TOAR-II HEGIFTOM: Description of homogenized Dobson Umkehr free-tropospheric ozone time series

General guideline: Use only short descriptions in the main. Details can be referenced to literature or hyperlinked to sites where the document can be accessed. New documents or material can be included here as an annex.

Availability

Please give the location where the homogenized time series can be obtained (ftp, data archive, website, doi, e-mail address contact person, etc.). Data format (hdf, NASA-Ames, etc.) Put its description in separate document (hyperlink) or annex (possibly refer to code or libraries for accessing the data).

There are several version of Umkehr processing software: the WinDobson is used at NOAA, Martin Stanek software is used at Belsk, and MeteoSwiss is using their own software for Arosa/Davos Umkehr data processing. All have the standard (aka WOUDC) and supplementary output formats.

The homogenized Umkehr ozone profile data for **NOAA stations at Boulder, US; MLO, US; OHP, France; and Lauder, New Zealand** (Petropavlovskikh et al., 2022, https://amt.copernicus.org/preprints/amt-2021-203/) are available from https://gml.noaa.gov/aftp/data/ozwv/Dobson/AC4/Umkehr/Optimized/ as monthly mean time series in the .csv and netcdf formats, and as daily profiles in AMES (WOUDC) format. The netcdf files for daily ozone profiles can be obtained from irina.petro@noaa.gov or peter.effertz@noaa.gov.

The homogenized data including total column ozone, N-values and Umkehr profiles for **Belsk, Poland** can be obtained from jkrzys@igf.edu.pl or bonia@igf.edu.pl. The data format for the Umkehrs is according to input/output used by the UMK04 retrieval. Two output options are possible: short (measured and retrieved total column ozone, the profile, indices describing the quality of the profile) and long (as short output plus N-values, a priori profiles, residuals, AK in 8x8 format). The data format is ".txt".

2 different homogenized Umkehr data records for **D051 Arosa/Davos, Switzerland** are available

1) from eliane.maillard@meteoswiss.ch as monthly mean time series and as daily profile in the .txt format.

2) from the LOTUS website (ftp://Phase-2_2022-2019/Umkehr/optimized/ARO_*) as monthly mean time series in the .csv format.

Data field description

- *Please describe shortly all the data fields (and their units) that are available, also auxiliary data fields.*

  WOUDC format (i.e. *.txt file for NOAA record)

  Daily Umkehr profiles in the WOUDC standard output have the following format:

DD MM YY M/A LAM TO_OB TO_RT LO3(10:1)*100 NUMIT SZA_b SZA_num RMSD(DIF) RMSD(CONV) RMSD(err) STN_num

Example:

30 12 18 1 3 289 2906 147 371 960 1848 3418 6491 7580 4496 2096 1649 3 3 10 0 3 16 67

where

DD is day

MM is month

YY is year

M/A is morning or afternoon (1/2)

TO_OB is observed total ozone (TO)

TO_RT is retrieved TO

LO3(10:1) is 100*ozone amount (DU*100) in Umkehr layers 10, 9,...1 (layer 1 is a double layer 0+1)

NUMIT is number of iterations

SZA_b is the SZA number for the first available measurement (1 is 60, 2 is 65, 3 is 70 etc)

SZA_nub is the number of measurements (12 is the maximum number)

RMSD(DIF) is the root-mean square deviation (RMSD) of the difference of the solution profile from the previous iteration

RMSD(CON) is the RMSD of the convergence of the forcing factor

RMSD(err) is the RMSD of the residual fit (difference between OB and RT N-values)

STN_num station number.

NOAA datasets

Supplementary Daily Umkehr ozone profile files (netcdf format, per request) have output for all quality assured (less than 5 iterations, RMSD(err) less than 100, not negative ozone) profiles. The files contain a collection of profiles for a single year. The following parameters are included: the date (Julian day of the year), time of the day (1 for AM or 2 for PM), total ozone observed (DU), total ozone integrated from profile (DU), pressure for 10 and 16 Umkehr layers (pressure is at the bottom of the layers, mbar), Umkehr profile in 10 and 16 layers (DU), type of observation (i.e. 3 is for Dobson C-pair wavelengths), measurement code (30 means AD zenith measurement with clear sky).

