northern Hemisphere, it is observed that the values in the significance level gradually decrease until September and then rising again.

Figure 3 shows the comparison between the OMI and OMPS measurements in the Southern Hemisphere, from 2012 to 2018, year by year. It can be seen that the significance level values are clearly higher than 0.05, which means that the measurements from both satellites year after year they are similar with significance levels higher than 95%.

It is observed in **Figure 2** and **Figure 3** that both statistical tests give good significance levels and that the comparison using the normal distribution gives better results. It is also observed in **Figure 2** and **Figure 3**, that both statistical tests present multiple coincidences, which are more evident when the confidence levels increase or decrease abruptly making peaks. These peaks do not appear consistently year after year, which means they are not stationary. But the fact that they coincide on very precise days means that both statistical tests record the same events.
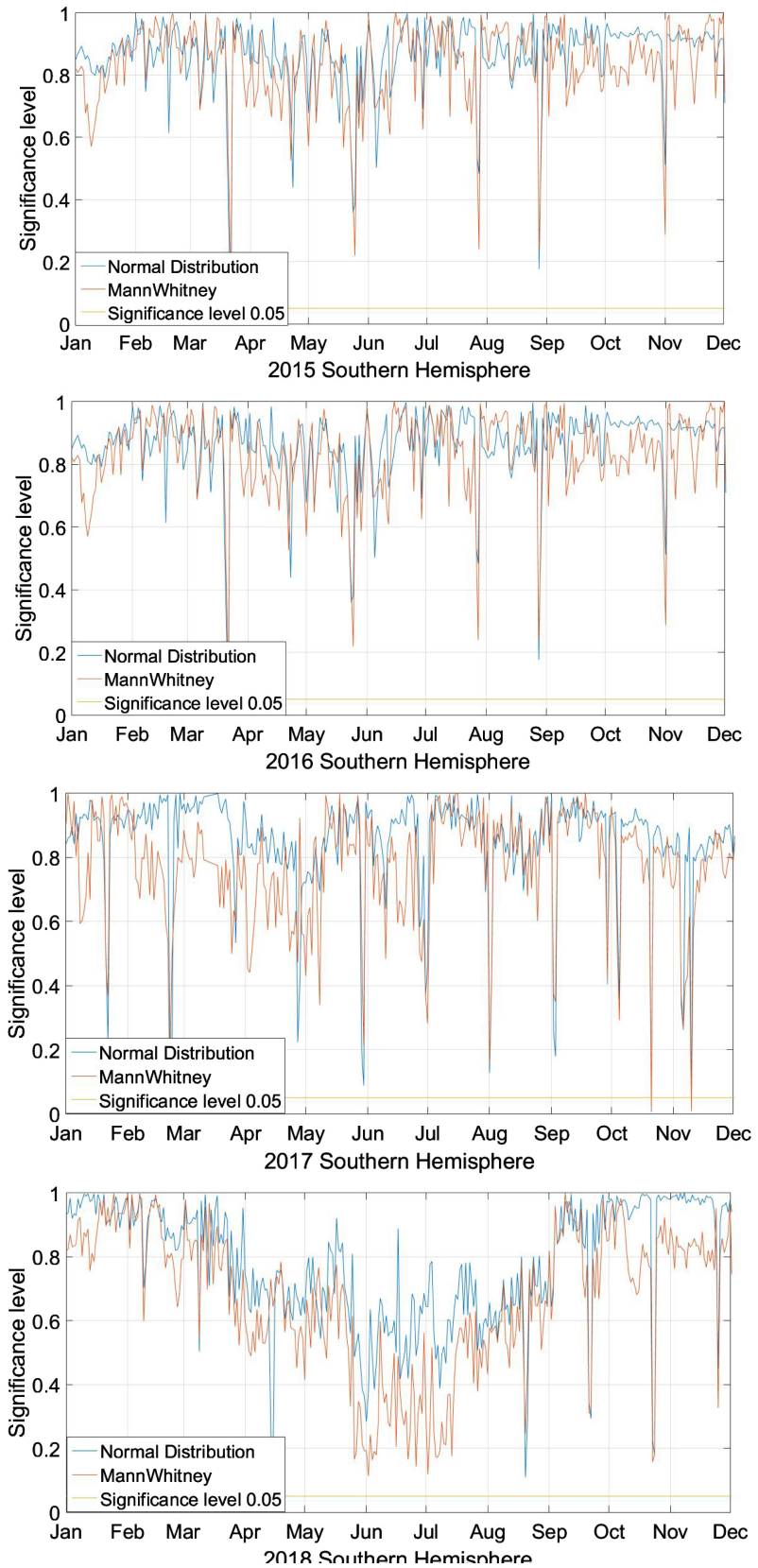
Figure 3. Statistical comparison of measurements between OMI and OMPS for the Southern Hemisphere from 2012 to 2018.

