

ALGOM WP4: Level 1c studies

Cédric Tétard and Didier Fussen

28-Jun-2016

WP 4: Development of a level 1c product of averaged spectra for minor species retrieval

Contents

1	Introduction	2
2	Averaged transmittances climatology	2
2.1	Virtual occultations	2
2.2	Preliminary step	2
2.3	Statistical analysis of the transmittance data set	4
2.3.1	Detection of mono/bimodal distributions	4
2.3.2	Outliers detection	4
2.4	Wavelength shift	6
2.5	Computation of weighted median transmittance	6
3	Description of the climatology	7
4	A preliminar bimodal case study	9
5	Conclusions	11
6	Annex	12

1 Introduction

In the frame of the ALGOM project, BIRA/IASB is involved in the WP4 “Development of a level 1c product of averaged spectra for minor species retrieval”. The goal of this work package is to develop a data processing model and to build a climatology of GOMOS averaged transmittances (also named below ‘virtual transmittances’) as a time-latitude series covering the entire duration of the mission (121 months from 04/2002 to 04/2012). Previous scientific studies have shown that by averaging transmittance signals it is possible to retrieve some minor but important constituents like sodium and OCIO from GOMOS data.

This document describes how the climatology has been built and presents some results.

2 Averaged transmittances climatology

2.1 Virtual occultations

Because of the weakness of the radiation source, the main drawback of instruments using stellar occultation method is the low signal-to-noise ratio (SNR). Thus, some absorbing species can not be retrieved using a single occultation whereas they are of great scientific interest. This is the case of OCIO ([2]), a proxy for the ClO abundance or atomic sodium (used for adaptive optics). A method called ‘virtualization’ is developed to increase the SNR and thus retrieve these species. The principle of virtualization is to combine several measurements spatially and temporally collocated in a time-latitude bin to obtain a “virtual” measurement, which is representative of this bin. The SNR of the virtual measurement is then sufficient to retrieve the species cited above and may be some others (hopefully).

A spectacular demonstration of the effectiveness of the virtualization method is shown in Fig.1. It clearly shows how the sodium doublet (589.16 nm and 589.76 nm) is emerging when the number of binned occultations is increased. As a consequence, it is possible to retrieve the concentration of sodium in the mesosphere from GOMOS virtual measurements ([1]).

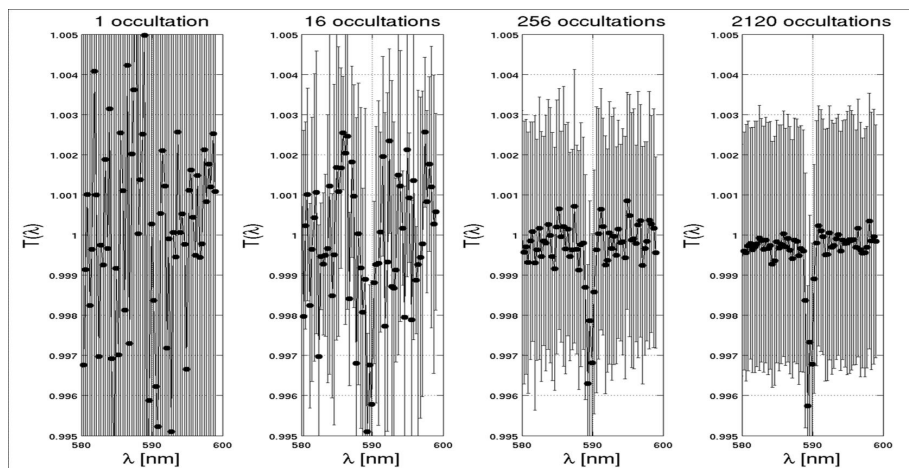


Figure 1: GOMOS weighted median transmittances as a function of the number of occultations used in the bin 70°N - 80°N in January 2005.

2.2 Preliminary step

Firstly, we have to state the criteria for the selection of single GOMOS measurements used to build the bins. We combine measurements co-located in 10° latitude bands with a temporal resolution of one month. This leads to a total of 121*18 bins for the entire GOMOS mission. Added to this, a distinction concerning the illumination conditions is done: we separate measurements performed

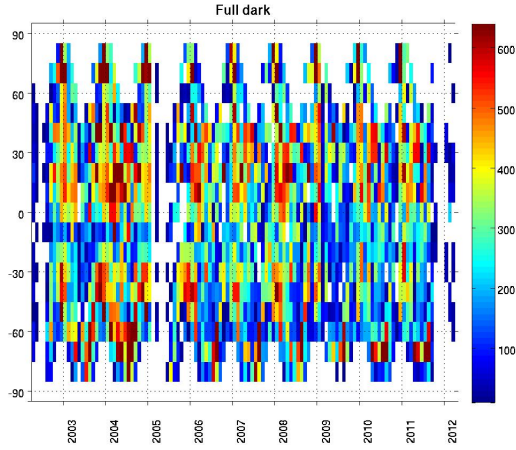


Figure 2: Monthly number of full dark limb occultation in April 2002 - April 2012 as a function of time and latitude.