Monthly mean files (.csv) contain the following information: Date (DD/MM/YYYY), Year, Month, Layer 1 through 10 mean ozone (DU), Layer 1 through 10 standard deviation (DU), Count (number of profiles). #N/A indicates missing data.

Belsk dataset

The following data from measurements taken at Belsk, Poland, by the Dobson spectrophotometer serial no.84, are available for the period March 1963-December 2020:

- Total column ozone (DU) – homogenized time series of the intra-day total column ozone measurements. The homogenization procedure accounting for the instrument aging, dependence on ozone absorption on temperature, stray light is described in the paper: Krzyścin, J. W., Rajewska-Więch, B., and Jarosławski, J.: Total column ozone measurements by the Dobson spectrophotometer at Belsk (Poland) for the period 1963–2019:

homogenization and adjustment to the Brewer spectrophotometer, Earth Syst. Sci. Data, 13, 4425–4436, https://doi.org/10.5194/essd-13-4425-2021, 2021

- N-values - homogenized time series based on the Umkehr observations for C pair during clear-sky conditions. The data are coded using standard for the UMK04 input
- Daily Umkehr ozone profiles (DU) – output of the UMK04 retrieval using the homogenized series of total column ozone and N values. The data are coded using standard for the UMK04 output: date (year, month, day), time of the day (1 or 2 for AM or PM), total ozone observed (DU), total ozone integrated from profile (DU), type of observation (i.e. 3 is for Dobson C-pair wavelengths), ozone content in 10 layers (DU), number of N values used the retrieval (max 10 for all possible N values between 70°-90°SZA, and 7 for the so-called short-Umkehr), indices for the profile quality. In addition to standard UMK04 output, the flag is added (see flagging section)

Extended output of the UMK04 retrieval contains also a priori profiles, residuals, AK values in 8x8 format.

Arosa/Davos

Daily Umkehr ozone profiles (DU) are made in 16 pressure-based layers. For each layer record the pressure (mbar), a priori (DU), and AK. Date (Julian day of the year), time of the day (1 for AM or 2 for PM), total ozone observed (DU), total ozone integrated from profile (DU), pressure for 10 Umkehr layers (pressure is at the bottom of the layers, mbar), Umkehr profile in 10 layers (DU), for 2) type of observation (i.e. 3 is for Dobson C-pair wavelengths), measurement code (for (1)clear sky or cloudy sky).

- *Describe the metadata that is available*

  NOAA data

  Monthly mean files have the header. Some station and instrument information is included in the netcdf file description of the fields.

  Further information about datasets and instruments used in homogenized records can be found in Petropavlovskikh et al., 2022, https://amt.copernicus.org/preprints/amt-2021-203/

  Belsk data

  Daily N-values and Umkehr ozone profiles are in standard format used by the UMK04 retrieval. Concerning the total column ozone data, the metadata are included as the data header (txt format).

  Arosa/Davos

  For 1) Some information is included in the file header.

  For 2) Some information is included in the netcdf file description of the fields
- *Are there derived products (e.g. tropospheric columns) available for HEGIFTOM? Where?*

  Both NOAA and Belsk files

  Umkehr layer 1 is representative of the tropospheric layer (surface to 250 mbar).

Description of homogenization procedure

Please provide a short description (few lines) of the steps taken to make your dataset (more) homogeneous/harmonized within the network. Give details in an Annex document or the references of the publications (documents) in which the homogenization procedure of the data is described.