4. Statistical Comparison as a Function of Latitude

In order to investigate the great variability observed, the set of possible statistical differences between satellite measurements needs to be calculated as a function of latitude. Annual average values for the degree of latitude of the TCO satellite measurements are calculated simultaneously and compared using statistical tests, the Normal distribution and the Mann-Whitney test.

Figure 4 presents the statistical comparison between the average values for latitude (degree by degree) of the TCO measurements recorded by TOMS and OMI in 2005, using both the Normal distribution and the Mann-Whitney test.

It can be seen in Figure 4 that in several latitude bands the significance level values are below 0.05. This is evident in the southern hemisphere between the −50˚ and −70˚ latitude band, on the equator between the −15˚ and +20˚ band, and in the northern hemisphere above +45˚ latitude. This implies that for these latitude bands the null hypothesis fails and therefore the measurements of both the TOMS and OMI systems do not observe the same statistical behavior. As OMI had just been put into orbit at that time, it can be considered that the TOMS measurements are the ones that are not reliable, except for the latitude bands between −40˚ and −15˚ and between +20˚ and +40˚. Therefore, the analysis as a function of latitude serves to explain the reason for the enormous variability seen in Figure 1.

Figure 5 presents the statistical comparison between the average values for each degree of latitude of the TCO measurements recorded by OMI and OMPS from 2012 to 2018, year by year, using the Normal distribution and the Mann-Whitney test.

It can be seen that the significance level values are clearly higher than 0.05 for the years 2012 to 2016. This means that with a confidence level higher than 95%, the measurements of both satellites are statistically similar.