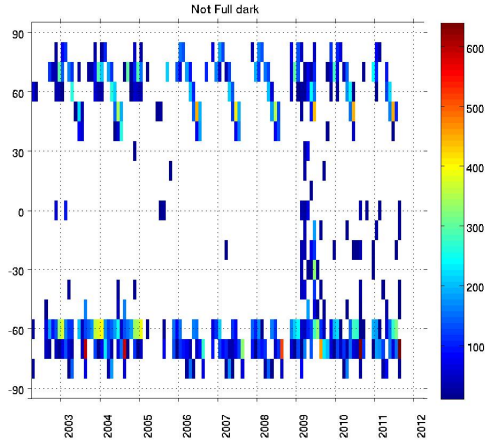


Figure 3: Same as Figure 2 for limb occultations with $SZA > 104^\circ$ and not bright limb conditions.

in full dark conditions ($SZA > 104^\circ$) from the others (with the restriction not to use bright limb measurements whose quality is known to be poor). Thus two climatologies will be built. Figures 2 and 3 show the distribution of GOMOS measurements used in this study. Note that both climatologies present a good coverage of winter poles and are missing coverage of summer poles.

Even if the quality of a single measurement depends on the occulted star, no distinction of stars will be done to build the bins because the combined measurements will be weighted with the inverse of their respective uncertainties.

Finally, for each bin, we build a transmittance data set:

$$[T_1(\lambda_i, z_j), T_2(\lambda_i, z_j), \dots, T_n(\lambda_i, z_j)]$$

where n is the number of individual GOMOS measurements in the bin, λ_i is the wavelength and z_j is the tangent altitude. Note that all the single transmittance spectra are interpolated onto a common vertical grid (from 0 to 100 km with a 1 km step).

Input data used are the transmissions corrected for scintillation and dilution effects and the covariance function of the transmission (GOM_EXT files).

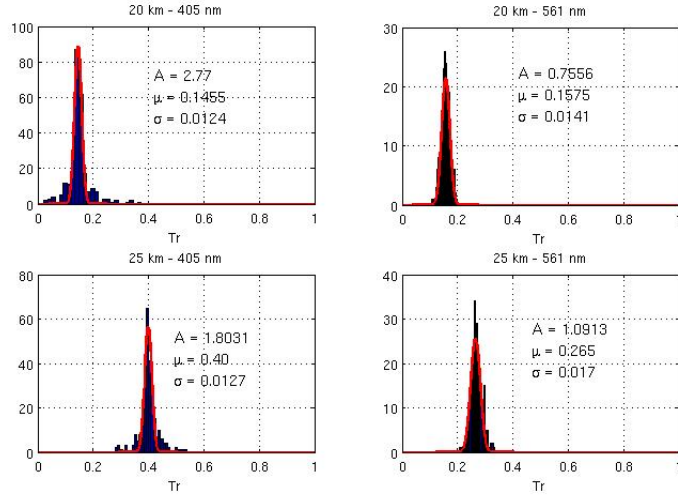


Figure 4: example of unimodal distributions (Sept 2002; [40°N-50°N]) for different altitudes and wavelengths.

2.3 Statistical analysis of the transmittance data set

2.3.1 Detection of mono/bimodal distributions

The method of virtualization requires a thorough statistical analysis. The homogeneity of the data set being of primary importance, a detection of multimodal distributions is performed by fitting the distribution with theoretical models (using the curve fitting matlab toolbox). Figure 4 shows an example of GOMOS transmittance distributions for a unimodal distribution suitably fitted by gaussian models (parameters are given in the legend of the plot).

An example of bimodal distributions is presented in Fig. 5. Each bimodal case will be studied separately by trying to understand why this particular bimodality and trying to properly separate each mode. For the example of Fig. 5, the bimodality is due to the presence of the vortex edge in this latitude band. Therefore, some measurements are inside the polar vortex and others are outside. All the bimodality case observed during the entire GOMOS mission are all in high latitude regions during winters.

Table 1 lists all the case of bimodality observed.