NOAA data

Umkehr ozone profile records are homogenized using simulated ozone record over the station (i.e. M2GMI model) as a reference to remove step changes in observations (N-values) when caused by changes in the stray light contribution or from other instrumental artifacts, or in association with data re-processing before and after instrument calibrations. The reference paper is in the reviews. Petropavlovskikh et al., 2022, https://amt.copernicus.org/preprints/amt-2021-203/

Belsk data

The standard approach is used i.e., homogenization is done using time series of the ratio between N(SZAi)- values (calculated at selected SZAi) and total column ozone. Two steps procedure is proposed. In the first step, the ratio is corrected for a slight decline in the ratio values since the early 2000s. After the first step, the corrected ratio values are examined for a step change in the data. To eliminate step changes in 1990, 1996, 2003 year, constant values are added to the time series. (Annex 1)

Arosa/Davos

1) D051 Umkehr ozone profile record is homogenized using simultaneous and collocated Dobson and Brewer ozone records. Steps changes are corrected on the raw data level (N values) under the express condition that the anomaly is confirmed by a technical issue in the metadata. The reference paper is in preparation (Maillard Barras et al., 2022).

2) D051 Umkehr ozone profile record is homogenized using simulated ozone record over the station (i.e. M2GMI model) as a reference to remove step changes in observations (N-values) when caused by changes in the stray light contribution or from other instrumental artifacts, or in association with data re-processing before and after instrument calibrations. The reference paper is in the reviews. https://amt.copernicus.org/preprints/amt-2021-203/#discussion

Data management

Flagging

- *Data cleaning (outlier removal)?*
  NOAA data
  Yes, only good quality data are provided in the output (i.e. less than 4 iterations, RMSD less than 100, no negative ozone, no missing observations between 70 and 90 degrees SZA)
  Belsk data
  YES, no data with missing observations at 90 degree SZA
  Arosa/Davos
  Yes, only good data are provided in the output (i.e. less than 4(3 for (1)) iterations, RMSD less than 100, no negative ozone, for 2) no missing observations between 70 and 90 degrees SZA
- *Flagging applied? Description of data flags*
  NOAA data
  No flagging
  Belsk data
  YES

1 - good standard Umkehr, all possible N values (10) between 70°-90 SZA, less than 4 iterations, and RMSE <1,
2 -good short Umkehr, all possible N values (7) between 80°-90 SZA, less than 4 iterations, and RMSE <1,
3 - bad Umkehr, number of N values different than 10 or 7 for SZA in the range 70-90 SZA and 80-90 SZA, respectively, or more than 3 iterations, or RMSE >1
Arosa/Davos
For 1) Clear sky day measurements considered: flagging using a nearby UV/VIS lux meter.

- *Data quality indicators?*
  Both NOAA and Belsk data
  No
  Arosa/Davos
  For 1)
  3=clear sky
  5=corrected for cloud effect

Uncertainties

- Which? Distinction random vs. systematic?
  Total uncertainty, generic
- How determined?
  NOAA data and Arosa/Davos
  Rodgers (2000) equations for measurement and smoothing errors, similar to calculations discussed in Bhartia et al, 2013 paper.
  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, 2533–2548, https://doi.org/10.5194/amt-6-2533-2013, 2013.
  Total uncertainties based on synthesized Umkehr data as discussed in Petropavlovskikh et al. (2005, see Figure 2 for monthly mean record and three cases with different normalization SZAs) and in Petropavlovskikh et al. (2022, Figure 1 c)
  Belsk data
  Total uncertainty for each layer was estimated from the statistics of the differences between the morning (am) and afternoon (pm) Umkehr profiles taken during one day (see Table1 in Annex 2)

Traceability

Are the data of an instrument traceable to a reference instrument? Traceable to SI units?

NOAA data

Yes, Dobson 083 instrument is WMO GAW world standard instrument. During intercomparisons, Umkehr observations are compared between the station and standard instrument and results are included in the report. The optical wedge calibration is done with standard lamps (NIST traceable).

NOAA data

All Dobson instruments are compared against either Dobson 083 (MLO, Boulder, Fairbanks) or regional standard (D064 in Hohenpeissenberg, Germany for OHP and Perth, and Dobson standard in Melbourne, Australia for Lauder).