It can also be observed that the levels of significance are higher in the Southern Hemisphere, where multiple precise coincidences are observed between both
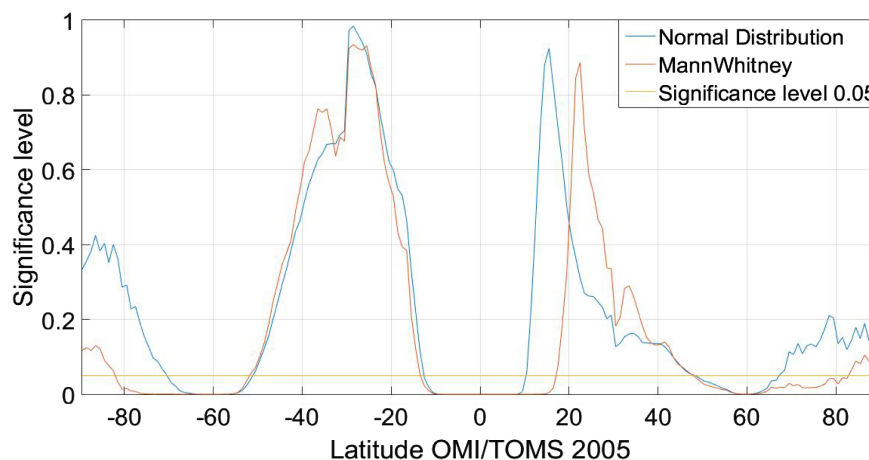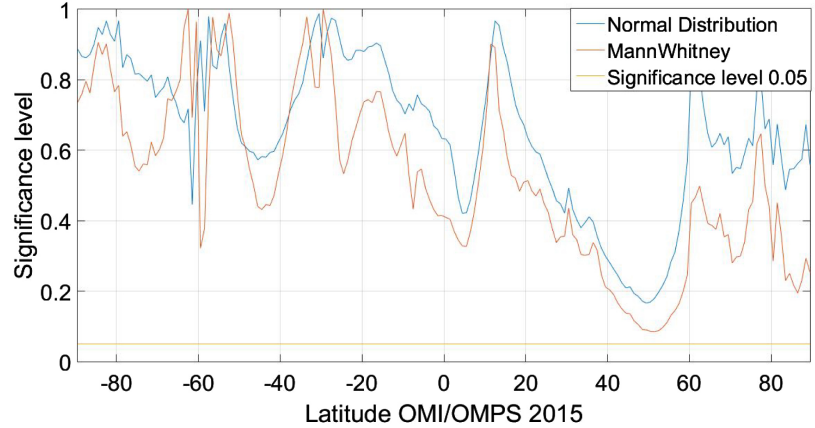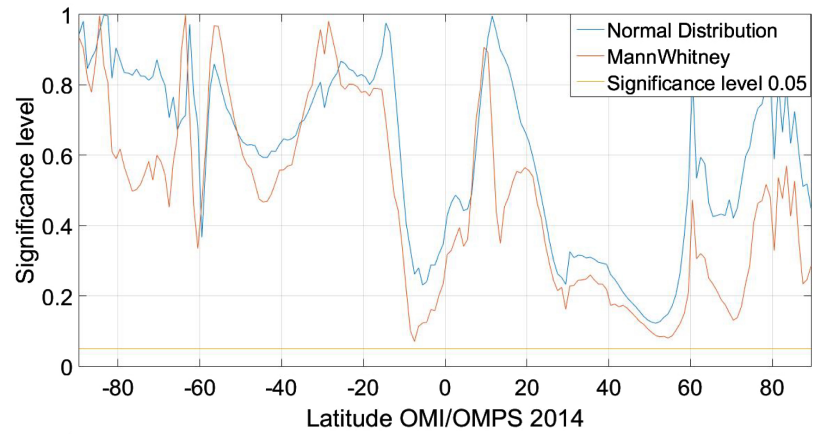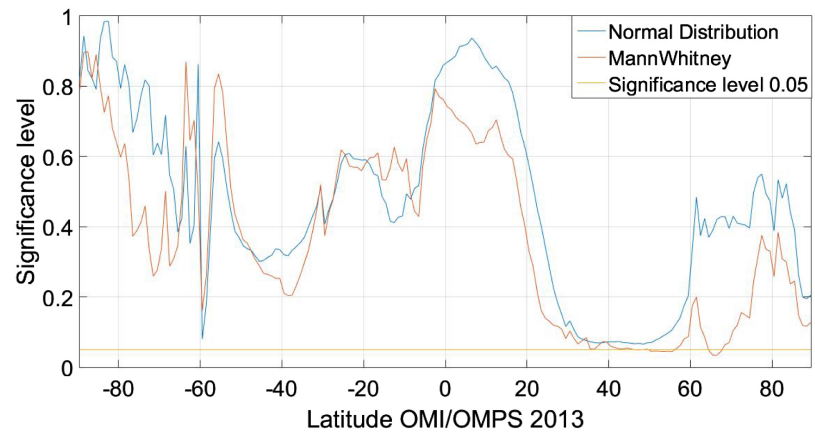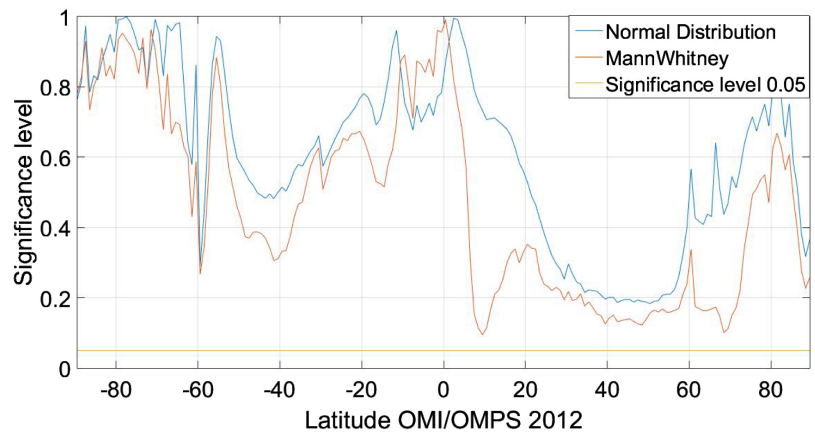