2.3.2 Outliers detection

The second step of the statistical analysis is the detection and rejection of outliers in each data set. An algorithm has been developed based on the well known Jackknife method. This algorithm provides very good results: almost all the outliers are detected (Fig. 6) but the CPU time required to test all the data set in one bin is very important. Indeed, in each bin, we have to apply this algorithm at each tangent altitude (101 levels) and each wavelengths (2336 for all the spectrometers).

The Jackknife method is quite simple and easy to implement (the matlab routine used is presented at the end of this document). Consider a dataset with N values. Remove the first value and compute the standard deviation based on the data subset with $(N-1)$ elements. Repeat this for the N values and you have a new data set of N standard deviations (bottom panel of Figure 6 shows these standard deviations). It appears that when you remove an outlier, the standard deviation decrease strongly. Then, you have to fix a threshold (solid lines in the bottom panel of Figure 6) to detect outliers. Based on visual analysis of several data set, the threshold has been fixed to the median of the distribution minus 3 standard deviations.

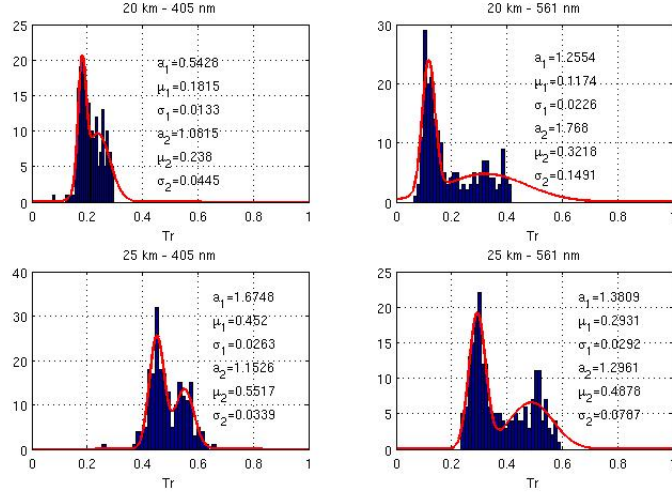


Figure 5: example of bimodal distribution (Sept 2002; [60°S-70°S]) for different altitudes and wavelengths.

year	month	latitude band	year	month	latitude band
2002	08	60°S - 70°S	2005	08	50°S - 60°S
2002	09	60°S - 70°S	2005	08	60°S - 70°S
2003	01	70°N - 80°N	2005	09	60°S - 70°S
2003	01	60°N - 70°N	2006	01	70°N - 80°N
2003	02	70°N - 80°N	2007	01	70°N - 80°N
2003	02	60°N - 70°N	2007	08	50°S - 60°S
2003	03	70°N - 80°N	2008	01	70°N - 80°N
2003	03	60°N - 70°N	2008	09	60°S - 70°S
2003	09	60°S - 70°S	2008	10	60°S - 70°S
2004	01	70°N - 80°N	2008	12	70°N - 80°N
2004	01	60°N - 70°N	2009	01	80°N - 90°N
2004	02	70°N - 80°N	2009	01	70°N - 80°N
2004	02	60°N - 70°N	2009	09	50°S - 60°S
2004	03	70°N - 80°N	2010	01	70°N - 80°N
2004	03	60°N - 70°N	2010	09	60°S - 70°S
2004	09	60°S - 70°S	2011	01	70°N - 80°N

Table 1: list of bimodal cases observed.

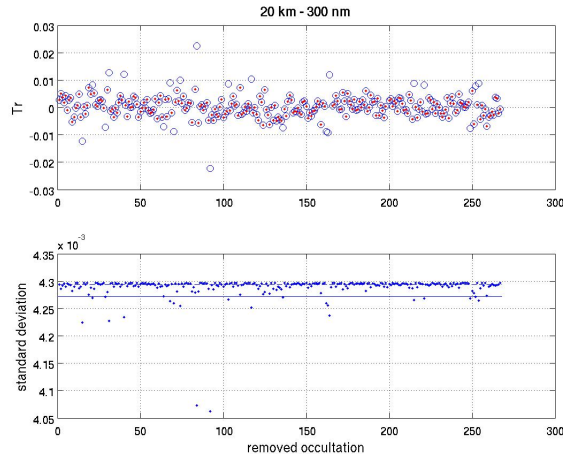


Figure 6: Illustration of the Jackknife method used to detect outliers for the 267 events in the bin [01/2006; 0°-10°N]. Top panel: blue circles = all transmittances at 20 km and 300 nm - red points = remaining transmittances after outliers correction. Bottom panel: distribution of the standard deviations, solid line is the threshold used (see text).