The **Belsk' Dobson** (serial no.84) has been calibrated against standard instrument since 1974 during regular (almost every four year) international inter-comparison campaigns including also calculations of new R-N tables after optical wedge calibrations. The calibrations supported long-term stability of total column ozone measurements (Krzyścin et al., 2021). There were no such inter-comparisons with the Belsk's Dobson concerning measurements of N-values for different SZAs.

Dobson 051 instrument at Arosa is regularly intercompared within the Dobson network wrt two regional secondary standard Dobson instruments (D064 from the Hohenpeissenberg Observatory (MOHp, Germany) and D074 from the Solar and Ozone Observatory in Hradec Kralove (SOO-HK, Czech Republic) (Stubi et al., 2021, https://doi.org/10.5194/amt-14-4203-2021). The optical wedge calibration is done with standard lamps (NIST traceable).

Internal consistency

- *Are the time series of different instruments within the network internally consistent?*

  NOAA data

  Yes, they are processed with the same UMK08 algorithm that is based on the source code (Fortran 77, WOUDC) but also incorporates the stray light correction look up tables (Petropavlovskikh et al, 2011). The Umkehr retrieval algorithm is incorporated in the WinDobson software. The optimization corrections to the N-values are applied outside of the WinDobson software. Corrections for NOAA record homogenization are published in Petropavlovskikh et al. (AMTD, under the review).

  Belsk data

  The source program (in FORTRAN 77) was copied from WOUDC resources.

  Arosa/Davos

  Processed with umk04 (Petropavlovskikh et al, 2005). For 1) No stray light correction.

- *References of intercomparison campaigns of different instruments within the network? Overall conclusions of such studies?*

  NOAA data

  GAW Report, 180. Towards a Better Knowledge of Umkehr Measurements: a Detailed Study of Data from Thirteen Dobson Intercomparisons. Conclusion: individual instruments have different out-of-band light rejection (aka stray light), which can result in relative biases (up to 6 % or larger) between Umkehr retrieved profiles measured simultaneously by several instruments.

  Belsk data

  Comparison between Dobson Umkehrs with those by the collocated Brewer at Belsk was possible for the period 2011-2016 (Annex 3). There was a good correspondence between the profiles by both spectrophotometers as biases (Brewer minus Dobson) are in the range -2.7% (Layer 2) and 2.2% (Layer 7). For Layer 1, bias is -1.7%. and the 10th-90th percentile range is of [-7.1%; 4.7%], i.e. close to the uncertainty range of the Dobson Umkehr in layer 1 (Annex2).

Arosa/Davos

The biases increase in the lower stratosphere and troposphere depending on station: 5% for D051/LKO.

- *World Meteorological Organization (WMO) - WMO, 2008 (WMO/TD-No. 1456) How much inconsistency is left over after homogenization: details in Annex or referencing.*

Under investigation.

External consistency

- *References of intercomparison studies between your technique and other techniques measuring free-tropospheric ozone? Overall conclusions of such studies?*

  NOAA data

  Petropavlovskikh et al (2021) shows that homogenized Umkehr profiles have less than +/- 5 % biases from other techniques (i.e. COH, MLS, SAGE III and ozonesonde) in the stratosphere. The biases increase in the lower stratosphere and troposphere depending on station: -5 % at Lauder, near zero at Boulder and MLO, 5% at OHP.

  Belsk data

  Up to now, there were no intercomparisons of the Belsk's tropospheric ozone by the Dobson spectrophotometer with other techniques.

Data quality indicators

- *Short description or referencing or hyperlinking to a document*

  See Annex 2
- *Factsheet of the performance of the instrument in field operation (only overall specifications, e.g. overall uncertainty xx%) (Table on one page)*

  The uncertainty ranges for the ozone content in Umkehr Layers are of about ± 5% for layers 2-10, and only slightly above this range for Layer 1 [-7.5%; 5.8%). (see Annex 2).