Figure 4. Statistical comparison between the average values by latitude of the TCO measurements recorded by TOMS and OMI in 2005.
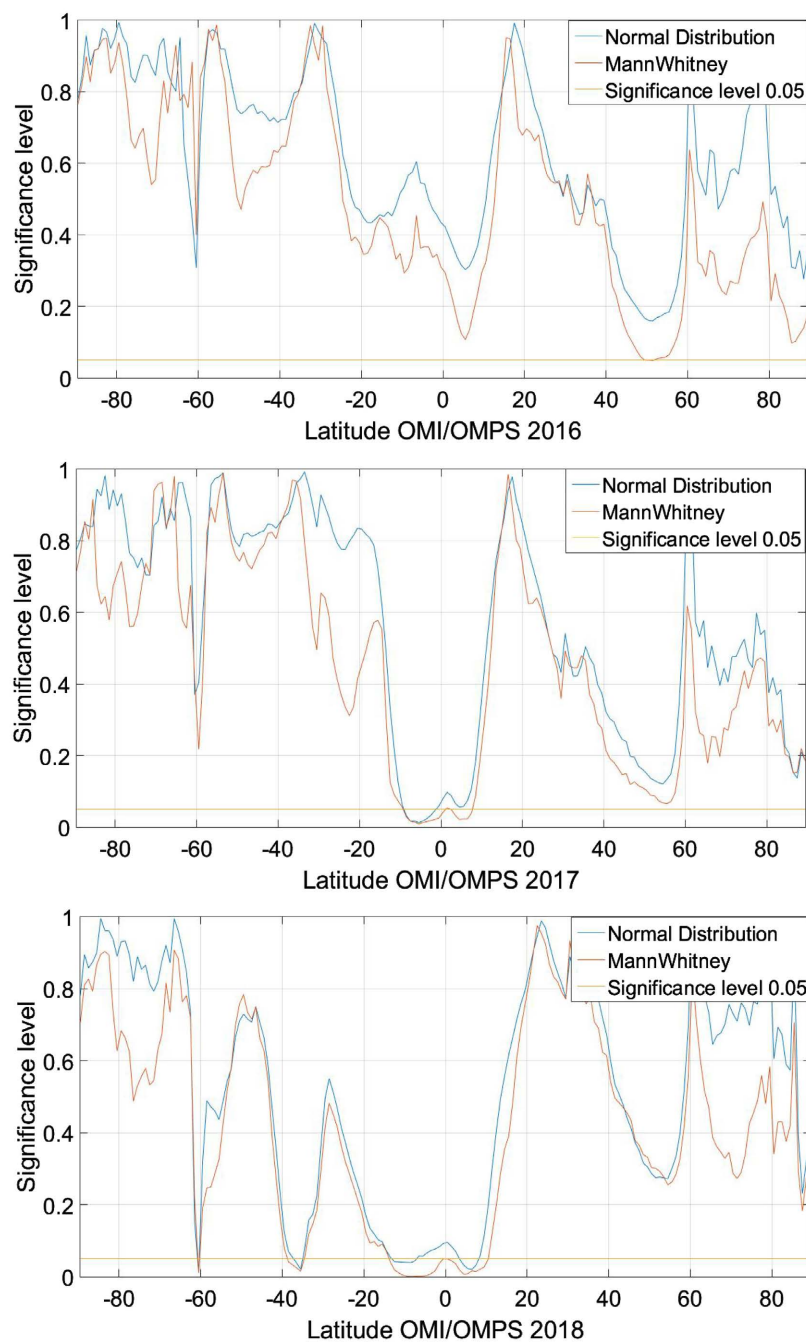
Figure 5. Statistical comparison between the average values by latitude of the TCO measurements recorded by OMI and OMPS from 2012 to 2018.