2.4 Wavelength shift

A problem due to slight variations of the wavelength assignments has been encountered for spectrometer B1 and B2. This problem induces a shift in wavelength of some absorption peaks characteristic of O_2 and H_2O . This shift may differ slightly from one measurement to another so much so that we need to realign properly each single transmittance spectra before combining them. It is done by applying a convolution of the shifted spectrum with a reference spectrum (occultation of sirius at the beginning of the GOMOS mission).

2.5 Computation of weighted median transmittance

Finally, for each tangent altitude and each wavelength, the weighted median transmittance is calculated instead of the usual mean because the median is known to be more robust against the presence of residual outliers. Furthermore, as the SNR of different measurements can differ considerably, we combine transmittances weighted with respect to the inverse of their estimated measurements errors.

A weighted median calculation starts by sorting the transmittance values in increasing order, and rearranging the associated weights in the same fashion. The cumulative distribution of these weights is subsequently evaluated. The weighted median is then the transmittance value corresponding to the 50% level of this cumulative weight distribution. The matlab code used is presented at the end of this document.

The uncertainty associated with the weighted median is calculated as the weighted median absolute deviation (WMAD). It is the median of the absolute value of the difference between each transmittance value and the weighted median. It represents in particular the spatial variability due to sampling.

Figure 7 shows an example of averaged transmittances for the bin [70°-80°S] in August 2003.

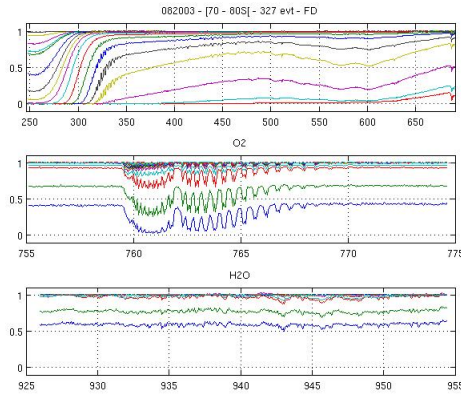


Figure 7: Example of averaged transmittance spectra for spectrometer A (top panel), spectrometer B1 (middle panel) and spectrometer B2 (bottom panel).

3 Description of the climatology

The GOMOS virtual climatology consists of 18 averaged spectra per month, corresponding to the 18 latitude bands. In practice, the all latitude bands are not covered (see Figure 2).

We show in Figure 8 averaged spectra (spectrometer A) for the year 2003 in the latitude band $[50^{\circ}\text{S} ; 60^{\circ}\text{S}]$ for tangent altitudes from 15 km to 50 km with a 5 km step. Several spectral features are clearly observed. In particular, the signature of ozone absorption in the Chappuis band (around 590 nm) and in the Hartley-Huggins band (below 400 nm). It is also interesting to note that any traces of residual scintillations are vanished.

Figure 9 and Figure 10 present the same climatology but for spectrometer B1 (755 to 774 nm) and spectrometer B2 (926 to 954 nm). The spectral signature of O_2 and H_2O are clearly visible.

The GOMOS virtual transmittance climatologies (Full dark and not full dark) are stored in directories like:

```
/FD/yyyy/mm/
/NFD/yyyy/mm/
```

where yyyy stands for the year (from 2002 to 2004) and mm for the month

In each of this directory, one can find at least 18 matlab files corresponding to the 18 latitude bands.

These files are named, for example, bin_042002_00_10N_FD.mat for the bin $[0^{\circ} - 10^{\circ}\text{N}]$ in april 2002 in full dark condition.

Each of these files consists in:

- Tbin: a matrix containing the transmittance values for each tangent altitude and wavelength (size= 101*2336).
- dTbin: a matrix containing the associated uncertainties (size = 101*2336).
- FPbin: a matrix containing the averaged transmittance values calculated using photometers signals (size = 101*2).
- dFPbin: a matrix containing the associated uncertainties (size = 101* 2).
- zin: a vector with 101 elements corresponding to the altitude levels (in km).
- wvl: a vector with 2336 elements corresponding to the wavelength grid (in nm).
- L: a vector containing the list of the single occultation list to build the averaged measurement (for a description of each of these occultation, see the file GomosList.mat hereafter).