List of homogenized sites (name, geographical location, period of observations)

Table 1. NOAA Dobson Umkehr data information: Name of the station, WMO code, dates of the record (month and year), geolocation of the ground-based stations.

Location	Site Code	Data Record (MM/YEAR)	Latitude	Longitude	Elevation (m)
Fairbanks, Alaska	FBK	03/1984 - 10/2020	64.86 N	147.85 W	133
Haute Provence, France	OHP	09/1983 - 12/2020	43.93 N	5.71 E	685
Boulder, Colorado	BDR	02/1978 - 12/2020	40.02 N	105.25 W	1634
Mauna Loa, Hawaii	MLO	05/1982 - 12/2020	19.53 N	155.58 W	3400
Perth, Australia	PTH	03/1969 - 07/2016	31.92 S	115.96 E	2
Lauder, New Zealand	LDR	02/1987 - 12/2020	45.04 S	169.68 E	370

Belsk, Poland, BEL, 03/1963 – 12/2020 , 51.84 N, 20.79 E, 173

Arosa/Davos, LKO/PMOD , 1/1956- 12/2021 , 46.77N 9.67E/46.81N 9.84E, 1860m/1550m

Belsk data

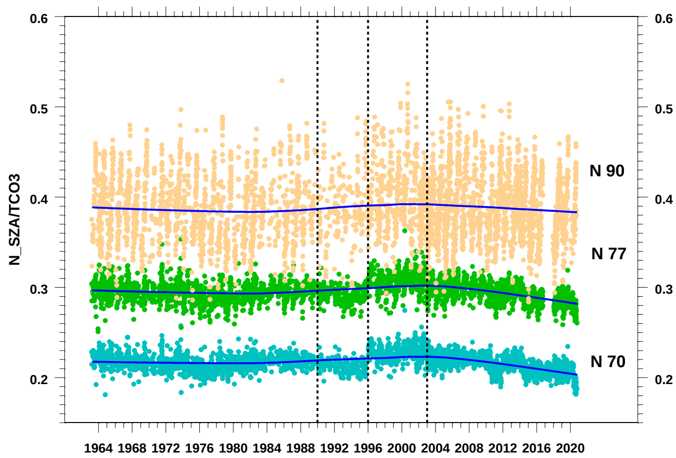
Annex 1 – Homogenization of the Umkehr series

Umkehr retrieval (UMK04) requires total column ozone (TCO) and series of N-values measured at selected solar zenith angles (SZA) ≥70° before and/or in the afternoon. Here, TCO were taken from the re-evaluated TCO intra-day values for the period 1963-2020. The long-term stability of TCO measurements was supported by frequent (almost every 4 years) intercomparisons of the Belsk's Dobson (serial no. 84) with the reference spectrophotometer. The corrections to the original TCO data accounted for less accurate Dobson observations under low solar elevation (mi>3), presence of clouds, and the temperature dependence of ozone absorption (Krzyścin et al., 2021). TCO measured at the moment closest to SZA=70° in the morning and afternoon were used by the Umkehr retrieval. For some days (late autumn-early winter), the noon SZA values at Belsk were above 70°. In such cases, the daily mean TCO values were used both for calculations of am and pm Umkehrs. Previously, corresponding TCO values for the Umkehr observations were calculated manually from the intra-day TCO series based on the Bass-Paur ozone absorption coefficients calculated at the fixed effective temperature of -46.3°C.

The Umkehr homogenization procedure follows standard approach (used in many previous papers) to examine time series of the ratio between N(SZA_i) values (calculated at selected SZA_i for C pair) and the corresponding TCO. The long-term smoothed series of the ratio (for SZA≥70°) should be a trendless line without step changes. Therefore, two steps homogenization procedure is proposed. In the first step, corrections to N values were added to remove a decline found in the smoothed pattern of the ratio that was found in the period 2004-2020 (Figure 1). The differences between the smoothed ratio values and the constant value (smoothed N value in 1963) were subtracted from the raw ratios. After this correction, the resulting smoothed time series of the ratios was close to a trendless line (Figure 2).