statistical tests. However, from the equator and towards the Northern Hemisphere the levels of significance are poor, and the graphs of both statistical tests do not coincide in form; which implies that satellites are not recording the same events with the same precision.

In 2017 the significance levels around the equator, between −10° and +10°, are less than 0.05, allowing to state with a confidence level of 95% that in this latitude band there is no coincidence between the measurements of both satellites.

This can be interpreted as one of the measurement systems beginning to fail.

For 2018, in addition to the band around the equator, the levels of significance are less than 0.05. There are narrow bands in the Southern Hemisphere where the levels of significance are also less than 0.05, which can be interpreted as a consequence of the mentioned failure being accentuated.

5. Conclusions

The comparison between satellite measurements of TCO has been carried out by performing two statistical tests. The first one can be considered a modification of the method used to compare two means of independent samples, that we have abbreviated simply as proof of the *Normal Distribution*. The second test was the Mann-Withney test.

An annual hemispheric statistical comparison was tested first. For both hemispheres, both statistical tests seem to give reasonably similar results. However, a great variability is observed in both hemispheres for both statistical tests.

To investigate the great variability that is observed, an annual statistical comparison was tested depending on latitude using both statistical tests. For each degree of latitude from −90˚ to +90˚, the significance level of the difference between the annual average values of the measurements from both satellites was calculated. In the case of TOMS-OMI, bands were found in which the significance level was less than 0.05; which means that, although hemispheric, the statistical comparison seemed acceptable. In detail, not all measurements from both satellites were statistically similar. These results are indicative of a likely failure in 2005 TOMS measurements.

Unlike the OMI-OMPS case between 2012 and 2016, the significance levels were higher than 0.05. Therefore, for this period it can be stated with a confidence level higher than 95% that the measurements of both satellites are similar. However, in the years 2017 and 2018, bands whose significance level was less than 0.05 were found, showing the start of a likely failure in OMI measurements.

The results of both statistical tests were found to be similar, which is proof that the results are consistent. The Normal Distribution test results in higher levels of significance than the Mann-Withney test. There are no elements to ensure that one of the tests is better than the other, it can be concluded that both are acceptable. The interesting contribution of the statistical analysis presented here is that it allows us to recognise a good level of confidence in satellite measurements. At the same time, it tells us that satellite measurements, like any type of measurement, have a certain degree of uncertainty that is difficult to quantify.

The limitation of this work is that the comparison is strictly between satellite measurements. It can be tested if measurements coincide with each other in certain periods of time, but it is not possible to test their accuracy. A priori, we have established that the reference measurements are those of the satellite that comes into operation.

To answer the question that we asked in the introduction of this study: "Can

satellite measurements be trusted?" This work has demonstrated that satellite measurements can be relied upon in terms of reproducibility. The ground station measurements include very few points on the Earth's surface and greatly differ from each other, they serve as indicators of the behavior of satellite measurements, but they cannot be considered infallible a priori in global validations.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Molina, M.J. and Rowland, F.S. (1974) Stratospheric Sink for Chlorofluoromethanes: Chlorine Atom-Catalyzed Destruction of Ozone. *Nature*, **249**, 810-812. https://doi.org/10.1038/249810a0

[2] Farman, J.C., Gardiner, B.G. and Shanklin, J.D. (1985) Large Losses of Total Ozone in Antarctica Reveal Seasonal ClO_x/NO_x Interaction. *Nature*, **315**, 207-210. https://doi.org/10.1038/315207a0