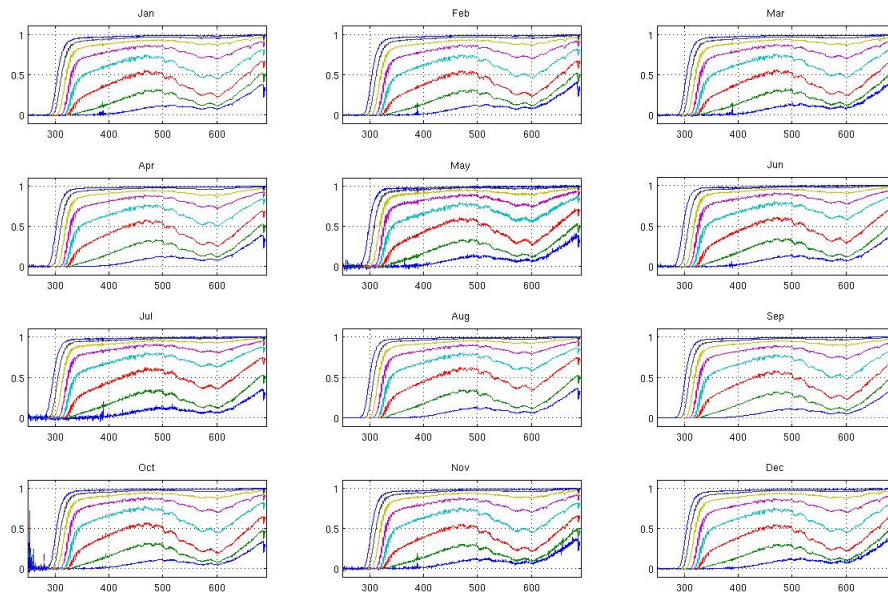


Figure 8: Climatology of averaged spectra for the year 2003 in the [50°S - 60°S] latitude band for the spectrometer A [248 nm-690 nm] from 15 to 50 km (5 km step).

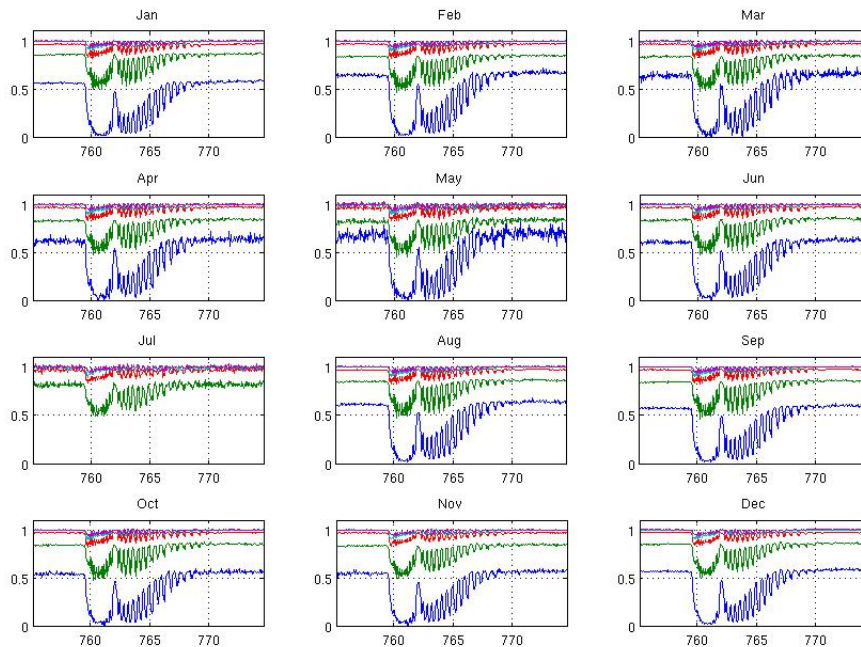


Figure 9: Same as Figure 8 but for spectrometer B1.