Next step of the homogenization of the Umkehr data is adding corrections for step changes in the ratio time series. These were found in 1990, 1996, and 2003. The mean value of the decline from the smoothed pattern was calculated for the period 1990-1995 and 1996-2002. For these periods, the mean values were subtracted from the time series of the ratio obtained after the first step of homogenization. The next smoothed curve of the corrected ratios was calculated for each SZA. Finally, this smoothed curve was close to a constant line and the ratio values were scattered around this line (Figure 3). In the homogenized data , there were no periods of several years in which the ratios were mostly above or below the line as it was observed after the first step of the homogenization (Figure 2).

Figure 1. Ratios between raw N-values (for SZA=70,77, and 90) and total column ozone measured at Belsk for the period 1963-2020 (points). Curve show the corresponding smoothed values by the



LOWESS smoother. Vertical dashed lines mark periods with step changes.

Figure 2. The same as Fig.1 but after elimination of a decline in the ratios for the period 2004-2020 (first step of the homogenization).

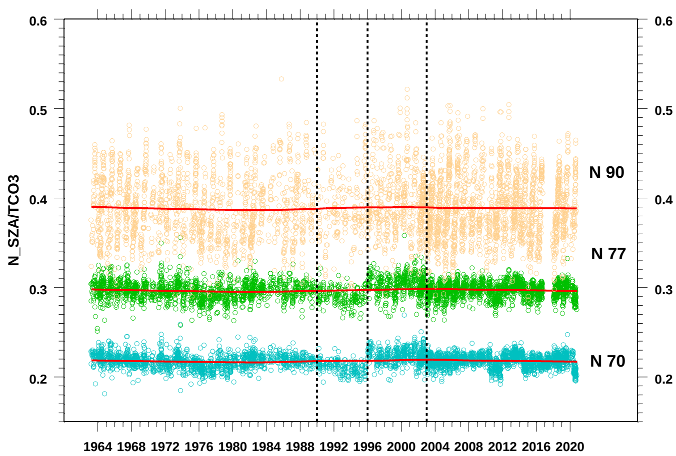
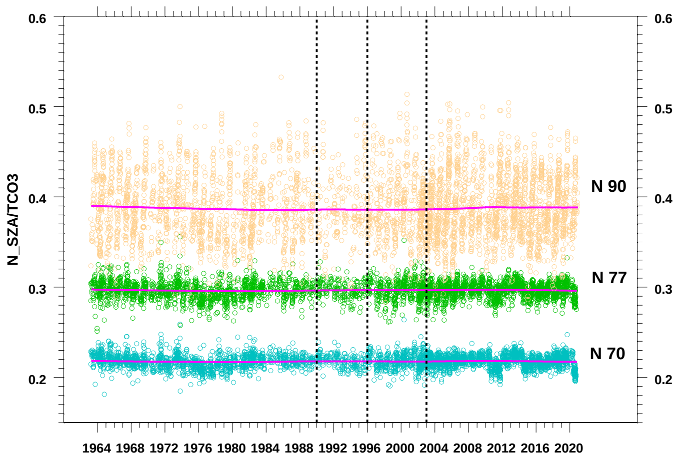


Figure 3. The same as Fig. 1 but the ratios were corrected for step changes in the ratios found after the first step of the homogenization (second step of the homogenization).