[3] Adcock, K.E., Fraser, P.J., Hall, B.D., Langenfelds, R.L., Lee, G., Montzka, S.A., Oram, D.E., Röckmann, T., Stroh, F., Sturges, W.T., Vogel, B. and Laube, J.C. (2021) Aircraft-Based Observations of Ozone-Depleting Substances in the Upper Troposphere and Lower Stratosphere in and above the Asian Summer Monsoon. *Journal of Geophysical Research: Atmospheres*, **126**, 1-18. https://doi.org/10.1029/2020JD033137

[4] Hofmann, D.J., Butler, J.H., Dlugokencky, E.J., Elkins, J.W., Masarie, K., Montzka, S.A. and Tans, P. (2006) The Role of Carbon Dioxide in Climate Forcing from 1979 to 2004: Introduction of the Annual Greenhouse Gas Index. *Tellus*, **58**, 614-619. https://doi.org/10.1111/j.1600-0889.2006.00201.x

[5] Butler, J.H. and Montzka, S.A. (2020) The Noaa Annual Greenhouse Gas Index (AGGI). NOAA Earth System Research Laboratory, Boulder, CO, USA. https://www.esrl.noaa.gov/gmd/aggi/aggi.html

[6] Keeble, J., Brown, H., Abraham, N.L., Harris, N.R.P. and Pyle, J.A. (2018) On Ozone Trend Detection: Using Coupled Chemistry-Climate Simulations to Investigate Early Signs of Total Column Ozone Recovery. *Atmospheric Chemistry and Physics*, **18**, 7625-7637. https://doi.org/10.5194/acp-18-7625-2018

[7] Ball, W.T., Alsing, J., Mortlock, D.J., Staehelin, J., Haigh, J.D., Peter, T., Tummon, F., Stübi, R., Stenke, A., Anderson, J., Bourassa, A., Davis, S.M., Degenstein, D., Frith, S., Froidevaux, L., Roth, C., Sofieva, V., Wang, R., Wild, J., Yu, P., Ziemke,

J.R. and Rozanov, E.V. (2018) Evidence for a Continuous Decline in Lower Stratospheric Ozone Offsetting Ozone Layer Recovery. *Atmospheric Chemistry and Physics*, **18**, 1379-1394. https://doi.org/10.5194/acp-18-1379-2018

[8] Balis, D., Kroon, M., Koukouli, M., Brinksma, E., Labow, G., Veefkind, J. and McPeters, R. (2007) Validation of Ozone Monitoring Instrument Total Column Ozone Measurements Using Brewer and Dobson Spectrophotometer Ground-Based Observations. *Journal of Geophysical Research*, **112**, D24S46. https://doi.org/10.1029/2007JD008796

[9] McPeters, R., Kroon, M., Labow, G., Brinksma, E., Balis, D., Petropavlovskikh, I., Veefkind, J.P., Bhartia, K. and Levelt, P.F. (2008) Validation of the Aura Ozone Monitoring Instrument Total Column Ozone Product. *Journal of Geophysical Research*, **113**, D15S14. https://doi.org/10.1029/2007JD008802

[10] Lalongo, I., Casale, G.R. and Siani, A.M. (2008) Comparison of Total Ozone and Erythemal UV Data from OMI with Ground-Based Measurements at Rome Station. Atmos. *Chemical Physics*, **8**, 3283-3289. https://doi.org/10.5194/acp-8-3283-2008

[11] Antón, M., López, M., Vilaplana, J.M., Kroon, M., McPeters, R., Bañón, M. and Serrano, A. (2009) Validation of OMI-TOMS and OMI-DOAS Total Column Ozone Using Five Brewer Spectroradiometers at the Iberian Peninsula. *Journal of Geophysical Research*, **114**, D14307. https://doi.org/10.1029/2009JD012003

[12] Mann, H.B. and Whitney, D.R. (1947) On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other. *Annals of Mathematical Statistics*, **18**, 50-60. https://doi.org/10.1214/aoms/1177730491