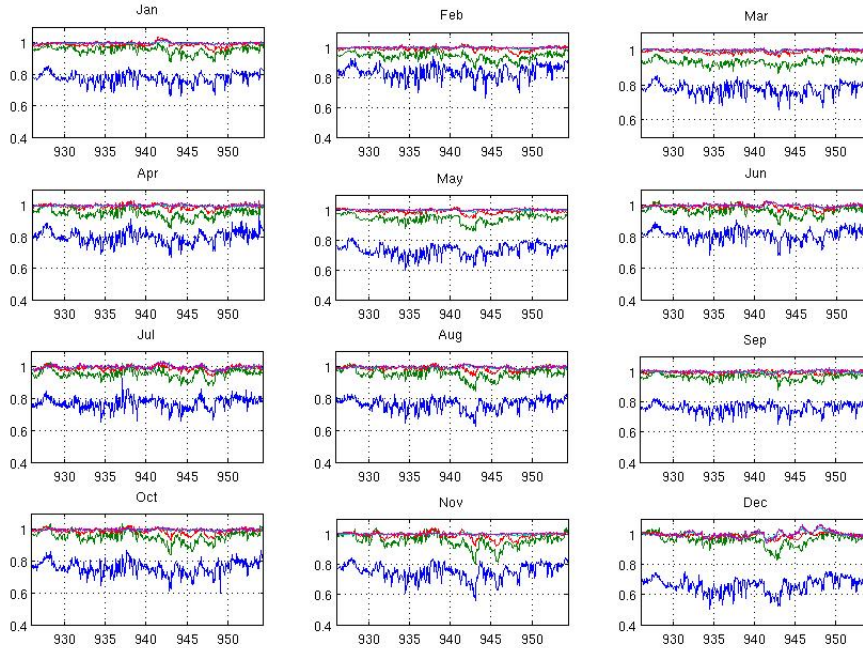


Figure 10: Same as Figure 8 but for spectrometer B2.

- StarOcc: contains some details and statistics about the stars used (for each star used: magnitude, effective temperature, number and percentage of occultations using this star in this bin).

The size of each file is about 3 Mbytes for a total size of 6.4 Gbytes. In the case of bimodal distribution, the files are almost identical except that there are two matrix for the transmittance (Tbin1 and Tbin2), for the uncertainties and for the list of occultations used in each mode.

The data set can be downloaded from the FTP site: <ftp://algom@ftp-ae.oma.be>. A password is required: *gomosAlgom*

The data format is Matlab binary files. For users not having MatLab, it is possible to convert them with scientific Python (*scipy*) or the “*matio*” library on Sourceforge.

Alternatively, it might be easier to make use of Octave (a free GNU MatLab clone), to analyze and/or to convert the data.

The authors wish to be associated to scientific publications making use of the data set.

4 A preliminar bimodal case study

In this section, we present a preliminary case study of a bimodal distribution of transmittances. This case occurred in September 2002 in the [60°S - 70°S] latitude band. A total of 213 GOMOS full dark occultations are combined together. The first step of the statistical analysis (see previously) has warned us that we were dealing here with bimodal distributions of transmittances. Therefore, this case has to be studied separately and we present here what it is possible to do with such bimodal distributions. The first step is to separate the 2 modes. The figure 11 shows the distribution of transmittance at 25 km and 561 nm for this bin. The bimodality is obvious and it also appears for other couples of tangent altitudes and wavelength. The parameter of the best fit (amplitudes of the modes, means and variances) are given in the legend of this figure. Based on this, we have separate the 2 modes using a transmittance threshold at 0.4. Of course, we have checked that this is coherent with several other couples of altitudes and wavelengths. Once the two modes are

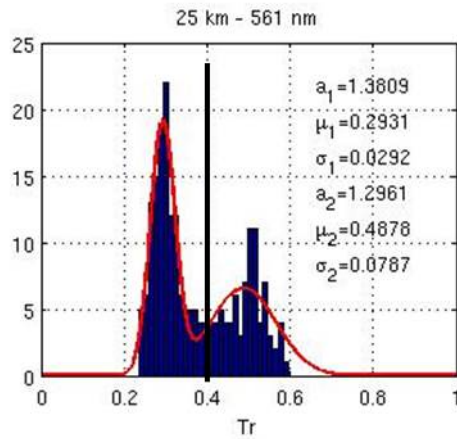


Figure 11: Histogram of the transmittance at 25 km and 561 nm for the bin [September 2002; 60°S - 70°S]. The red line is the best fit. The values of parameters are given in the legend (amplitudes of the mode, means and variances). The black line is the threshold used to separate the 2 modes.

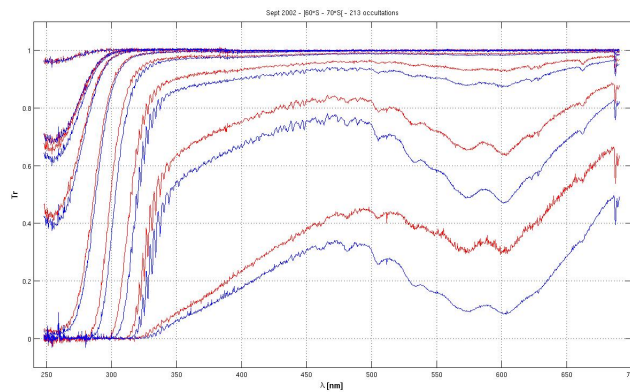


Figure 12: Averaged transmittance spectra for the two modes of the bin [September 2002; 60°S-70°S].

separated clearly, we can compute averaged transmittance spectra correspondig to this two modes (Figure 12).