Annex 2 - Uncertainty of the Umkehr profile

Total uncertainty in the Umkehr profiles over Belsk is estimated empirically by examining set of differences between am and pm profiles taken during one day. Table 1 shows the difference between am and pm values (for all Umkehr layers and total column ozone) in percent of the daily mean value, i.e. (pm+am)/2. The following statistical characteristics (N=639 for the period 1963-2020) are used, mean value, standard deviation, median, 10th and 90th percentile. It is assumed that the range between 10th and 90th percentile provides the uncertainty. This value probably overestimates the "true" uncertainty as it is possible that the profile could change during the day. However, the Umkehr profiles are taken during perfect clear-sky days with a stable weather conditions and we do not expect also abrupt changes in the stratosphere. It is possible to eliminate days with large intraday changes in the ground-based profiles examining differences in the satellite ozone profile for three consecutive days (*t*-1, *t*, *t+1*) and select only days (*t)* for calculations of the ratio differences when the change between the triad values are below a threshold. This is left for future consideration.

It is supposed that the am and pm profiles should be quite similar as the Dobson total column ozone (last row in Table 1) changes only slightly during a day (i.e., mean=0.2% with the uncertainty range between -0.7% and 1.1%. The uncertainty ranges for all layers are about ± 5%, and only slightly above this range for Layer 1 [-7.5%; 5.8%].

Table 1. Statistical characteristics of the relative differences, Δ, between am and pm Dobson Umkehr and total column ozone measured during one day at Belsk for the period 1963-2020, Δ=(am –pm)/(am+pm)/2 *100% .

Layer	Mean	SD	Median	10th per	90th per
Difference between Ozone Content in the Umkehr Layer					
1	-0.8	5.9	-0.6	-7.5	5.8
2	-0.6	3.8	-0.6	-5.5	3.9
3	-0.1	2.9	-0.1	-3.5	3.3

4	0.6	3.4	0.5	-3.5	4.7
5	0.8	2.1	0.9	-1.6	3.3
6	0.8	2.0	0.8	-1.6	3.3
7	0.7	2.3	0.7	-1.9	3.3
8	0.4	3.6	0.4	-3.9	4.7
9	0.1	3.3	0.2	-3.7	4.0
10	0.0	1.7	0.0	-2.0	2.0
Difference between Total Column Ozone					
	0.2	0.8	0.3	-0.7	1.1

Annex 3 – Comparison between the Dobson and Brewer Umkehr profiles.

The column ozone monitoring with the Brewer spectrophotometer serial no. 64 (BS64) mark II (single monochromator) was launched at Belsk in 1991 but the Umkehr observations began in 2010. The full series of the Brewer Umkehrs (2010 -2021) is under construction. At the moment, a comparison with the concurrent Dobson data is available for the period 2011-2016 (N=328 inluding am and pm Dobson-Brewer pairs). Statistics of the relative differences between Brewer and Dobson Umkehrs and total column ozone is shown in Table 2. There is a good agreement between the instruments. Standard deviations and the uncertainy ranges (10th-90th percentile of the Brewer-Dobson differences) are similar to those for the relative differences between am and pm Dobson profiles for one day (Table 1). This allows to determine a universal uncertainty range pertaining the Umkehr retrieval for any ground-based spectrophotometer measuring N-values regardless of its type.

Table 2. Statistical characteristics of the relative differences, Δ, between Brewer and Dobson ozone content in Umkehr Layers and the column amount of ozone measured simultaneously by both instruments at Belsk for the period 2011-2016, Δ= (Brewer-Dobson)/(Brewer+Dobson)/2 *100%

Layer	**Mean**	**SD**	**Median**	**10th per**	**90th per**
Difference between Ozone Content in the Umkehr Layer					
1	-1.7	5.2	-2.1	-7.1	4.7
2	-2.7	3.9	-2.9	-6.9	2.4
3	-0.1	3.1	-0.2	-3.7	3.7
4	0.3	2.8	0.2	-3.0	3.4
5	0.0	2.0	0.1	-2.4	2.4
6	0.8	2.0	0.9	-1.8	3.1
7	2.2	2.7	2.0	-1.1	5.8
8	2.0	3.4	1.8	-1.7	6.3
9	1.4	3.6	1.4	-2.6	5.4
10	0.7	2.0	0.7	-1.6	3.1
Difference between Total Column Ozone					
11	0.0	1.0	0.0	-1.3	1.0