Then it is interesting afterwards to try to understand the reasons of the bimodal distribution of the transmittances. For this case, the responsible is probably the polar vortex. Indeed, you can see in the figure 13 that the occultations used to bin the mode with the highest transmittance values (in red) are located between about 150° west and 50° east (with some exceptions). It is well known that the air inside the polar vortex are poor in NO_2 compared with the one outside. Furthermore, inside the polar vortex, the air is older than outside. As a result, the air density is smaller inside the polar vortex. These two effects can explain why the transmittance inside the polar vortex exceeds the one outside. To validate this explanation, it only remains to check the position of the polar vortex.

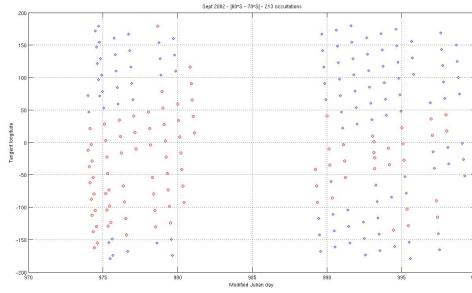


Figure 13: Longitude of the tangent points of the occultations used to build the two modes of the bin [September 2002; 60°S-70°S].

5 Conclusions

This technical note explains how the GOMOS averaged transmittances spectra have been build. The great quality of this product appears clearly and it can be useful for many studies. We can quote for example the retrieval of some metallic species in the mesosphere (such as Na or Mg). It is also possible to compare climatology of ozone (for example) building using averaged occultations with the one build using the operational product. Many other studies can be considered where a direct retrieval (from averaged transmittances and look-up tables) of small constituents is the most efficient method, at the edge of the instrument capacities.

Presently, years 2002 to 2005 have been fully processed and validated. It is expected to complete the full mission in the forthcoming months.

In addition, a super bin climatology will be built. It will consist in combining all the years (from 2002 to 2012) to create a reference year. This super bin climatology could be considered as a reference for low Earth orbit (LEO) instruments.

References

- [1] D. Fussen, F. Vanhellemont, C. Tétard, N. Matashvili, E. Dekemper, N. Loodts, C. Bingen, E. Kyrölä, J. Tamminen, V. Sofieva, A. Hauchecorne, F. Dalaudier, J.-L. Bertaux, G. Barrot, L. Blanot, O. Fanton d'Andon, T. Fehr, L. Saavedra, T. Yuan, and C.-Y. She. A global climatology of the mesospheric sodium layer from gomos data during the 2002 to 2008 period. *Atmospheric Chemistry and Physics*, 10(19):9225–9236, 2010.
- [2] C. Tétard, D. Fussen, F. Vanhellemont, C. Bingen, E. Dekemper, N. Matashvili, D. Pieroux, C. Robert, E. Kyrölä, J. Tamminen, V. Sofieva, A. Hauchecorne, F. Dalaudier, J.-L. Bertaux, O. Fanton d'Andon, G. Barrot, L. Blanot, A. Dehn, and L. Saavedra de Miguel. Oclo slant column densities derived from gomos averaged transmittance measurements. *Atmospheric Measurement Techniques*, 6(11):2953–2964, 2013.

6 Annex

Algorithm 1 Main program

```
clear
clc
load ../GomosList.mat
phi=85:-10:-85; zin=0:100; nbz=numel(zin); nbwave=2336; ifig=0;
for year=2002:2012
    moisdeb=1; moisfin=12;
    if year==2002; moisdeb=4; end
    if year==2012; moisfin=4; end
    for mois=moisdeb:moisfin
        latlist=1:18;
        for ilat=latlist
            L=find(geolocation(:,1)==year & geolocation(:,2)==mois & ...
                geolocation(:,11)>103.9 & geolocation(:,9)>=phi(ilat)-5 & geolocation(:,9)<phi(ilat)+5);
            nbOcc=numel(L);
            if nbOcc>10
                Tvor = nan(nbz,nbwave,nbOcc); dTvor = Tvor;
                lambda = nan(nbwave,nbOcc);
                %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
                % import/interp transmissions
                %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
                for i=1:nbOcc
                    filename=OccPath(L(i,:),);
                    EXT = coda_fetch(coda_open(filename));
                    lam = EXT.ext_nom_wav_assignment.nom_wl;
                    nb_proc = EXT.sph.num_lv2proc;
                    refstar_meas = EXT.ext_summary_quality.no_ref_star_comp;
                    sdp = EXT.ext_summary_quality.sdp_extract;
                    if nb_proc~=1 && refstar_meas~=2 && sdp==1
                        for j=1:nb_proc
                            T(j,:) = EXT.ext_mds(j).trans;
                            dT(j,:) = EXT.ext_mds(j).covar_func;
                            z(j) = EXT.ext_ads(j).tangent_alt;
                        end
                        z = z*1e-3;
                        Tint = interp1(z,T,zin);
                        dTint = interp1(z,dT,zin);
                    else
                        Tint = nan(nbz,nbwave);
                        dTint = Tint;
                    end
                    Tvor(:,i) = Tint;
                    dTvor(:,i) = dTint;
                    lambda(:,i) = lam;
                    clear z T dT Tint dTint lam
                    coda_clearall
                end
                %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
                % bimodality detection
                %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
                altj=0;
                for alt=26:5:36
                    wvl=1100; % detection performed at 600 nm
                    Ytest = squeeze(Tvor(alt,wvl,:));
                    [y,x]=hist(Ytest,30);
                    [param1,gof1,info1]=fit(x',y', 'gauss1');
```

Algorithm 2 Main program (following)

```
    altj=altj+1;
    sse(altj)=gof1.sse;
end
if sum(sse)>1000
    warning('maybe bimodal distribution : make a visual check')
    ifg=ifg+1;
    Bimo(ifg,:)=year,mois,phi(ilat),sum(sse);
    Bwarn = 1;
else
    Bwarn = 0;
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% outliers detection
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if Bwarn == 0
    Tclean=NaN(nbz,nbwave,nbOcc);
    dTclean=dTvor;
    for i=1:nbz
        for j=1:nbwave
            if numel(find(isfinite(squeeze(Tvor(i,j,:))))==1)>5
                [out,De,Ttri]=Jack(squeeze(Tvor(i,j,:)),2);
                Tclean(i,j,:)=Ttri; dTclean(i,j,out)=NaN;
            end
        end
    end
    Lfini=isnan(Tclean);
    dTclean(Lfini)=NaN;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% weighted medians
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    Tbin=NaN(nbz,nbwave); dTbin=Tbin;
    for i=1:nbz
        for j=1:nbwave
            [phi(ilat),mois,i,j,nbOcc]
            T=squeeze(Tclean(i,j,:));
            dT=squeeze(dTclean(i,j,:));
            Lt=(T~=0);
            [tb,dtb]=wpercentile(T(Lt),1./sqrt(dT(Lt)));
            Tbin(i,j)=tb; dTbin(i,j)=dtb;
        end
    end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% save file
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    if phi(ilat)>0 hemis='N'; else hemis='S'; end
    fln=['bin_',num2str(mois,'%2i'),num2str(year,'%2i'),'_',...
        num2str(min(abs(phi(ilat)-5),abs(phi(ilat)+5)),'%2i'),'_',...
        num2str(max(abs(phi(ilat)-5),abs(phi(ilat)+5)),'%2i'),hemis,'_NFD.mat'];
    save(fln,'zin','wvl','L','Tbin','dTbin','lambda')
end
end
end
end
end
```

Algorithm 3 Jackknife algorithm

```
function [outliers,DEV,Ttri]=Jack(T,st)
if nargin==1
    st=3;
end
if ~isempty(T)
    N = numel(T);
    for i = 1:N
        Tech=T;
        Tech(i)=[];
        DEV(i)=nanstd(Tech);
        NUM(i)=i;
    end
    Med=nanmedian(DEV);
    Etype=nanstd(DEV);
    L=DEV<Med - st*Etype;
    outliers=unique(NUM(L));
else
    outliers=0;
end
Ttri=T;
Ttri(outliers)=NaN;
```

Algorithm 4 Weighted median

```
function [wmed,gauss]=wpercentile(x0,w0,p,imet);
dyn=0.001;
wmed = NaN;
gauss = wmed;
l = isfinite(x0);
x = x0(l);
if not(exist('w0'))
    w=ones(size(x));
elseif isempty(w0)
    w=ones(size(x));
    w=w0(l);
end
w=w0(l);
if not(exist('imet','var'))
    l1=find((w./max(w)>dyn) & w>0);
    w1=w(l1);
    x1=x(l1);
    n1=length(x1);
    if n1 > 1
        i1=round(w1./min(w1));
        cum=cumsum(i1);
        m1=cum(end);
        pop=zeros(1,m1);
        for j=1:n1
            pop((cum(j)-i1(j)+1):cum(j))=x1(j);
        end
        if not(exist('p','var'))
            p=[(1-erf(1/sqrt(2)))/2 0.5 (1+erf(1/sqrt(2)))/2];
            p=100*p;
            dum=percentile(pop',p);
            wmed=dum(2);
            gauss=dum(3)-dum(1);
        else
            wmed=percentile(pop',p);
            gauss=[];
        end
    elseif n1==1
        wmed = x1;
        gauss = w1;
    end
end
end
```
